Online-Journals.org

iJOE International Journal of Online and Biomedical Engineering

iJOE | eISSN: 2626-8493 | Vol. 21 No. 13 (2025) | OPEN ACCESS

https://doi.org/10.3991/ijoe.v21i13.58335

PAPER

Artificial Neural Networks as a Support and Learning Tool in Medical Practice

Leonardo Moraes Armesto (✉), Thabata Roberto Alonso, Daniel Souza Ferreira Magalhães, Laurita dos Santos

Scientific and Tecnological Institute at Universidade Brasil (UNIVBRASIL), São Paulo, Brasil

leonardo.armesto@ub.edu.br

ABSTRACT

This study compared the performance of an artificial neural network (ANN), implemented as a convolutional neural network (CNN), with biological neural networks (BNNs) represented by medical students, residents, and specialists. The task consisted of classifying magnetic resonance imaging (MRI) scans in both binary (physiological vs. pathological) and multiclass settings (physiological, Chiari malformation, cortical degeneration, and brainstem glioma). The CNN, trained in MATLAB, achieved 100% accuracy (acc) (AUC = 0.99) in binary classification and 72.5% acc (AUC = 0.78) in multiclass classification, consistently outperforming all human groups, whose maximum acc reached 50%. Additional metrics (precision, recall, and F1-score) confirmed the network's robustness, while statistical analyses (chi-square test, 95% CI) revealed no significant correlation between participants' expertise and diagnostic performance. These findings demonstrate the superior diagnostic capacity of CNNs and emphasize their potential as complementary tools in medical practice. Moreover, they highlight the educational relevance of CNNs, suggesting their role in supporting the development of anatomical and diagnostic skills and in bridging knowledge gaps during medical training.

KEYWORDS

artificial neural network (ANN), diagnostic support systems, pattern recognition, medical education

1 INTRODUCTION

The use of biomedical instruments and devices has intensified significantly across various healthcare and patient care settings [1]. While continuously evolving in innovation—enhancing diagnosis, prevention, treatment, and rehabilitation—there remains a broad field where technological development aligns with operational and clinical needs [2].

In 2023, the global biomedical device market was valued at US$525.8 billion and is projected to reach US$957.8 billion by 2033, with an estimated annual growth

Armesto, L. M., Alonso, T. R., Magalhães, D. S. F., dos Santos, L. (2025). Artificial Neural Networks as a Support and Learning Tool in Medical Practice. *International Journal of Online and Biomedical Engineering (iJOE)*, 21(13), pp. 151–168. https://doi.org/10.3991/ijoe.v21i13.58335

Article submitted 2025-08-19. Revision uploaded 2025-09-09. Final acceptance 2025-09-10.

rate of 6.2% from 2024 to 2033 [3]. This growth reinforces the complementary role of biomedical devices, particularly those powered by artificial intelligence (AI), in supporting diagnostic processes and increasing acc across medical specialties [1]. In addition to accelerating treatment delivery, AI helps optimize healthcare resources by enabling earlier diagnoses [4]. Despite this technological advancement, foundational sectors such as technical and higher education still lack adequate information systems and computational tools to support effective teaching and learning processes. This gap can impact the development of well-trained, technically skilled professionals [5].

In response, academic curricula have increasingly emphasized hands-on learning and the integration of AI tools—especially machine learning and, more specifically, deep learning based on artificial neural networks (ANNs)—to enhance medical education and training [6]. ANNs are mathematical models inspired by biological neural networks (BNNs) capable of processing information through artificial neurons [7] and [8]. Among these, convolutional neural networks (CNNs) are widely used for pattern recognition in medical imaging. Recent studies [9–11] highlight how medical education has embraced AI as a tool to enhance diagnostic reasoning and investigative skills. This aligns with the concept of Medicine 4.0, where ongoing education and the integration of machine learning are viewed as essential for modern medical practice [12,13].

While CNN-based pattern recognition is well established, its application in educational contexts remains underexplored [8]. CNNs have demonstrated high acc in several medical domains, including pulmonology, gynecology, and neurology, confirming their value in clinical diagnostics [14–17]. However, unlike humans, CNNs lack reasoning and contextual adaptation, which can limit their robustness when dealing with noisy or uncertain images. Recent work [18] highlights that preprocessing choices, feature selection, and CNN architecture strongly influence performance, but further progress may require borrowing concepts from adaptive and nonlinear control. Techniques such as fuzzy control, backstepping, and neural adaptive control could improve CNN adaptability and learning speed, while outputfeedback and optimal control approaches may open pathways for combining human reasoning with machine pattern recognition. From an educational perspective, this integration is particularly relevant: although studies support the role of AI in teaching and learning [7,9–11,19,20], little is known about how ANNs can specifically enhance medical education. Some evidence indicates that AI tools can foster knowledge retention and the development of core competencies in healthcare professionals [19]. Moreover, deep learning—through its multilayered architecture and repetitive training—mirrors aspects of biological learning, where reinforcement mechanisms are central to the functioning of BNNs [20–22].

In this context, the present study compares the performance of ANNs and BNNs in classifying brain MRI images. Beyond evaluating diagnostic accuracy, it investigates how CNNs can complement medical education by benchmarking human and machine performance. This work also contributes to the ongoing discussion on integrating AI into medical curricula, supporting the formation of agile and tech-savvy healthcare professionals. As medicine increasingly incorporates information systems and AI into daily practice, understanding how professionals interact with these tools becomes essential. The paper is structured as follows: Section 2 details the methodology, including the CNN network architecture, systematization and learning flow, data acquisition and preprocessing procedures, study participants, questionnaire application, and the comparison between networks. Section 3 presents the results of the BNN and ANN performance analyses. Section 4 discusses the findings, and Section 5 provides the conclusions and perspectives for future research.

2 METHODOLOGY

The methodological design of this study began with the development of an ANN using a CNN architecture, implemented in MATLAB. Problem-based questions were then formulated using neurological images representing both healthy and pathological conditions. Inparallel, a group of participants—referred to here as BNNs—was recruited. This group included individuals with varying levels of expertise in neurology: medical students, neurology residents, and certified neurologists. Both the ANN and the BNN participants were presented with the same set of problem questions based on the same medical images. However, only the ANN received preprocessed images, following established literature protocols to enhance model performance and ensure fair comparison. Based on the ANN's output, confusion matrices were generated and acc values were calculated. These results were then compared with the BNN responses to evaluate which network (artificial or biological) demonstrated superior classification performance. Additionally, a domain-based comparison was conducted to identify which anatomical, physiological, and pathological dimensions each network handled with higher accuracy.

All results were tabulated and subjected to statistical analysis for comparative interpretation. This section presents a detailed description of the CNN architecture developed for the study, including input classes, activation functions, the learning algorithm, and classification layers. It also outlines the strategy used to divide the dataset into training, validation, and testing sets. The process of image selection and preprocessing is described, along with the repository used for CNN training. Additionally, the recruitment procedures for the BNN group are explained, including the criteria used to define participants' levels of neuroradiological expertise. The section also covers the development and implementation of the image-based problem questions administered to both the ANN and BNN groups, as well as the methods employed to obtain and analyze the resulting data.

2.1 CNN network architecture

A CNN developed for this study followed a standard deep learning design for image classification, as reported in the literature [23–26]. The network consisted of an input layer, two convolutional layers interleaved with pooling layers, and a final fully connected layer. Convolutional blocks used 3×3 filters with increasing feature maps (16, 32, and 64), ReLU activation, batch normalization, and 2×2 max-pooling. A dropout layer (30%) was applied to mitigate overfitting. The output layer employed Softmax for probability-based classification into binary (two-class) or multiclass (four-class) tasks.

A deterministic approach was used, with fixed architecture parameters to ensure reproducibility. The input to the network consisted of 256×256-pixel grayscale images. Figure 1 and Table 1 illustrate the overall CNN structure and the specific characteristics of each layer.

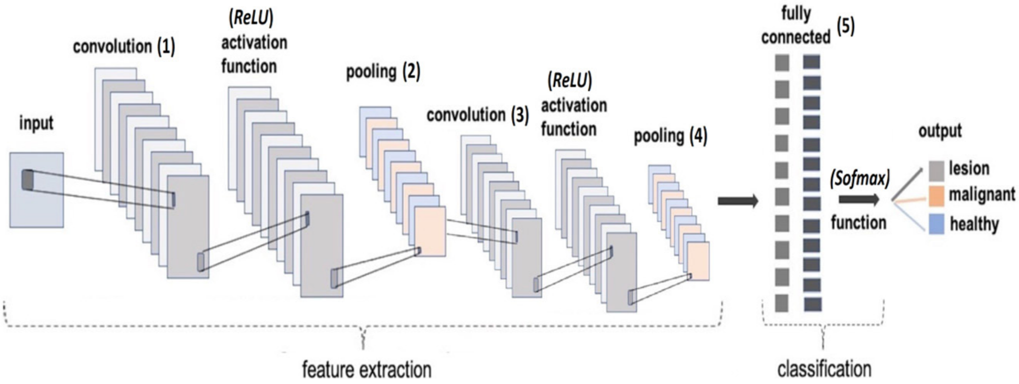


Fig. 1. CNN architectural structure

Training was conducted using MATLAB's Neural Networks Toolbox on an Intel Core i5 processor (2 GHz) and 16 GB of RAM. The training process included 40 iterations (four per epoch) across 10 epochs, using mini-batches of 32 images. The learning rate was randomly selected from a uniform distribution within the range of 10^{-5} to 10^{-2}. Optimization was carried out using both the Adam and Stochastic Gradient Descent (SGD) algorithms. Shuffling occurred at the start of every epoch ('Shuffle,' 'every-epoch'), and validation was performed in parallel with training to monitor model performance in real time.

Table 1. CNN structural specification

Type:	*Convolutional.*
Function:	*Extract basic features from imputimages.*
Filter Size:	*Determined dynamically – capture relevant patterns.*
Activation:	*ReLU (Rectified Linear Unit) – introduce non-linearity.*
Input:	*256×256 pixel grayscale images.*
Output:	*Convolutional feature maps.*
Type:	*Pooling.*
Function:	*Reduce the dimensionality of feature maps.*
Pool Size:	*MaxPooling or AveragePooling: determined dynamically – preserves important detail.*
Input:	*Feature maps from Layer 1*
Output:	*Feature maps reduced in size.*
Type:	*Convolutional.*
Function:	*Extract complex and abstract features from images.*
Filter Size:	*Determined dynamically – captures patterns: high level.*
Activation:	*ReLU (Rectified Linear Unit).*
Input:	*Reduced feature maps from Layer 2.*
Output:	*Convolutional feature maps.*

(Continued)

Table 1. CNN structural specification *(Continued)*

Type:	*Pooling.*
Function:	*Further reduce dimensionality while retaining key features*
Pool Size:	*MaxPooling or AveragePooling: determined dynamically*
Input:	*Feature maps from Layer 3.*
Output:	*Feature maps reduced in size.*
Type:	*Fully Connected.*
Function:	*Classify extracted features into target categories.*
Neurons:	*Number based on classification task (2 or 4 classes).*
Classification:	*Softmax generates a probability distribution over classes.*
Input:	*Vector of features from the layers.*
Output:	*Probabilities of belonging to each class.*

2.2 Systematization and learning flow

The CNN was trained and validated using separate datasets, with images propagated through the network using a feedforward process to compute the output. The difference between the predicted and actual values (error) was used in the backpropagation process to update the network's internal parameters. Optimization was performed using both the Adam and SGD algorithms, which adjusted the network's weights to minimize the loss function. The feedforward algorithm is characterized by a unidirectional flow of information—from input to output—without any feedback loops or cyclical connections influencing the inputs [26].

After the training and validation phases, the CNN was testing dataset consisting of previously unseen images to assess its generalization ability and classification accuracy. To enhance the reliability of the evaluation, a 10-fold cross-validation approach was applied during training and validation, followed by a separate testing round. This systematic flow is illustrated in Figure 2.

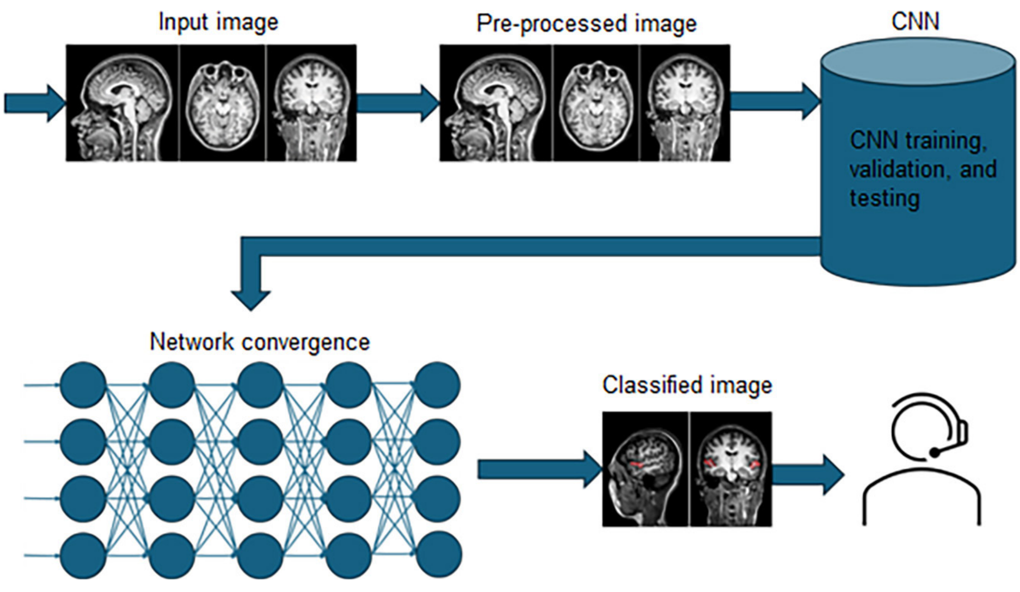


Fig. 2. CNN learning-analytical flow

All images were appropriate (no images were added that were translated, rotated, blurred, or representative of other conditions/types that did not correspond to the study). Examples of images composed in the databases for extraction to the entry (a) and exit (b) questionnaires can be seen in Figure 2.

2.3 Data acquisition and preprocessing

The CNN was trained using a dataset of 200 images, evenly distributed across four classes. The physiological (healthy) images (n = 50) were sourced from a database provided by the Neuroimaging Laboratory at the State University of Campinas, under ethics committee approval (opinion number 3,657,509). These images had been previously used in scientific research [27]. The remaining 150 pathological images representing three commonly studied neurological conditions often evaluated through magnetic resonance imaging (MRI) [28–30]: Chiari malformation (n = 50), Alzheimer's disease (n = 50), and brainstem glioma (n = 50), were obtained through public access in open databases [31] and [32] from different individuals. All images were acquired from distinct individuals.

The data was divided as follows: 80/20—training and testing. Of these, the training data was further divided into 80/20—training and validation. This percentage distribution is repeated in each iteration, randomly, with the images separated in the indicated proportion. For all cases (of input diversity), a balanced group of images from each class was considered.

Prior to input into the CNN, images underwent a standardized preprocessing pipeline: application of a Gaussian filter to reduce noise, followed by a sharpening filter to enhance edges; intensity normalization to the [0, 1] range; resizing to 256×256 pixels; and systematic renaming within organized folders by class (1 to 50). It is important to note that, for the BNN participants, the same images were presented without any of the preprocessing steps described above, preserving the raw visual input for human assessment.

The differentiation in the steps described for image preprocessing between BNN and CNN is justified by the structural differences between digital and cognitive processing. CNNs rely on pixel patterns and lack innate mechanisms for contextual perception or attentional selectivity, a fact that differentiates them from humans in their visual system. In detail, just as sharpening filters help highlight edges and subtle but significant details, and Gaussian filters smooth out unwanted noise, improving learning stability and network generalization, humans interpret images using context, memory, previous experiences, and complex visual mechanisms. These capabilities allow them to perceive relevant information in natural images, whereas the application of artificial filters could alter the appearance and interfere with contextual interpretation, resulting in a failure to represent the true perception. Thus, by selectively using CNN filters, CNN and BNN can be compared under conditions that enhance their respective digital learning capabilities.

2.4 Study participants

In this study, individuals were considered representatives of BNNs. A total of 60 participants were targeted, divided into three groups: 40 undergraduate medical students (average class size), 10 neurology residents, and 10 board-certified neurologists.

All ethical principles involving human subjects were strictly observed. The study protocol was approved by the Ethics Committee of Universidade Brasil (approval number: 6,097,741) and conducted in accordance with Resolution No. 466/12 of the National Commission for Ethics in Research (CONEP), under the Brazilian National Health Council. Eligible participants were adults who provided informed consent (ICF) and met the following inclusion criteria: (i) enrollment in a medical program with prior completion of a neurology course (for students); (ii) current enrollment in a neurology residency program (for residents); or (iii) certified specialization in neurology (for professionals).

Exclusion criteria included: failure to provide informed consent via the electronic questionnaire form; voluntary withdrawal or discomfort regarding participation; enrollment in a non-medical course or lack of exposure to neurology content (for students); residency in a medical specialty unrelated to neurology; and professional practice outside of neurology or neuropediatrics. Additionally, participants who failed to complete both the entry and exit questionnaires were excluded from the final analysis.

2.5 Questionnarie application

The development of the questionnaire aimed to evaluate participants' ability to recognize patterns in neurological images, based on the level of academic or professional anatomical and physiological knowledge required at each stage of medical training. As highlighted in prior studies [1,10,33], effective healthcare practice depends on a foundation of well-structured academic and clinical training, supported by tools that enhance diagnostic reasoning and decision-making. The continuous assessment of knowledge and technical competencies is therefore essential to improving diagnostic performance and delivering patient-centered care [34,35].

In this context, an analytical questionnaire was constructed and applied to the BNN groups. It comprised five multiple-choice questions. The first four questions focused on technical knowledge in neuroradiology, while the fifth aimed to assess participant motivation. The technical questions addressed the following areas: question 1: fundamentals of morphofunctional/anatomophysiological conformation; question 2: structural, dimensional, and volumetric recognition; question 3: density/signal analysis and zonal detection; and question 4: neuroanatomical topography and planimetric delimitation.

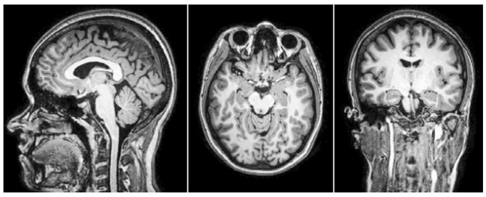
Only questions 1 and 2 were the same for all BNN groups, as they targeted foundational knowledge considered essential regardless of academic level. Questions 3 and 4 addressed similar themes but were adapted in wording and complexity to each group's expertise level. To allow for comparative evaluation between BNN and ANN, the analysis in this study focused exclusively on questions 1 and 2, as they were consistently applied across all human groups and reflected core pattern recognition skills. The same image-based input used for the BNN questionnaire was presented to the CNN, which classified the images and generated results via confusion matrices, providing corresponding accuracy values.

In detail, the methodological definition for choosing the two aforementioned questions is justified by the research proposal, which aimed to compare the results obtained for the same questions between BNN and CNN, as well as within BNN groups, establishing equivalence and more targeted results. The topics covered in each question (1 and 2) analyzed were: Morphofunctional/anatomophysiological fundamentals and structural, dimensional, and volumetric recognition, respectively.

Thus, aiming at the comparative equivalence of correct answers, both between the BNN groups and between BNN and ANN, this study focused on the analysis of the same questions that also addressed basic knowledge: questions 1 and 2. The questions sought to analyze, in the groups and in the CNN, the recognition of patterns focused on knowledge subjects.

The BNN groups' responses were collected as correct or incorrect, while the CNN's classifications were converted into the same binary outcome using its confusion matrices. A chi-square test was applied for bivariate analysis to examine potential correlations between performance acc and knowledge level across both ANN and BNN groups. The questionnaire content is detailed in Table 2.

Table 2. Issues applied to networks

Questionnaire Image – BNN Bank Image Example – ANN	Statement
	*Is the image representative of a physiological (normal) presentation?* **a)** *Yes (correct)* **b)** *No (incorrect)*
This question does not include a new image; it refers back to the image shown in Question 1.	*If the image was classified as non-physiological in the previous question, which anatomical region is affected? If classified as physiological, indicate "not applicable."* **a)** *Not Applicable (correct)* **b)** *Prefrontal cortex (incorrect)* **c)** *Cerebellum (incorrect)* **d)** *Bulb (incorrect)*

2.6 Comparison between networks

The comparative analysis between the ANN and the BNN groups was performed using data tabulated in Microsoft Excel. Descriptive statistics were applied, and results were expressed as absolute and relative frequencies. Bar charts were used to visualize the distribution of correct and incorrect responses across groups and questions. To assess statistical associations between variables (responses to each question) among the BNN subgroups—and between BNN and ANN—a bivariate analysis using the Chi-square test (χ^2) was conducted.

For the ANN, confusion matrices were generated based on its classifications. From these matrices, the following performance metrics were calculated: acc, precision (pre), recall (re), F1-score, sensitivity (VPR), and false positive rate (FPR, or 1-specificity).

$$acc(\%) = \frac{VP + VN}{VP + VN + FP + FN} \times 100 \tag{1}$$

$$pre(\%) = \frac{VP}{VP + FP} \times 100 \tag{2}$$

$$rec(\%) = \frac{VP}{VP + FN} \times 100 \tag{3}$$

$$F1(\%) = 2 \times \frac{Precision \times Recall}{Precision + Recall} \times 100 \tag{4}$$

$$VPR(sensitivity) = \frac{VP}{VP + FN} \tag{5}$$

$$FPR(1 - specificity) = \frac{FP}{FP + VN} \tag{6}$$

3 RESULTS

This section presents the performance analysis of both BNNs and the ANN in answering the classification tasks based on neuroradiological images.

3.1 BNN performance

The performance of the BNN groups (medical students, neurology residents, and neurology specialists) was evaluated through absolute and relative frequency analysis. A bivariate chi-square (χ^2) test was used to assess the association between professional level and number of correct responses per question. Table 3 summarizes the acc results for each group across both questions. For question 1, which required identifying whether the image showed a physiological brain, the correct answer rate varied from 30% to 41.18% across groups. For question 2, which involved identifying the affected anatomical region (or selecting "Not Applicable" for physiological images), the acc ranged from 38.24% to 50%.

Table 3. Accuracy (Correct Answers) per group and question

Group (BNN)	Question 1	Question 2
Medical studen	41.18%	38.24%
Resident in Neurology or Related Field	30.00%	50.00%
Specialist in Neurology or Related Field	40.00%	50.00%

Statistical analysis was conducted at a 5% significance level ($p < 0.05$) with a 95% confidence interval. The p-values for the chi-square tests were $p = 0.8136$ for question 1 and $p = 0.2037$ for question 2. In both cases, there was no statistically significant difference between the groups, indicating that the null hypothesis (H_0)—that there is no association between the professional level and performance—could not be rejected. This suggests that increased training or specialization in neurology did not result in significantly improved acc in the classification tasks evaluated.

3.2 ANN performance

The ANN was applied to two distinct tasks—corresponding to Question 1 (binary classification) and Question 2 (multiclass classification)—with 18 variations tested for each, containing six empirical values for each of the following parameters: Random seeds, maxepochs and iterations. Each question used a separate input configuration, with the ANN trained, validated, and tested independently for both tasks, as detailed in Table 4. These tests aimed to identify the influence of such parameters on the network's performance.

Table 4. Accuracy values obtained from different empirical values proposed for parameters (random seed, maxepochs, and iterations)

Parameters	Empirical Values	Accuracy (acc)	
		Question 1	Question 2
Random seeds	250	100%	65.0%
	500	100%	65.0%
	750	100%	70.0%
	1000	100%	65.0%
	1250	100%	72.5%
	1500	100%	55.0%
Maxepochs	10	100%	72.5%
	12	100%	60.0%
	14	100%	70.0%
	16	100%	62.5%
	18	100%	67.5%
	20	100%	70.0%
Iterations	10	100%	72.5%
	15	100%	72.5%
	20	100%	72.5%
	25	100%	72.5%
	30	100%	72.5%
	35	100%	72.5%

Thus, based on Table 4, convolutional architecture models [36–38], as well as others focused more specifically on applied and experimental work [18,39,40], having analyzed the scripts for a convolutional architecture model, the best-performing configuration, particularly for question 2, was achieved using a random seed of 1250, 10 training epochs, and 10 iterations, showing optimal acc without requiring increased complexity or additional training time. For question 1, the ANN consistently achieved 100% acc across all tested parameter combinations, indicating strong stability and reliability in binary classification. Additionally, modifying the number of layers in the CNN architecture did not lead to acc improvements but did increase computational cost.

The performance was further assessed using confusion matrices (Figures 3 and 4), providing detailed insights into classification behavior across both tasks. For question 2, while overall acc remained high, some misclassifications were observed, particularly between anatomically adjacent structures (e.g., cerebellum vs. bulb). The class-wise performance metrics and training/validation/testing splits are presented in Table 5. These findings validate the CNN's capacity for accurate image-based diagnosis, especially when properly tuned, and highlight its comparative advantage over BNN groups in classification precision.

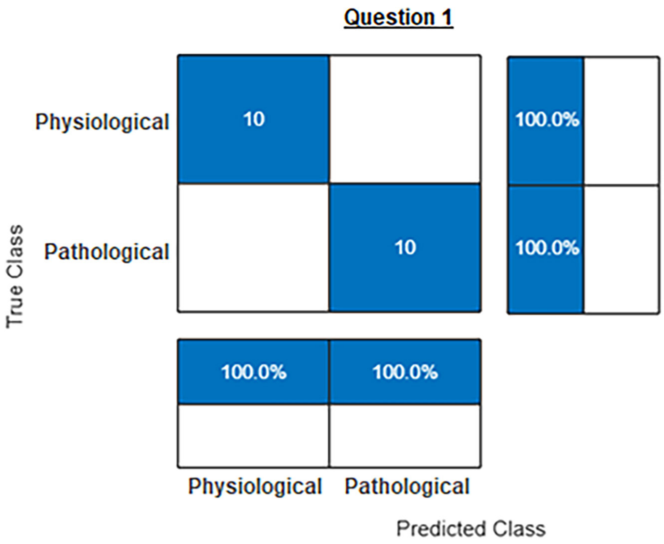


Fig. 3. Confusion matrices by question 1 – ANN

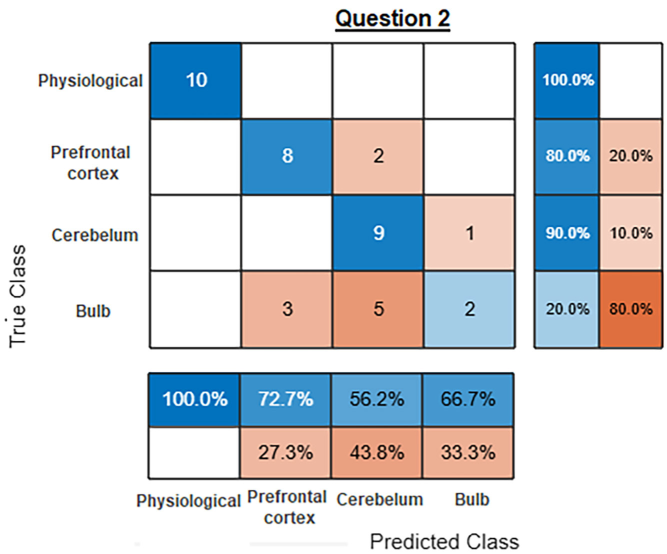


Fig. 4. Confusion matrices by question 2 – ANN

Table 5. Distribution: training/validation/testing, by question

Question	Class	Training/Validation N°* → 80% (64/16)		Test N°* → 20%
Q1	2	– Physiological; – Pathological.	– Physiological; – Pathological.	– Physiological; – Pathological.
Q2	4	– Physiological; – Cortical degen; – Chiari; – Glioma.	– Physiological; – Cortical degen; – Chiari; – Glioma.	– Physiological; – Cortical degen; – Chiari; – Glioma.

From the confusion matrices (Figures 3 and 4), the *acc* obtained for each question were calculated to compare ANN and BNN performance. Additionally, precision (*pre*), recall (*rec*), and F1-score (*F*1) metrics were derived for further benchmarking and discussion (Table 6).

Table 6. Parameters obtained from ANN

Parameters	Question 1	Question 2
Accuracy (*acc*)	100%	72.5%
Precision (*pre*)	100%	73.9%
Recall (*re*)	100%	72.5%
F1-score (*F*1)	100%	69.3%

To provide a more comprehensive evaluation of model performance, the ROC (Receiver Operating Characteristic) curves were also analyzed (see Figure 5). These curves demonstrate the trade-off between true positive rate (TPR, or sensitivity) and false positive rate (FPR, or 1 – specificity), allowing us to assess the classifier's behavior across different decision thresholds.

For question 1 (binary classification), the area under the curve (AUC) was approximately 0.99, which reflects excellent discriminatory ability between physiological and pathological images. The curve closely follows the upper-left boundary of the plot, indicating that the CNN correctly classified nearly all test images. For question 2 (multiclass classification), the AUC value was 0.78. Although lower than the binary classification performance, this value still exceeds the 0.5 baseline of random guessing, indicating acceptable classification ability for improvement. Possible improvements include adjusting classification thresholds, refining the model architecture, or increasing the training dataset.

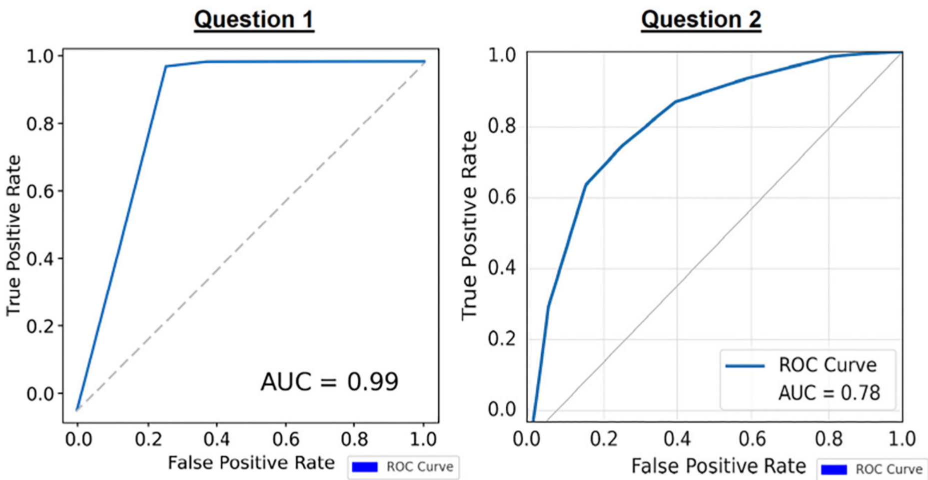


Fig. 5. ROC curves – ANN

For question 1, the ANN showed perfect classification (95% CI: [1.00, 1.00]), while the BNN groups achieved the following: Medical Students: 95% CI [31.53%, 50.83%]; Neurology Residents: 95% CI [21.02%, 38.98%]; Neurology Specialists: 95% CI [30.40%, 49.60%]. For question 2, the ANN achieved a 95% CI of [63.75%, 81.25%]. The BNN groups presented: Medical Students: 95% CI [28.71%, 47.76%] and Neurology Residents and Specialists: 95% CI [40.20%, 59.80%].

Finally, chi-square analyses were conducted to assess statistical significance in the ANN classification results per class. Using a 5% significance level and 95% confidence interval, no statistically significant differences were observed in ANN classification acc for either question 1 (p-value = 0.6756) or question 2 (p-value = 0.9997).

These findings confirm the consistency of the model's predictions across different classes. The comparative statistical results for ANN and BNN are summarized in Figure 6.

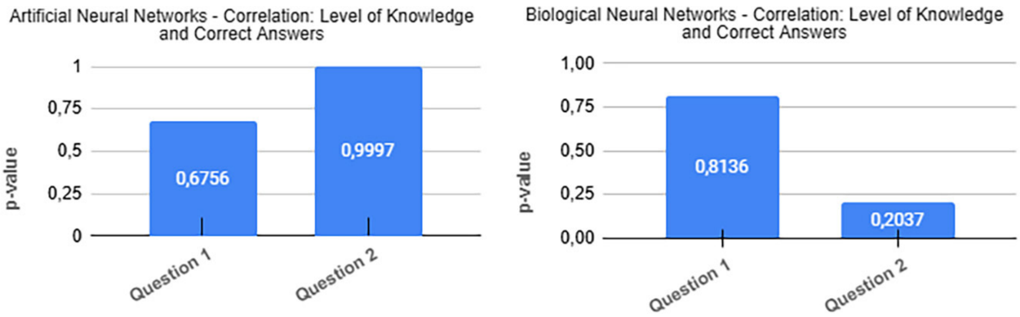


Fig. 6. Network correlation: correct answers obtained and level of knowledge about the subject addressed

4 DISCUSSION

This section compared the performance of the BNN groups and the ANN, highlighting key differences in acc and their implications. For question 1, which assessed the ability to identify relational aspects between anatomical structures and their areas of integration, BNN group 1 (medical students) achieved an acc of 41.18%, group 2 (residents in neurology or related fields) obtained 30%, and group 3 (specialists) reached 40%. The ANN, in contrast, achieved 100% acc in distinguishing between physiological and pathological classes. Interestingly, medical students and specialists performed similarly, surpassing residents. This may reflect the balance between ongoing training in neuroanatomy for students and routine clinical exposure for specialists, whereas residents often face workload-related constraints that limit opportunities for detailed image analysis. These findings highlight the superior performance of the CNN compared to human participants across different training levels.For question 2, which involved more complex classification (proportion, extent, dimensionality, and volume of brain structures), group 1 reached 38.24% accuracy, while groups 2 and 3 both achieved 50%. Again, the ANN outperformed the BNNs with 72.5% accuracy. Although the performance gap narrowed in this multiclass task, the CNN maintained a clear advantage. Confidence interval (CI) analysis further reinforced the robustness of CNN predictions (see Figure 6): perfect acc for question 1 with CI [1.00, 1.00], and consistently higher intervals for question 2 ([63.75%, 81.25%]) compared to BNNs.

The chi-square analysis showed no statistically significant association between participants' professional levels and correct/incorrect answers for either question (p-values: 0.8136 and 0.2037 for BNNs; 0.6756 and 0.9997 for ANN), suggesting that domain expertise did not predict classification performance in this setting. These findings align with prior literature where CNNs have demonstrated high acc in diverse medical imaging tasks. For example, CNNs achieved 97.72% acc in pulmonary pattern recognition [14], competitive results in cytopathology [15], and high metrics in pharmacogenomics [41], ductal carcinoma [42], COVID-19 [43], and brain tumor detection via MRI [17]. Such evidence reinforces the applicability of CNNs across medical domains.

Beyond performance comparisons, the results have important implications. Clinically, CNNs demonstrate strong potential for decision support, particularly in contexts requiring rapid and precise recognition, such as emergency care or triage, and in resource-limited environments lacking specialists. Educationally, CNNs offer valuable opportunities to strengthen learning strategies. By providing immediate, standardized, and quantitative feedback, CNNs can complement problem-based learning, peer instruction, flipped classrooms, and prototyping-based teaching [7–10]. They can be integrated into curricula to personalize training in neuroanatomy and diagnostic imaging, identifying individual difficulties, adapting case complexity, and allowing repetitive practice in controlled environments. This integration could accelerate skill acquisition, improve consistency, and expand access to high-quality training resources.

When comparing CNNs and BNNs, it is important to recognize their distinct cognitive processes. Human reasoning relies on accumulated knowledge, contextual experiences, and interpretative flexibility, while CNNs are limited to detecting and analyzing visual patterns [44,45]. Despite these differences, continuous and varied exposure to cases can progressively enhance human recognition skills, narrowing the performance gap with CNNs [46]. In educational contexts, this finding supports the use of CNNs as pedagogical tools to strengthen consistency in human learning, even when prior experiences differ [47,48].

The dataset focused on a controlled subset of neurological conditions, limiting generalizability. The number of participants and images was constrained by availability, reducing statistical power. Furthermore, factors such as cognitive load, response time, and participant motivation—all relevant in real diagnostic practice—were not measured. Potential biases include the restricted participant pool, differences in workload and clinical exposure, and the artificial testing conditions (e.g., absence of time constraints and clinical information). These aspects may have influenced the results. Future studies should therefore expand to multicenter databases, include a broader range of conditions and participant profiles, investigate real-time diagnostic performance, and align experiments more closely with clinical and educational contexts. Such efforts will strengthen external validation and ensure that CNN applications in health and medical education are both reliable and generalizable.

5 CONCLUSION

This study compared CNNs with BNNs, represented by medical trainees and specialists, in classifying neurological imaging patterns. CNNs consistently outperformed human participants, highlighting their potential as complementary tools in both diagnostics and education. Importantly, humans and CNNs rely on distinct mechanisms: humans draw on anatomical knowledge and prior experience, whereas CNNs detect statistical patterns without contextual understanding. Combining these complementary strengths could enhance acc and reliability beyond either system alone. Future research should therefore explore hybrid human–machine models, as well as adaptive control strategies, to build more robust and interactive educational and diagnostic frameworks. By advancing this synergy, CNNs may enrich medical training, refine diagnostic competencies, and support innovative practices in healthcare.

6 ACKNOWLEDGMENT

The authors thank Dr. Enio Galli Ferlin, from the Department of Neurology at ABC Medical School (FMABC), for his valuable neuroradiological guidance and recommendations to relevant research groups, and the Brazilian Federal Agency for Support and Evaluation of Graduate Education (CAPES/BRAZIL) for their support. L.M.A. is grateful to Universidade Brasil (UNIVBRASIL) for the doctoral scholarship granted, which contributed significantly to the development of this study, as well as to Faculdade Focus for the financial support provided for this publication. L.d.S. would like to thank the National Council for Scientific and Technological Development (CNPq grant No. 304967/2024-6) for their financial support. (Corresponding authors: Leonardo Moraes Armesto and Laurita dos Santos).

7 REFERENCES

[1] A. A. Rajathi, A. A. S. Raj, P. Rajalakshmy, and K. G. J. Nigel, "Review on biomedical instrumentation/biomedical engineering and its various applications," *International Journal of Advanced Research in Electrical*, vol. 3, no. 3, pp. 7707–7719, 2014.

[2] A. Kumari, S. Tanwar, S. Tyagi, and N. Kumar, "Fog computing for Healthcare 4.0 environment: Opportunities and challenges," *Computers and Electrical Engineering*, vol. 72, pp. 1–13, 2018. https://doi.org/10.1016/j.compeleceng.2018.08.015

[3] Allied Market Research, "Medical devices market size," *Share|Forecast Report* – 2033, 2024.

[4] D. R. Pai, "Complexities of simultaneously improving quality and lowering costs in hospitals; comment on hospitals bending the cost curve with increased quality: A scoping review into integrated hospital strategies," *Health Policy and Management*, vol. 12, no. 1, pp. 1–3, 2023. https://doi.org/10.34172/ijhpm.2022.7442

[5] R. Saxena, J. L. Shobe, and B. L. Mcnaughton, "Learning in deep neural networks and brains with similarity-weighted interleaved learning," *Proceedings of the National Academy of Sciences Journal*, vol. 119, no. 27, pp. 1–11, 2022. https://doi.org/10.1073/pnas.2115229119

[6] G. M. Ven, H. T. Siegelmann, and A. S. Tolias, "Brain-inspired replay for continual learning with artificial neural networks," *Nature*, vol. 11, no. 4069, pp. 1–14, 2020. https://doi.org/10.1038/s41467-020-17866-2

[7] T. Pircher, E. Schlucker, and A. Feigenspan, "The structure dilemma in biological and artificial neural networks," *Nature*, vol. 11, no. 5621, pp. 1–16, 2021. https://doi.org/10.1038/s41598-021-84813-6

[8] O. S. Eluyode and D. T. Akomolafe, "Comparative study of biological and artificial neural networks," *European Journal of Applied Engineering and Scientific Research*, vol. 2, no. 1, pp. 36–46, 2013.

[9] Z. Lin, P. Chen, and L. Chen, "Artificial Intelligence in education: A review," *IEEE Access Multidisciplinary*, vol. 8, no. 6, pp. 75264–75278, 2020. https://doi.org/10.1109/ACCESS.2020.2988510

[10] L. C. Lobo, "Inteligência artificial, o Futuro da Medicina e a Educação Médica," *Revista Brasileira de Educação Médica*, vol. 42, no. 3, pp. 3–8, 2018. https://doi.org/10.1590/1981-52712015v42n3rb20180115editorial1

[11] J. R. Silva, F. B. Medeiros, W. S. Bessa, and E. L. M. Bezerra, "Uso das tecnologias de informação e comunicação no curso de medicina da UFRN," *Revista Brasileira de Educação Médica*, vol. 39, no. 4, pp. 537–541, 2015. https://doi.org/10.1590/1981-52712015v39n4e02562014

[12] F. H. C. Ferraz, "As tecnologias na medicina 4.0 e a resistência médica aos novos sistemas," *Revista Científica Multidisciplinar Núcleo do Conhecimento*, vol. 1, no. 2, pp. 5–18, 2022. https://doi.org/10.32749/nucleodoconhecimento.com.br/tecnologia/tecnologias-na-medicina

[13] A. G. L. S. Mesquita, “The importance of continuing education: Professional improvement facing the challenges of the 21st century,” *Research, Society and Development*, vol. 10, no. 9, pp. 1–9, 2021.
[14] L. O. Silva and A. D. F. Santos, “Redes neurais convolucionais aplicadas na detecção de pneumonia através de imagens de Raio-X,” in *XIV Congresso Brasileiro de Inteligência Computacional – Artigos Apresentados Nas Sessões Técnicas*, Marabá, 2019, pp. 1–8. https://doi.org/10.21528/CBIC2019-24
[15] L. B. Croffi and G. A. Bertoli, “Redes neurais convolucionais com especificidade para diagnóstico de carcinoma em exames de Papanicolau,” *Revista Fatec*, vol. 10, no. 75, pp. 1–6, 2020.
[16] D. R. Sarvamangala and R. V. Kulkarni, “Convolutional neural networks in medical image understanding: A survey,” *Evolutionary Intelligence*, vol. 15, no. 1, pp. 1–22, 2022. https://doi.org/10.1007/s12065-020-00540-3
[17] Z. Rguibi, A. Hajami, D. Zitouni, A. Elqaraoui, and A. Bedraoui, “CXAI: Explaining convolutional neural networks for medical imaging diagnostic,” *Electronics*, vol. 11, no. 1775, pp. 1–19, 2023. https://doi.org/10.3390/electronics11111775
[18] C. Jie, L. Jiawei, W. Shulin, and Y. Sheng, “Feature selection in machine learning: A new perspective,” *Neurocomputing*, vol. 300, pp. 70–79, 2018. https://doi.org/10.1016/j.neucom.2017.11.077
[19] F. A. Mohammed, K. K. Tune, B. G. Assefa, M. Jett, and S. Muhie, “Medical image classifications using convolutional neural networks: A survey of current methods and statistical modeling of the literature,” *Machine Learning & Knowledge Extraction*, vol. 6, no. 1, pp. 699–735, 2024. https://doi.org/10.3390/make6010033
[20] R. Siqueira-Batista *et al.*, “As redes neurais artificiais e o ensino da medicina,” *Revista Brasileira de Educação Médica*, vol. 38, no. 4, pp. 548–556, 2014. https://doi.org/10.1590/S0100-55022014000400017
[21] G. R. Yang and X. J. Wang, “Artificial neural networks for neuroscientists: A primer,” *Neuron*, vol. 109, no. 4, pp. 1049–1070, 2020. https://doi.org/10.1016/j.neuron.2021.01.022
[22] A. M. Zador, “A critique of pure learning and what artificial neural networks can learn from animal brains,” *Nature*, vol. 10, no. 3770, pp. 1–7, 2019. https://doi.org/10.1038/s41467-019-11786-6
[23] X. Tia, K. H. Park, and Q. Liu, “Deep learning influences on higher education students’ digital literacy: The meditating role of higher-order thinking,” *International Journal of Engineering Pedagogy (iJEP)*, vol. 13, no. 6, pp. 33–49, 2023. https://doi.org/10.3991/ijep.v13i6.38177
[24] A. Hassan, H. Pathan, S. S. Kotamjani, S. M. A. Hamza, and R. Rastogi, “Analyzing the deep learning-based mobile environment in educational institutions,” *International Journal of Interactive Mobile Technologies (iJIM)*, vol. 18, no. 9, pp. 155–166, 2024. https://doi.org/10.3991/ijim.v18i09.49029
[25] R. E. Hatimi, C. F. Choukhan, A. Moussaid, and M. Esghir, “Predicting global education quality: A comprehensive machine learning approach using World Bank Data,” *International Journal of Engineering Pedagogy (iJEP)*, vol. 14, no. 4, pp. 24–37, 2024. https://doi.org/10.3991/ijep.v14i4.48205
[26] Y. Lecun and Y. Bengio, “Convolutional networks for images, speech, and time-series,” *Association for Computing Machinery*, vol. 6, pp. 255–258, 1995.
[27] H. Ávila, V. R. Silva, and D. S. F. Magalhães, “Volume estimation of the brain, white matter, and gray matter using FreeSurfer and FSL: Consistency between methods,” *Research Biomedical Engineering*, vol. 35, pp. 257–263, 2019. https://doi.org/10.1007/s42600-019-00027-w

[28] I. A. Araújo, E. F. Amorim, L. M. M. Santana, E. P. Alencar, M. I. A. Araújo, and P. E. Fernandes, "Malformação de Arnold-Chiari: Uma Revisão Bibliográfica," *Journal of Medicine and Health Promotion*, vol. 2, no. 3, pp. 651–660, 2017.

[29] M. C. C. Gozzano, L. S. S. Milare, and M. O. Bernardo, "Glioma de tronco cerebral," *Revista Faculdade de Ciências Médicas de Sorocaba*, vol. 20, no. 3, pp. 187–203, 2018. https://doi.org/10.23925/1984-4840.2018v20i3a13

[30] J. G. Neto, M. G. Tamelini, and O. V. Forlenza, "Diagnóstico diferencial das demências," *Revista Psiquiatria Clínica*, vol. 32, no. 3, pp. 119–130, 2005. https://doi.org/10.1590/S0101-60832005000300004

[31] Radiopaedia.org©, "Recurso de radiologia de edição aberta revisado por pares em rápido crescimento, compilado por radiologistas e outros profissionais de saúde de todo o mundo," *Radiopaedia*, 2024.

[32] Universe.roboflow.com©, "Recurso de radiologia de edição aberta revisado por pares em rápido crescimento, compilado por radiologistas e outros profissionais de saúde de todo o mundo," *Roboflow Universe*, 2025.

[33] F. O. L. Cavalcanti and F. L. Guizardi, "Educação continuada ou permanente em saúde? Análise de Produção Panamericana da Saúde," *Trabalho, Educação e Saúde*, vol. 16, no. 1, pp. 99–122, 2018. https://doi.org/10.1590/1981-7746-sol00119

[34] G. Singh and R. Singh, "Domains of learning: Art of learning in medical education program," *Era's Journal of Medical*, vol. 7, no. 1, pp. 79–85, 2020. https://doi.org/10.24041/ejmr2020.14

[35] C. Bruen *et al*., "Medical student experiences of Case-Based Learning (CBL) at a multicultural medical school," *BMC Medical Education*, vol. 25, no. 152, pp. 1–11, 2025. https://doi.org/10.1186/s12909-024-06585-7

[36] M. Jhala and J. Mathur, "The association between deep learning approach and case based learning," *BMC Medical Education*, vol. 19, no. 106, pp. 1–4, 2019. https://doi.org/10.1186/s12909-019-1516-z

[37] H. Kwon *et al.*, "Deep learning-based automated measurement of cervical length in transvaginal ultrasound images of pregnant women," *IEEE Journal of Biomedical and Health Informatics*, vol. 29, no. 6, pp. 3979–3988, 2025. https://doi.org/10.1109/JBHI.2024.3433594

[38] B. Kim, T. S. Mathai, K. Helm, P. A. Pinto, and R. M. Summers, "Classification of multi-parametric body MRI series using deep learning," *IEEE Journal of Biomedical and Health Informatics*, vol. 28, no. 11, pp. 6791–6802, 2024. https://doi.org/10.1109/JBHI.2024.3448373

[39] J. Bernal *et al.,* "Deep convolutional neural networks for brain image analysis on magnetic resonance imaging: A review," *Artificial Intelligence In Medicine*, vol. 95, pp. 64–81, 2019. https://doi.org/10.1016/j.artmed.2018.08.008

[40] X. Lei and Y. Yang, "Deep learning-based magnetic resonance imaging in diagnosis and treatment of intracranial aneurysm," *Computational and Mathematical Methods in Medicine*, vol. 2022, no. 1, pp. 1–7, 2022. https://doi.org/10.1155/2022/1683475

[41] J. M. Vaz and S. Balaji, "Convolutional neural networks: Concepts and applications in pharmacog," *Molec. Divers*, vol. 25, pp. 1569–1584, 2021. https://doi.org/10.1007/s11030-021-10225-3

[42] W. Voon *et al.,* "Performance analysis of seven Convolutional Neural Networks (CNNs) with transfer learning for Invasive Ductal Carcinoma (IDC) grading in breast histopathological images," *Nature – Scientific Report*, vol. 12, no. 19200, pp. 1–19, 2022. https://doi.org/10.1038/s41598-022-21848-3

[43] D. Alsaeed and S. F. Omar, "Brain MRI analysis for Alzheimer's disease diagnosis using CNN-Based feature extraction and machine learning," *Sensors*, vol. 22, no. 8, pp. 1–16, 2022. https://doi.org/10.3390/s22082911

[44] W. Nie, Z. Yu, L. Mao, A. B. Patel, Y. Zhu, and A. Anandkumar, "Bongard-LOGO: A new benchmark for human-level concept learning and reasoning," in *34th Conference on Neural Information Processing Systems (NeurIPS 2020)*, Vancouver, Canada, vol. 4, 2020, pp. 1–22.

[45] C. F. Cadieu *et al.,* "Deep neural networks rival the representation of primate IT cortex for core visual object recognition," *PLoS – Computational Biology*, vol. 10, no. 12, p. e1003963, 2014. https://doi.org/10.1371/journal.pcbi.1003963

[46] R. M. Battleday, J. C. Peterson, and T. L. Griffiths, "From convolutional neural networks to models of higher-level cognition (and back again)," *Annals of the New York Academy of Science*, vol. 1505, no. 1, pp. 55–78, 2021. https://doi.org/10.1111/nyas.14593

[47] R. Geirhos, D. H. J. Janssen, H. H. Schutt, J. Rauber, M. Bethge, and F. A. Wichmann, "Comparing deep neural networks against humans: Object recognition when the signal gets weaker," *NeurIPS*, vol. 2, pp. 1–31, 2018.

[48] R. Geirhos, K. Meding, and F. A. Wichmann, "Beyond accuracy: Quantifying trial-by-trial behaviour of CNNs and humans by measuring error consistency," *NeurIPS*, vol. 3, pp. 1–26, 2020.

8 AUTHORS

Leonardo Moraes Armesto is an engineer and educator, with a master's degree in Bioengineering and a PhD candidate in Biomedical Engineering at the São Paulo State Institute of Science and Education (2021 and 2025), working as a Full Professor at the Scientific and Technological Institute, Universidade Brasil, working mainly in physical-mathematical fundamentals and tools in medical education. (E-mail: leonardo.armesto@ub.edu.br).

Thabata Roberto Alonso is a pharmacist and educator, with a master's degree in Bioengineering from the Institute of Science and Education of the State of São Paulo, University of Brazil (2024), is a Professor at UniBTA, with interestsin pharmacology and technology for health education.

Daniel Souza Ferreira Magalhães is a Physicist, with a master's and doctorate from the State University of Campinas, Brazil (2003, 2005, and 2009), and postdoctoral degree from the Faculty of Medicine at the State University of Campinas (2013), is a Full Professor at the Scientific and Technological Institute, Universidade Brasil, with interests in medical imaging, medical software development, and bioengineering.

Laurita dos Santos is a Professor at the Scientific and Technological Institute at Universidade Brasil. Her research interests are in biomedical signal and image processing, including nonlinear analysis and complex systems with application in bioengineering systems. (E-mail: laurita.santos@ub.edu.br).