DOI: https://doi.org/10.15276/hait.09.2026.08
UDC 004.9

Comparative evaluation of deep neural networks for brain tumor classification from magnetic resonance imaging

Dmytro I. Uhryn[1)]
ORCID: https://orcid.org/0000-0003-4858-4511; d.ugryn@chnu.edu.ua. Scopus Author ID: 57163746300
Yurii O. Ushenko[1)]
ORCID: https://orcid.org/0000-0003-1767-1882; y.ushenko@chnu.edu.ua. Scopus Author ID: 6701840218
Artem O. Karachevtsev[1)]
ORCID: https://orcid.org/0009-0000-6226-6822; a.karachevtsev@chnu.edu.ua. Scopus Author ID: 36925155800
Yurii O. Halin[1)]
ORCID: https://orcid.org/0009-0006-9629-9896; halin.yurii@chnu.edu.ua
[1)] Yuriy Fedkovych Chernivtsi National University, 2, Kotsiubynskoho Str. Chernivtsi, 58002, Ukraine

ABSTRACT

The growing prevalence of brain tumors and the complexity of magnetic resonance image interpretation create a need for automated decision support tools that can improve diagnostic accuracy and reduce the workload on medical specialists. Modern deep learning approaches demonstrate significant potential in medical image analysis, but the selection of appropriate model architectures requires systematic comparative evaluation considering accuracy, robustness, and computational efficiency. The aim of this study is to evaluate the performance of modern deep learning architectures for multiclass brain tumor classification based on magnetic resonance images and to determine their suitability for practical implementation under different computational conditions. The research is based on the application of transfer learning using several convolutional neural network architectures. The dataset was prepared through cleaning, normalization, and augmentation procedures to improve model generalization and robustness. The models were trained and evaluated using multiple quality criteria, including classification performance, stability under data distortions, and interpretability analysis based on visualization of decision regions. Computational complexity and inference efficiency were also analyzed to assess deployment feasibility. The comparative evaluation demonstrated that deep neural networks are capable of reliably distinguishing between different tumor types and normal brain conditions. One architecture showed the highest overall classification performance and robustness to noisy input data, while architecture provided a more balanced trade off between computational efficiency and prediction quality, making it suitable for resource constrained environments. Visualization analysis confirmed that the models focus on diagnostically relevant regions of the images, supporting the validity of their predictions. The study confirms the effectiveness of deep learning models for automated brain tumor classification and highlights their potential for integration into intelligent medical decision support systems. The obtained results demonstrate practical value for clinical applications, educational purposes, and further research in medical image analysis, particularly in scenarios requiring accurate and fast preliminary diagnosis.

Keywords: Deep learning; magnetic resonance imaging; brain tumor classification; medical image analysis; convolutional neural networks; decision support systems

For citation: **Uhryn D. I., Ushenko Yu. O., Karachevtsev A.O., Halin Yu. O. "Comparative evaluation of deep neural networks for brain tumor clas-sification from magnetic resonance imaging".** ***Herald of Advanced of Information Technology.*** **2026; Vol.9 No.1: 100–113 DOI: https://doi.org/10.15276/hait.09.2026.08**

INTRODUCTION

Brain tumors remain one of the most challenging clinical problems, as timely and accurate diagnosis largely determines the effectiveness of subsequent treatment. According to GLOBOCAN 2022 estimates, hundreds of thousands of new cases of central nervous system (CNS) tumors are registered worldwide each year, with a standardized incidence rate of approximately 3–4 cases per 100,000 population [1]. Similar trends are confirmed by data from the Central Brain Tumor Registry of the United States (CBTRUS, 2017–2021), which indicate a persistently high prevalence of primary CNS tumors in the United States and significant variability in their distribution by age and tumor type [2].

The current World Health Organization classification (WHO CNS5, 2021) emphasizes the key role of molecular diagnostics, which substantially increases the requirements for the accuracy of the initial radiological assessment [3]. Magnetic resonance imaging (MRI) is the primary imaging modality for suspected brain tumors; however, the interpretation of large volumes of images is a complex and resource-intensive process that strongly depends on physician experience. In this context, computer vision and deep learning methods demonstrate significant potential. According to recent systematic reviews, convolutional neural networks (CNNs) are capable of achieving high accuracy, sensitivity, and specificity in brain tumor classification tasks based on MRI images [4], [5], [6]. However, these studies emphasize the need for external validation and algorithm standardization prior to widespread clinical application [7].

An important resource for the development and comparison of algorithms is the international Brain Tumor Segmentation (BraTS) challenge, which in recent years has included tasks not only related to classification but also to missing modality synthesis and pediatric cohort analysis. Owing to large open MRI datasets and transparent evaluation metrics, these challenges provide conditions for reproducibility and accelerate the transfer of technologies into clinical practice [8].

The aim of this paper is to perform a comparative evaluation of modern deep learning models for multiclass brain tumor classification based on MRI images, as well as to develop a web-based application for integrating the obtained results into practical workflows. The study includes domain analysis; data preparation (cleaning, preprocessing, and augmentation); design and training of three CNN architectures: DenseNet121, ResNet50 and EfficientNet-B0; multi-criteria performance evaluation; and integration of the models into a web interface. The choice of architecture is motivated by their proven effectiveness in medical image classification, as reported in recent publications [6], [9], [10].

The expected outcome of the study is the development of a tool that provides high accuracy and fast initial diagnosis, reduces risks associated with the human factor, and can be applied in clinical practice as well as in educational processes and research projects [11], [12], [13], [14], [15].

LITERATURE REVIEW AND PROBLEM STATEMENT

The diagnosis of brain tumors based on MRI images remains one of the key application areas of artificial intelligence methods in medicine. Recent publications show a clear trend toward the use of deep learning approaches, particularly convolutional neural networks (CNNs), which demonstrate high performance in multiclass classification and tumor segmentation tasks [16], [17].

According to the systematic review by Kazerooni et al. (2021), CNN-based models combined with data preprocessing and augmentation techniques can achieve accuracies exceeding 95% in recognizing major types of brain tumors [4]. Similar findings are reported by Pei, Vidyaratne, and Iftekharuddin (2022), who emphasize that ResNet, DenseNet, and EfficientNet architectures are among the most effective for medical applications, as they provide a balance between classification accuracy and computational complexity [5], [18], [19], [20].

Practical studies also confirm the effectiveness of CNNs in diagnostic tasks. For example, Zhou et al. (2020) applied a multi-level CNN for the classification of three tumor types (meningioma, glioma, and pituitary adenoma), achieving an accuracy of over 96% [6]. Shboul et al. (2021) demonstrated that even moderately deep models can successfully classify gliomas, providing a decision support tool for radiologists [9], [21], [22].

An important stage in the advancement of this field is the development and analysis of large open datasets, among which the international Brain Tumor Segmentation Challenge (BraTS) project is of particular importance. In recent editions of the challenge, emphasis has been placed not only on segmentation but also on synthetic data generation and multi-institutional result comparison, which enhances model reproducibility and reliability [8], [23], [24]. This confirms the prospects of using open datasets for training and evaluating algorithms in medical diagnostics.

At the same time, recent reviews (Ranjbarzadeh et al., 2023) highlight several limitations, including the limited availability of annotated data, high hardware requirements for model training, and the lack of standardized evaluation protocols [7]. In addition, Wen et al. (2021) emphasize that clinical deployment of such algorithms must be accompanied by rigorous validation standards, as even minor diagnostic errors may have critical consequences for patients [10].

To summarize current approaches and identify their strengths and limitations, a comparative table (Table 1) was compiled, presenting key results of studies on brain tumor classification based on MRI images [25], [26].

Thus, the literature analysis indicates significant progress in the application of CNNs for brain tumour classification using MRI images. The most promising architectures are considered to be ResNet, DenseNet, and EfficientNet, which demonstrate high accuracy while maintaining relatively moderate computational costs. However, further research is required for widespread clinical implementation, focusing on improving model generalisability, employing multicentre datasets, and integrating solutions into user-friendly applied tools [27].

Table 1. **Distribution of positions (professions) between clusters**

Authors and Year	Model/Architecture	Data Type	Classification Task	Accuracy	Limitations
Zhou et al., 2020 [6]	Multi-level CNN	MRI (T1/T2)	3 classes (meningioma, glioma, pituitary adenoma)	~96%	Limited dataset, need for external validation
Kazerooni et al., 2021 [4]	CNN with data augmentation	MRI (multicenter)	Multiclass	>95%	Lack of standardized evaluation metrics
Shboul et al., 2021 [9]	Medium-depth CNN	MRI (gliomas)	Binary/multiclass	~93%	Limited generalization to other tumor types
Pei et al., 2022 [5]	ResNet, DenseNet, EfficientNet	MRI (BraTS)	Multiclass	94–97%	High computational requirements
Ranjbarzadeh et al., 2023 [7]	Review study	MRI (various datasets)	Generalized analysis	–	Data scarcity, difficulty of standardization highlighted
Bakas et al., 2023 [8]	BraTS Challenge	Large multicenter MRI datasets	Segmentation and classification	Model-dependent	High entry barrier for clinical adoption

***Source*: compiled by the authors**

The conducted literature review shows that the use of deep models, particularly convolutional neural networks, represents the most promising approach for brain tumour classification from MRI images. Recent studies confirm that ResNet, DenseNet, and EfficientNet architectures achieve the highest accuracy rates (94–97%) while maintaining a relatively acceptable level of computational demand [4], [5], [6].

Nevertheless, several challenges hinder practical implementation:

- limited availability of labelled data and the high cost of its preparation;
- issues with model generalisation when transitioning from laboratory settings to clinical scenarios;
- high hardware requirements for training and inference;
- lack of standardised evaluation protocols, complicating comparisons between studies [7–10].

Particular attention should be given to the BraTS project, which provides an international benchmark for model evaluation, although even its results require adaptation to real clinical conditions [8].

In conclusion, despite considerable progress in the field of automated brain tumour classification, tasks related to enhancing reliability, generalisability, and practical usability of models in clinical practice remain open. This underscores both the scientific foundation and practical need for further research on the effectiveness of different deep learning architectures, their comparative evaluation, and subsequent integration into applied tools, including web-based applications for supporting medical decision-making.

RESEARCH AIM AND OBJECTIVES

Timely and accurate diagnosis of brain tumours is one of the key factors influencing treatment efficacy and reducing mortality. Traditional methods of MRI image analysis require considerable effort from radiologists, while the human factor may lead to errors in determining the type of neoplasm. The application of deep machine learning models opens new opportunities for automating tumour classification; however, their effectiveness largely depends on the quality of data preparation and the choice of neural network architecture.

Although deep learning architectures such as ResNet, DenseNet, and EfficientNet have been previously applied to brain tumor classification, this study provides a systematic comparative evaluation of these models using a standardized preprocessing and augmentation pipeline, combined with robustness testing on noisy MRI data. Additionally, computational efficiency and model interpretability are analyzed, demonstrating practical applicability for both high-performance and resource-constrained environments. The integration of these results into a web-based decision support framework highlights the contribution of this work to clinical and educational applications, which has not been previously addressed in the literature.

The aim of this study is to systematically evaluate and compare deep neural network architectures for brain tumor classification using magnetic resonance imaging, considering data preprocessing, model robustness, computational efficiency, and interpretability.

The objectives of this study are:

1) to develop and implement a data preprocessing pipeline, including image normalization, cleaning, class balancing, and augmentation techniques to improve model robustness and generalization;

2) to investigate the impact of dataset limitations and class imbalance on classification performance and to apply appropriate strategies for improving accuracy, particularly for underrepresented tumour types;

3) to perform a comparative evaluation of deep neural network architectures (ResNet50, DenseNet121, EfficientNet-B0) in terms of classification accuracy, recall, computational efficiency, and robustness to noise;

4) to analyze the interpretability of model decisions using visualization techniques (Grad-CAM) in order to assess their suitability for medical diagnostic applications.

MATERIALS AND METHODS

The dataset was split into training, validation, and test subsets in a 70 % / 15 % / 15 % ratio to prevent data leakage. The training subset contained 70 % of images from all classes, the validation subset 15 %, and the test subset 15 %. The split was performed with class stratification, preserving the original class proportions in each subset, and ensuring independence between subsets to allow a correct evaluation of model performance.

The method for preparing brain tumour data from MRI images is described as follows:

Step 1. Feature extraction and selection. At the preprocessing stage, when transforming input data into a feature set for a machine learning model, the primary task is to identify the most informative characteristics that enable effective training. Feature extraction techniques are employed to automatically obtain relevant features without manual design.

In this study, pretrained convolutional neural networks (DenseNet121, ResNet50, and EfficientNet-B0) were applied in a transfer learning mode. This approach allows the reuse of filters already trained on large datasets for extracting low-level features (edges, corners, textures), while fine-tuning is performed only at higher levels. This significantly reduces training time and enhances classification performance, even with limited data availability.

1. ResNet (Residual Networks). The key idea of ResNet lies in the use of residual connections (skip-connections), which overcome the vanishing gradient problem in very deep networks. Formally, instead of learning a direct mapping

$$H(x) \approx F(x).$$

ResNet learns a residual function:

$$H(x) = F(x) + x,$$

where x is the input to the block, $F(x)$ represents a nonlinear transformation (convolutions, normalization, ReLU), and $H(x)$ is the block output.

Thus, the model learns to construct only the “deviation” from the identity mapping, which simplifies optimization. This enables training of networks with hundreds or even thousands of layers without accuracy degradation.

2. DenseNet (Densely Connected Convolutional Networks). DenseNet introduces dense connections (dense blocks), in which the output of each layer is used as input for all subsequent layers. Mathematically, it can be expressed as:

$$x_l = H_l([x_0, x_1, \dots, x_{l-1}]),$$

where x_l is the output of the l-th layer; $x_0, x_1, \dots, x_{l-1}$ represent the concatenation of all previous outputs; and $H_l(\cdot)$ denotes a composition of convolution, normalisation, and activation.

This mechanism:

- ensures feature reuse;
- reduces the number of parameters (as filters do not need to be duplicated);
- improves gradient propagation in deep networks.

3. EfficientNet-B0. EfficientNet optimizes the network architecture using compound scaling, which simultaneously considers three parameters: depth, width, and input image resolution. The scaling formulas are defined as:

$$d = \alpha^{\phi}, \omega = \beta^{\phi}, r = \gamma^{\phi},$$

where d denotes depth (number of layers), ω is width (number of channels), r is the input image resolution, ϕ is the scaling coefficient, and α, β, γ are parameters that determine the scaling proportions.

Thus, the model is scaled in a balanced manner, rather than in a single dimension (e.g., depth alone), allowing high accuracy with a minimal number of parameters.

Neural networks perform automatic feature selection through optimization of the filter weights Wduring training. The convolution operation is formalised as:

$$xy_{i,j,k} = \sum_{m=1}^{M} \sum_{n=1}^{N} \sum_{c=1}^{C} W_{m,n,c,k} \cdot x_{i+m,j+n,c},$$

where x is the input image, W is the convolution kernel, and $y_{i,j,k}$ is the activation at position (i,j) in the k-th output channel.

During training, the weights $W$are adjusted to focus on the most informative image features (e.g., tumour edges or textural patterns). Consequently, ResNet, DenseNet, and EfficientNet represent different strategies for constructing efficient deep models for automatic feature extraction, collectively enabling high classification performance in medical tasks even under limited computational resources.

Step 2. Class balancing. The dataset exhibited a class imbalance issue. To compensate for this imbalance, a class weighting method was applied in the loss function. Each class $i$is assigned a weight ω_i, inversely proportional to the number of samples $N_i$in that class:

$$\omega_i = \frac{N_{\text{total}}}{C \cdot N_i},$$

where N_{total} is the total number of samples in the dataset, and C is the number of classes. Consequently, smaller classes receive higher weights, increasing the penalty for misclassification of these classes.

These weights are then integrated into the CrossEntropyLoss function in PyTorch:

$$L = -\sum_{i=1}^{C} \omega_i \ y_i \log(\hat{y}_i),$$

where y_i is the true class label (one-hot encoded), and $\hat{y}_i$ is the predicted probability for class i. This approach allows the model to train considering class imbalance, thereby improving accuracy on underrepresented categories.

Step 3. Data cleaning and preparation. Corrupted or duplicate images can mislead the neural network and reduce the model's generalization ability. To address this, corrupted files and duplicates were identified and removed during data preparation. Duplicates were detected using MD5 hashing:

$$\text{MD5}(I_i) = \text{MD5}(I_j) \implies I_i = I_j,$$

where I_i, I_j are two images, and $\text{MD5}(\cdot)$ is the hashing function.

Removing duplicates and corrupted samples ensures that the loss function more accurately represents the training data:

$$L = \frac{1}{N} \sum_{i=1}^{N} l(\hat{y}_i, y_i),$$

where $l(\cdot)$ is the local loss function (e.g., CrossEntropy), and $\hat{y}_i$ – is the model prediction for image I_i.

To standardize image dimensions, all images were resized to 224 × 224 pixels, ensuring compatibility with pretrained models and proper functioning of convolutional layers:

$$I' = \text{Resize}(I, H, W), H = W = 224.$$

Following standardization, pixel values were normalized per color channel using the mean μ_k and standard deviation σ_k from ImageNet statistics:

$$\hat{x}'_{i,j,k} = \frac{x_{i,j,k} - \mu_k}{\sigma_k}, k \in \{R, G, B\}.$$

This normalization guarantees that pixel values have a consistent scale ($E\left[\hat{x}'_{i,j,k}\right] = 0$, $\text{Var}[\hat{x}'_{i,j,k}] = 1$), which stabilises gradients and accelerates optimiser convergence:

$$\nabla_\theta L = \frac{1}{N} \sum_{i=1}^{N} \nabla_\theta \ l(\hat{y}_i, y_i).$$

Thus, data cleaning, size standardization, and pixel normalization significantly improve neural network training quality and its generalization capability on unseen data.

Step 4. Data Augmentation. To increase the size and diversity of the dataset, image augmentation techniques were applied. Augmentation generates new variations from existing samples, helping the model become more robust to changes in object position, orientation, and imaging conditions. Formally, an augmented image I'can be represented as a transformation function Tapplied to the original image I:

$$I' = T(I), T \in \{\text{Flipping, Rotation, Noise, Brightness/Contrast}\}.$$

The applied transformations include:

1. Horizontal and vertical flipping:

$$I'_{x,y} = I_{W-x,y}(\text{horizontal}), I'_{x,y} = I_{x,H-y}(\text{vertical}).$$

This ensures that the model is insensitive to object orientation.

2. Rotation:

$$I' = R_\theta(I),$$

where R_θ is a rotation matrix by an angle θ. Small rotations prevent the model from memorising only the “ideal” viewpoint of an object.

3. Gaussian noise:

$$I'_{i,j,k} = I_{i,j,k} + e, e \sim \mathcal{N}(0, \sigma^2).$$

Random noise is added to pixels, enhancing model robustness to noisy real-world data.

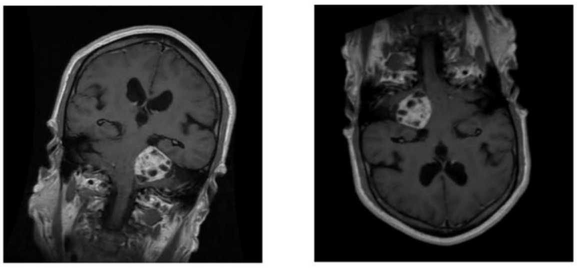
ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

4. Random brightness and contrast adjustment:

$$I'_{i,j,k} = \alpha I_{i,j,k} + \beta,$$

where α is the contrast coefficient and β is the brightness shift. This allows the model to adapt to images captured under varying lighting conditions.

The use of these transformations contributes to the creation of a more diverse and balanced training dataset, enhancing the model's generalization ability and stability when processing new data. Fig. 1 illustrates an example of an original image alongside the results of the applied augmentations.



(a) (b)

Fig. 1. **Visualization of an image before (a) and after (b) augmentation**

***Source*: compiled by the authors**

In this study, the Brain Tumor MRI Dataset was utilized, containing magnetic resonance images of the brain with labelled tumour classes (glioma, meningioma, pituitary) as well as samples without pathology (no tumour). The chosen dataset is hosted on the Kaggle platform and is composed of real MRI images, as indicated by the author [7]. Its use is justified by its open access, substantial data volume, and high-quality labelling, which enable effective training and evaluation of deep learning models. Images are provided in .jpg format, with varying dimensions (mostly 512 × 512 pixels), and were pre-normalised and resized to a standard 224 × 224 pixels before being fed into the models.

The dataset comprises a total of 7,023 images, distributed across the classes as follows:

- Glioma Tumour – 1,621 images;
- Meningioma Tumour – 1,645 images;
- Pituitary Tumour – 1,757 images;
- No Tumour – 2,000 images.

The dataset is publicly available, licensed openly, and contains no personalized information, and therefore is not subject to privacy restrictions. This allows free use for academic and research purposes.

The data meets key quality criteria:

- Accuracy – images correctly correspond to the assigned labels;
- Completeness – all necessary classes are represented;
- Consistency – all images follow a standardized format;
- Relevance – the dataset is not outdated with respect to current medical standards, although it is not updated in real time;
- Validity – clinically significant images are included;
- Uniqueness – each image has a unique identifier;
- Provenance – data were obtained from real MRI studies in medical institutions, as confirmed by the author.

Furthermore, the dataset already included a predefined split into training and testing sets in separate directories. The training set was additionally divided into training and validation subsets in an 80:20 ratio. This allowed not only monitoring training performance during validation but also objectively evaluating model accuracy on an independent test set.

RESEARCH RESULTS

The selection of a machine learning algorithm was based on the characteristics of the MRI image classification task and the model quality requirements outlined in the previous section. To improve accuracy and reduce computational costs, transfer learning was employed: models pretrained on a large image dataset (ImageNet) were adapted to the specific medical data. This approach enabled high performance even with a limited number of training samples.

For brain tumour image classification, three architectures were selected: ResNet50, DenseNet121, and EfficientNet-B0, which have proven effective in medical diagnostics due to their high accuracy and robustness to variations in the data.

When developing machine learning models, it is important to define evaluation metrics in advance. In this study, Accuracy, F1-Score, Recall, Specificity, and AUC-ROC were used to assess classification performance

Recall is particularly critical in medical applications, as it reduces the risk of missing tumour cases, while Specificity helps avoid false diagnoses in healthy patients. The AUC-ROC measures the model's ability to discriminate between classes across different thresholds. Model robustness was assessed by testing on data with noise and distortions. To enhance this property, augmentation methods (flipping, shifts, rotation) were applied, enabling the model to generalise better to varied inputs.

Another important consideration is the architecture complexity and resource requirements, as these determine training speed and the feasibility of practical deployment. A comparison of the three models based on key characteristics is presented in Table 2.

ResNet50 is well-suited for tackling complex tasks; however, it requires substantial computational resources. DenseNet121 is more parameter-efficient but has higher memory demands. EfficientNet-B0 represents a balanced solution under resource-constrained conditions due to its small number of parameters and fast training. The final choice of model depends on the available resources and the specific requirements of the task.

The models were trained over a number of epochs, each comprising a training phase, validation for quality assessment, learning rate adjustment, and early stopping checks. During training, accuracy and validation accuracy (val_accuracy) metrics were monitored to track model performance at each step. Early stopping was employed to prevent overfitting.

Fig. 2 presents the training results of the DenseNet121 model over 105 epochs, showing the dynamics of Loss and Accuracy on both training and validation sets.

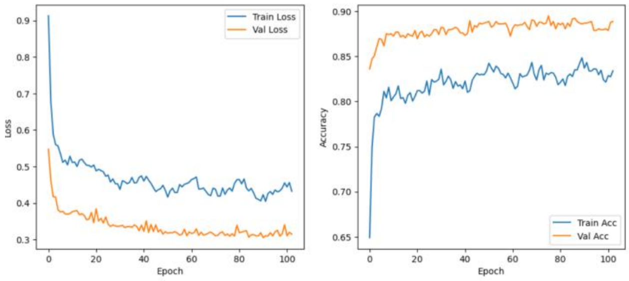


Fig. 2. **Training process of DenseNet121**
***Source*: compiled by the authors**

The model exhibited a stable training process without signs of overfitting: validation accuracy reached approximately 86%, loss gradually decreased, and the differences between the training and validation metrics can be attributed to data augmentation (in particular, the addition of noise to training data). This indicates that the model effectively generalizes information and performs reliably on new images.

Fig. 3 shows the training results of the ResNet50 model over 120 epochs. The model demonstrated high performance, maintaining a balance between training and validation metrics without signs of overfitting, and achieved over 91% accuracy on the validation set.

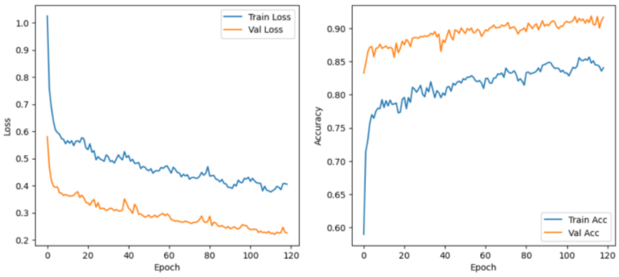


Fig. 3. **Training process of ResNet50**
***Source*: compiled by the authors**

Fig. 4 presents the training graph of the EfficientNet-B0 model over 89 epochs. Due to the absence of improvements during the last 15 epochs, the early stopping mechanism was activated.

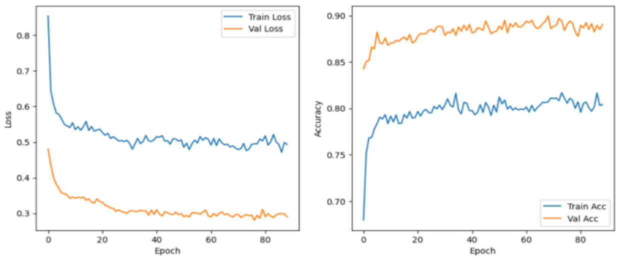


Fig. 4. **Training process of EfficientNet-B0**
***Source*: compiled by the authors**

Table 2. **Comparison of model complexity and training requirements**

Model	Number of parameters	GPU requirements (batch size = 32)	Training time	Features
DenseNet121	~8 million	~3–4 GB	Medium	Architecture with dense connections between layers, improving efficiency and reducing parameter count.
ResNet50	~25.6 million	~4–5 GB	Medium	Deep model with residual connections; high accuracy for complex tasks; high memory and computational demands.
EfficientNet-B0	~5.3 million	~2–2.5 GB	Fast	Optimized architecture using neural network blocks, achieving good efficiency with a small number of parameters.

***Source*: compiled by the authors**

In summary, the ResNet model demonstrated the most stable and balanced performance, although all three architectures proved suitable for the task of brain tumour classification.

For the classification of brain tumours from MRI images, ResNet50 achieved the best overall results. It reached 90% accuracy, the highest F1-scores, and an AUC-ROC of 0.9863. ResNet50 performed particularly well on the pituitary and no tumour classes, achieving high precision and recall values.

The EfficientNet-B0 model achieved slightly lower accuracy but exhibited more balanced recall across all classes. Among the three architectures, it performed best on glioma (recall = 0.83) and meningioma (recall = 0.80). Overall, the most challenging task for all models was classifying glioma and meningioma, likely due to their visual similarity.

DenseNet121 showed lower recall for meningioma (0.71) and glioma (0.79) but performed very well on no tumour (0.96) and pituitary (0.96) samples.

Given the partial class imbalance in the dataset, the F1-score is a key metric, as it accounts for both precision and recall. For the meningioma class, all models produced lower F1-scores (0.73–0.80), which may be due to the smaller representation of this class or its similarity to other tumour types. In contrast, F1-scores for no tumour and pituitary classes exceeded 0.90, indicating high model confidence in these predictions (Fig. 5 – Fig. 7).

The confusion matrices presented in Fig. 8 – Fig. 10 provide a detailed view of classification performance across all tumour categories for the DenseNet121, ResNet50, and EfficientNet-B0 models.

ResNet50 achieved the highest accuracy and the most balanced performance, while EfficientNet-B0 demonstrated nearly comparable results.

When tested on noisy MRI images, ResNet50 showed the greatest robustness, maintaining almost full accuracy, making it the most reliable option for practical applications under non-ideal or noisy conditions. EfficientNet-B0 also demonstrated good stability, particularly in detecting glioma and no tumour, whereas DenseNet121 was the most sensitive to noise .

In deep learning systems, the “black box” problem often arises, where a model achieves high accuracy but the decision-making process is unclear. To enhance trust and interpretability in image classification, the Grad-CAM method (Gradients-weighted Class Activation Mapping) was applied. This technique generates a heatmap highlighting important regions of the image that influenced the model’s decision [11]. It allows clinicians to assess whether the model focuses on relevant areas and to identify potential errors.

DenseNet121

	precision	recall	f1-score	support
glioma	0.92	0.79	0.85	299
meningioma	0.76	0.71	0.73	304
notumor	0.91	0.96	0.94	381
pituitary	0.85	0.96	0.90	300
accuracy			0.86	1284
macro avg	0.86	0.86	0.86	1284
weighted avg	0.86	0.86	0.86	1284

AUC-ROC: 0.9740

Fig. 5. **Classification report of the model DenseNet121**

***Source*: compiled by the authors**

ResNet50

	precision	recall	f1-score	support
glioma	0.96	0.81	0.88	299
meningioma	0.79	0.81	0.80	304
notumor	0.92	0.98	0.95	381
pituitary	0.94	0.98	0.96	300
accuracy			0.90	1284
macro avg	0.90	0.90	0.90	1284
weighted avg	0.90	0.90	0.90	1284

AUC-ROC: 0.9863

Fig. 6. **Classification report of the model ResNet50**

***Source*: compiled by the authors**

EfficientNet-B0

	precision	recall	f1-score	support
glioma	0.96	0.83	0.89	299
meningioma	0.80	0.80	0.80	304
notumor	0.92	0.97	0.95	381
pituitary	0.90	0.96	0.93	300
accuracy			0.89	1284
macro avg	0.89	0.89	0.89	1284
weighted avg	0.90	0.89	0.89	1284

AUC-ROC: 0.9778

Fig. 7. **Classification report of the model EfficientNet-B0**

***Source*: compiled by the authors**

Fig. 12 presents Grad-CAM examples for four MRI images. ResNet50 demonstrated the best interpretability, with highlighted regions more accurately aligning with pathological structures. EfficientNet-B0 and DenseNet121 also correctly classified the tumours, but their heatmaps were more diffuse and occasionally misaligned, particularly for pituitary tumours. This may be because the models consider not only the tumour itself but also secondary brain changes, such as deformations or inflammatory responses, which also hold diagnostic significance.

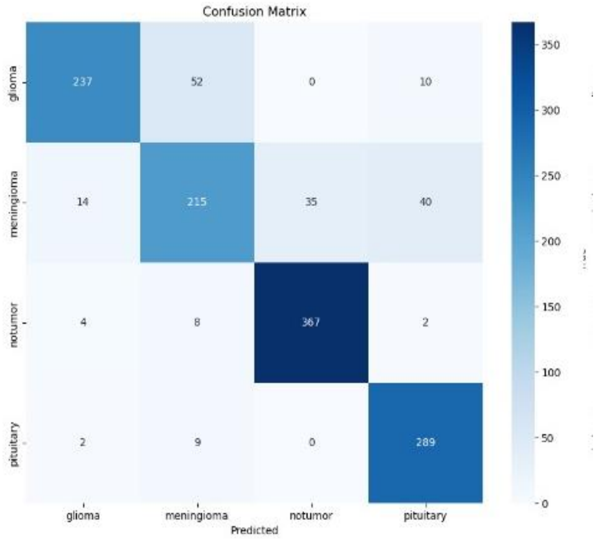


Fig. 8. **Confusion matrix of the model DenseNet121**
***Source*: compiled by the authors**

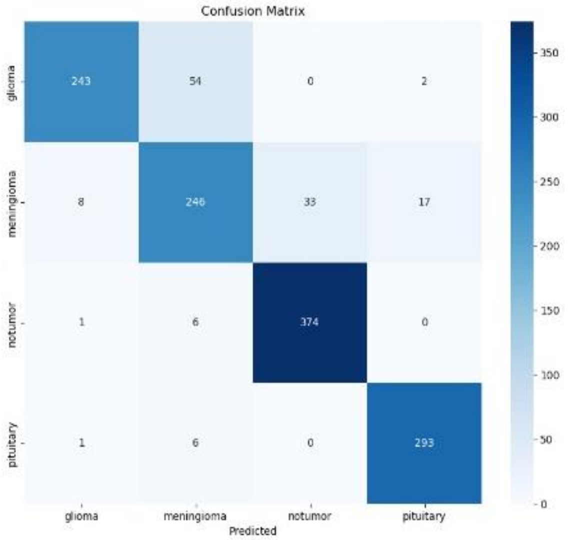


Fig. 9. **Confusion matrix of the model ResNet50**
***Source*: compiled by the authors**

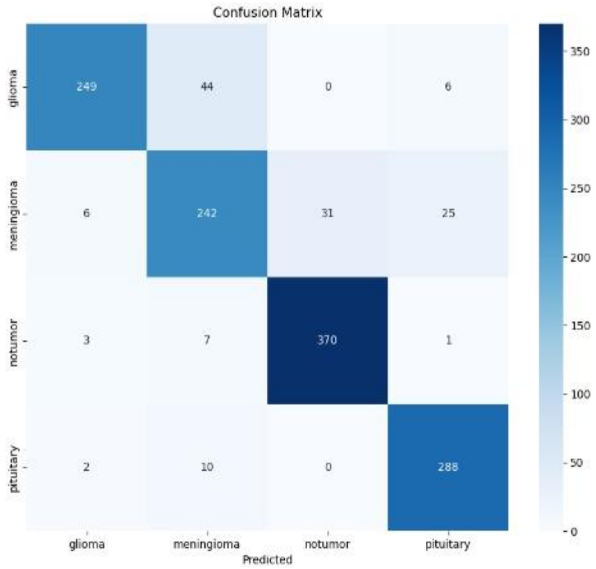


Fig. 10. **Confusion matrix of the model EfficientNet-B0**
***Source*: compiled by the authors**

The primary criterion for deploying a model in a production environment is its compliance with established quality metrics, including accuracy, recall, F1-score, and domain-specific requirements. In medical image classification tasks, particularly for brain tumour detection, recall is prioritised to minimise the likelihood of missing a pathology. Comparison of the trained models showed that ResNet50 and EfficientNet-B0 achieved the best performance in terms of accuracy and recall. ResNet50 also demonstrated high robustness to noise and good interpretability, enhancing trust in the model.

In addition to quality metrics, inference speed and computational resource requirements should be considered when selecting a model. Table 3 presents a comparison of the models according to these parameters.

DenseNet121

	precision	recall	f1-score	support
glioma	0.74	0.85	0.79	299
meningioma	0.76	0.43	0.55	304
notumor	0.97	0.79	0.87	381
pituitary	0.63	0.97	0.77	300
accuracy			0.76	1284
macro avg	0.78	0.76	0.74	1284
weighted avg	0.79	0.76	0.75	1284

ResNet50

	precision	recall	f1-score	support
glioma	0.96	0.69	0.80	299
meningioma	0.71	0.80	0.76	304
notumor	0.94	0.94	0.94	381
pituitary	0.84	0.98	0.91	300
accuracy			0.86	1284
macro avg	0.86	0.85	0.85	1284
weighted avg	0.87	0.86	0.85	1284

EfficientNet-B0

	precision	recall	f1-score	support
glioma	0.90	0.73	0.81	299
meningioma	0.80	0.54	0.64	304
notumor	0.86	0.96	0.91	381
pituitary	0.71	0.98	0.82	300
accuracy			0.81	1284
macro avg	0.82	0.80	0.80	1284
weighted avg	0.82	0.81	0.80	1284

Fig. 11. **Classification report of the models on noisy data**
***Source*: compiled by the authors**

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

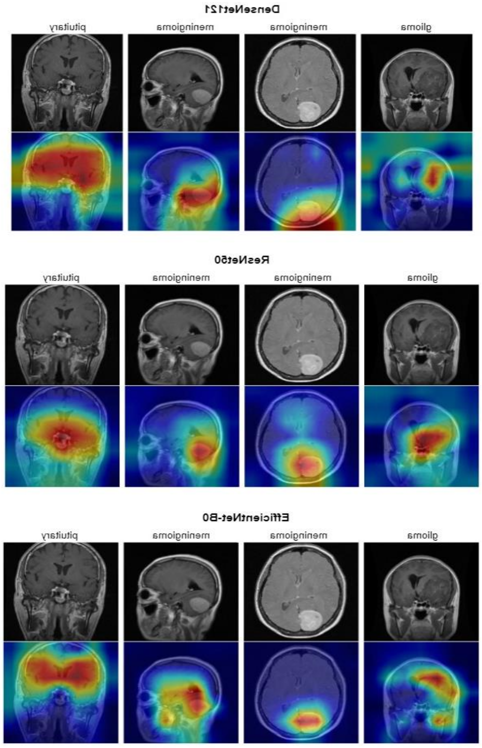


Fig. 12. **Grad-CAM visualisation of MRI images with tumour type classification**
***Source*: compiled by the authors**

Table 3. **Comparison of models for deployment**

Model	Number of parameters	FLOPs (224×224)	RAM During Inference
DenseNet121	~8.0 million	~2.9 GFLOPs	Medium (~60–120 MB)
ResNet50	~25.6 million	~4.1 GFLOPs	High (~100–200 MB)
EfficientNet-B0	~5.3 million	~0.39 GFLOPs	Low (~30–60 MB)

***Source*: compiled by the authors**

ResNet50 has a large number of parameters and high computational requirements, which ensure high accuracy but increase inference time. In cases of limited resources or the need for fast processing, EfficientNet-B0 is the optimal choice. If the priority is maximum accuracy and robustness and resources allow, ResNet50 is recommended.

The models are intended for use as assistive tools for preliminary diagnosis, which can be presented as a prediction panel. Figures 13, 14, and 15 illustrate how the different models classify the same tumour image. While all models correctly identified the class, the confidence levels varied.

Thus, for the given example, the predicted probability was 0.74 for DenseNet121, 0.92 for ResNet50, and 0.97 for EfficientNet-B0.

CONCLUSIONS

This study aimed to systematically evaluate the effectiveness of three deep learning architectures — ResNet50, DenseNet121, and EfficientNet-B0 — for automated brain tumor classification from MRI images. The research addressed key objectives including data preprocessing, class balancing, model training, performance evaluation, and interpretability analysis.

1. ResNet50 demonstrated the highest overall accuracy (90–91%), the best F1-score and AUC-ROC (0.9863), and the greatest robustness to noisy MRI data, making it the most reliable option for clinical applications requiring high precision. Grad-CAM visualizations confirmed that the model focused on relevant tumor regions, enhancing interpretability and trust.

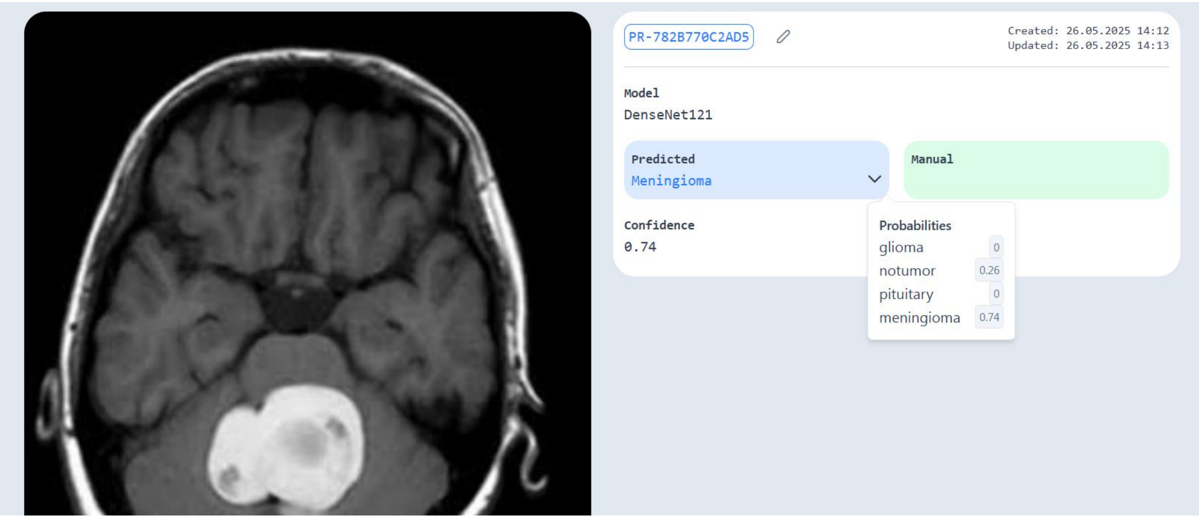


Fig. 13. **Prediction of the DenseNet121 model**
***Source*: compiled by the authors**

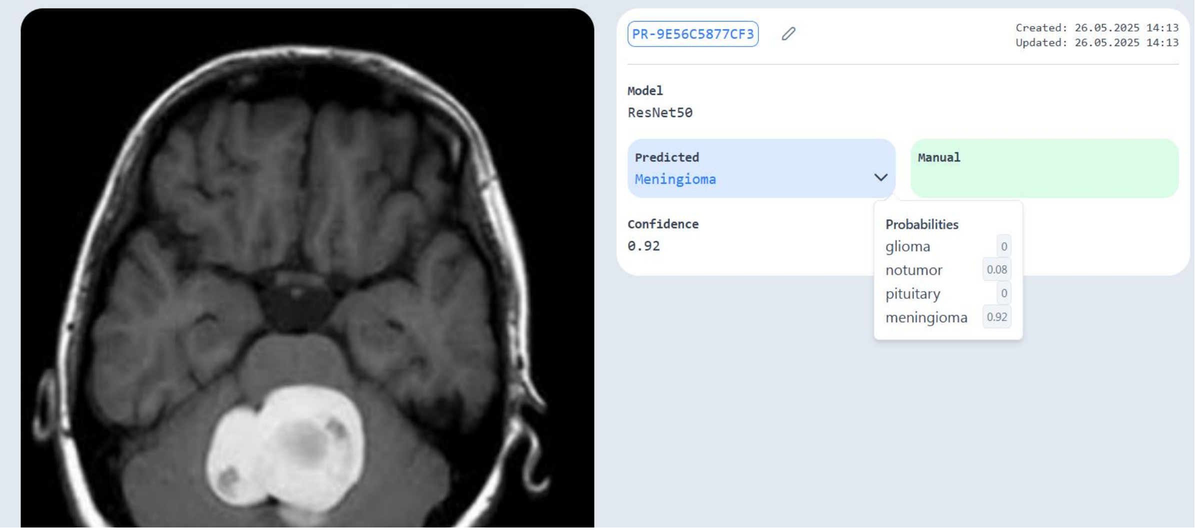


Fig. 14. **Prediction of the ResNet50 model**
***Source*: compiled by the authors**

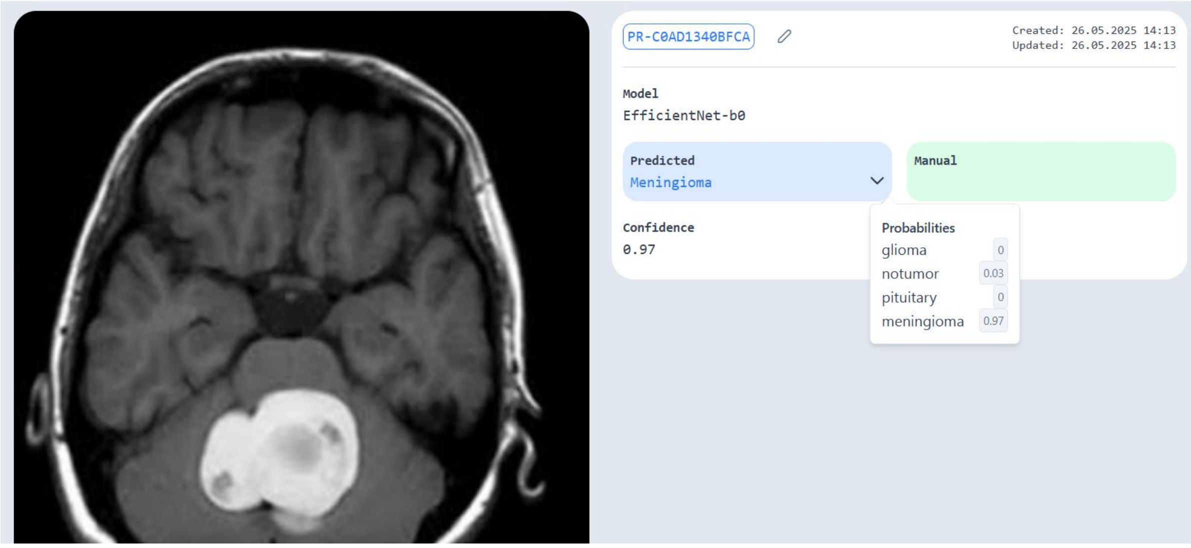


Fig. 15. **Prediction of the EfficientNet-B0 model**
***Source*: compiled by the authors**

2. EfficientNet-B0 showed slightly lower accuracy but balanced recall across all tumor classes, combined with significantly lower computational requirements. This makes it particularly suitable for deployment in resource-constrained environments or scenarios requiring fast inference.

3. DenseNet121 was effective in classifying pituitary tumors and healthy cases but exhibited lower recall for glioma and meningioma and higher sensitivity to noise. Its Grad-CAM visualizations were less informative, limiting its practical utility in clinical diagnostics.

This work provides a systematic comparative evaluation of widely used CNN architectures using standardized preprocessing, augmentation, and robustness testing, combined with analysis of computational efficiency and model interpretability. Integration of these results into a web-based decision support framework highlights the practical applicability for both clinical and educational settings, which has not been fully addressed in previous studies.

Further work could focus on multi-center datasets to improve model generalization, integration of multimodal imaging data, and development of adaptive lightweight architectures for real-time clinical deployment.

USE OF ARTIFICIAL INTELLIGENCE

Used for the correct formulation of some sentences and translation into English.

ACKNOWLEDGMENTS

This research has been funded by the National Research Fund of Ukraine (Grant No. 2023.03/0174**).**

REFERENCES

1. Bray, F., Laversanne, M., Sung, H., et al. “Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries”. *CA: A Cancer Journal for Clinicians*. 2024; 74 (3): 229–263, 2024. DOI: https://doi.org/10.3322/caac.21834.

2. “American Brain Tumor Association”. *Brain Tumor Statistics: CBTRUS Statistical Report 2017–2021*. 2022.

3. Louis, D. N., Perry, A., Wesseling, P., et al. “The 2021 WHO classification of tumors of the central nervous system: a summary”. *Neuro-Oncology*. 2021; 23 (8): 1231–1251. DOI: https://doi.org/10.1093/neuonc/noab106.

4. Hamza, A. & Damaševičius, R. “Deep learning for brain tumor segmentation and classification: A systematic review of methods and trends”. *Computers, Materials & Continua*. 2025; 86 (1): 1–41. DOI: https://doi.org/10.32604/cmc.2025.069721.

5. Zarenia, E., Far, A. A. & Rezaee, K. “Automated multi-class MRI brain tumor classification and segmentation using deformable attention and saliency mapping”. *Scientific Reports*. 2025; 15: 8114. https://doi.org/10.1038/s41598-025-92776-1.

6. Khan, H. A., Jue, W., Mushtaq, M. & Mushtaq, M. U. “Brain tumor classification in MRI image using convolutional neural network”. *Mathematical Biosciences and Engineering*. 2020; 17 (5): 6203–6216. DOI: https://doi.org/10.3934/mbe.2020328.

7. Saluja, S., Munesh, T., & Saha, A. “Deep CNNs for glioma grading on conventional MRIs: Performance analysis, challenges, and future directions”. *Mathematical Biosciences and Engineering*. 2024; 21 (4): 5250–5282, DOI: https://doi.org/10.3934/mbe.2024232.

8. Zhang, Y., Han, Y. & Zhang, J. “MAU-Net: Mixed attention U-Net for MRI brain tumor segmentation”. *Mathematical Biosciences and Engineering*. 2023; 20 (12): 20510–20527. DOI: https://doi.org/10.3934/mbe.2023907.

9. Özcan, H., Emiroğlu, B. G., Sabuncuoğlu, H., Özdoğan, S., Soyer, A. & Saygi, T. “A comparative study for glioma classification using deep convolutional neural networks”. *Math Biosci Eng*. 2021; 18 (2): 1550–1572. DOI: https://doi.org/10.3934/mbe.2021080.

10. Wen, P. Y., Weller, M., Lee, E. Q. et al. “Glioblastoma in adults: A Society for Neuro-Oncology (SNO) and European Society of Neuro-Oncology (EANO) consensus review on current management and future directions”. *Neuro-Oncology*. 2021; 22 (8): 1073–1113. DOI: https://doi.org/10.1093/neuonc/noaa106.

11. Uhryn, D., Ushenko, Y., Kovalchuk, M. & Bilobrytskyi, D. “Modelling a system for intelligent forecasting of trading on stock exchanges”. *Security of infocommunication systems and internet of things (SISIOT)*. 2023; 1 (2): 02002. DOI: https://doi.org/10.31861/sisiot2023.2.02002.

12. Levkivskyi, D., Vysotska, V., Chyrun, L., Ushenko, Y., Uhryn, D. & Hu, C. “Agile methodology of information engineering for semantic annotations categorization and creation in scientific articles based on NLP and machine learning methods”. *International Journal of Information Engineering and Electronic Business (IJIEEB)*. 2025; 17 (2): 1–50. DOI: https://doi.org/10.5815/ijieeb.2025.02.01.

13. Uhryn, D. I., Karachevtsev, A. O., Shevchuk, S. F. & Uhryn, A. D. “Modeling and forecasting of stock market processes”. *Herald of Advanced Information Technology*. 2024; 7 (1): 85–98. DOI: https://doi.org/10.15276/hait.07.2024.7.

14. Uhryn, D. I., Karachevtsev, A. O., Ilin, V. A., Halin, Y. O. & Shkidina, K.S. “Investigation of the efficiency of neural network models for developing a classifier of ophthalmic pathologies”. *Applied Aspects of Information Technology*. 2025; 8 (1): 102–112. DOI: https://doi.org/10.15276/aait.08.2025.8.

15. Uhryn, D., Ushenko, Y. & Iliuk, O. “Application of a genetic algorithm to solve a nonlinear problem of optimizing the diagnosis of unstable angina.” In *Lecture Notes on Data Engineering and Communications Technologies*, 2025; 194: 284–293. DOI: https://doi.org/10.1007/978-3-031-84228-3_24.

16. Burov, Y., Vysotska, V., Chyrun, L., Ushenko, Y., Uhryn, D. & Hu, Z. “Intelligent network architecture development for E-Business processes based on ontological models”. *International Journal of Information Engineering and Electronic Business*. 2024; 16 (5): 1–54. DOI: https://doi.org/10.5815/ijieeb.2024.05.01.

17. Mediakov, O., Vysotska, V., Uhryn, D., Ushenko, Y. & Hu, C. “Information technology for generating lyrics for song extensions based on transformers”. *International Journal of Modern Education and Computer Science*. 2024; 16 (1): 23–36. DOI: https://doi.org/10.5815/ijmecs.2024.01.03.

18. Isong, A., Stephen, B. U.-A., Asuquo, P., Ihemereze, C. & Enang, I. “Machine learning based cloud computing intrusion detection”. *Advanced Information Systems*. 2026. 10 (1): 115–125. DOI: https://doi.org/10.20998/2522-9052.2026.1.13.

19. Ajit, A., Acharya, K. & Semanta, A. “A review of convolutional neural networks”. *International Conference on Emerging Trends in Information Technology and Engineering*. 2020. p. 1–5. DOI: https://doi.org/10.1109/ic-ETITE47903.2020.049.

20. Youssef, A., Nichol, A. A., Martinez-Martin, N. et al. “Ethical considerations in the design and conduct of clinical trials of Artificial Intelligence”. *JAMA Netw Open*. 2024; 7 (9): e2432482. DOI: https://doi.org/doi:10.1001/jamanetworkopen.2024.32482.

21. MacIntyre, M. R., Cockerill, R. G., Mirza, O. F. & Appel, J. M. “Ethical considerations for the use of artificial intelligence in medical decision-making capacity assessments”. *Psychiatry Research*. 2023;328: 115466, https://www.scopus.com/authid/detail.uri?authorId=57210131886.
DOI: https://doi.org/10.1016/j.psychres.2023.115466.

22. Rajkomar, A., Dean, J. & Kohane, I. “Machine learning in medicine”. *N Engl J Med*. 2019; 380 (14): 1347–1358. DOI: https://doi.org/10.1056/NEJMra1814259.

23. Larson ,D. B., Magnus, D. C., Lungren, M. P., Shah, N. H., Langlotz, C. P. “Ethics of using and sharing clinical imaging data for artificial intelligence: a proposed framework|. *Radiology*, 2020; 295 (3): 675-682. DOI: https://doi.org/10.1148/radiol.2020192536.

24. Suri, A. & Summers, R. M. “Privacy, please: safeguarding medical data in Imaging AI using differential privacy techniques”. *Radiology: Artificial Intelligence*. 2024; 6: 1. DOI: https://doi.org/10.1148/ryai.230560.

25. Hidayat, E. Y., Hastuti, K. & Muda, A. K. “Artificial intelligence in digital image processing: A bibliometric analysis”. *Intelligent Systems with Applications*. 2025; 25: 200466. DOI: https://doi.org/10.1016/j.iswa.2024.200466.

26. Mittal, M., Arora, M., Pandey, T. & Goyal, L. M. “Image segmentation using deep learning techniques in medical images”. *Advancement of machine intelligence in interactive medical image analysis*. 2020. p. 41–63. DOI: https://doi.org/10.1007/978-981-15-1100-4_3.

27. Kushnerov, O., Niesheva, A., Shcherbachenko, V., Sokol, V. & Amiraslanov, K. “Neural network modeling and forecasting of imbalances in ukraine’s labor market under extreme conditions”. *Advanced Information Systems*. 2026; 10 (1): 94–105. DOI: https://doi.org/10.20998/2522-9052.2026.1.11.

Conflicts of Interest: The authors declare that they have no conflict of interest regarding this study, including financial, personal, authorship, or other interests, which could influence the research and its results presented in this article.

Received 26.12.2025
Received after revision 12.02.2026
Accepted 19.02.2026

DOI: https://doi.org/10.15276/hait.09.2026.08
УДК 004.9

Порівняльне оцінювання глибоких нейронних мереж для класифікації пухлин мозку за зображеннями магнітно-резонансной томографії

Угрин Дмитро Ілліч[1)]
ORCID: https://orcid.org/0000-0003-4858-4511; d.ugryn@chnu.edu.ua. Scopus Author ID: 57163746300
Ушенко Юрій Олександрович[1)]
ORCID: https://orcid.org/ 0000-0003-1767-1882; y.ushenko@chnu.edu.ua. Scopus Author ID: 6701840218
Карачевцев Артем Олегович[1)]
ORCID: https://orcid.org/0009-0000-6226-6822; a.karachevtsev@chnu.edu.ua. Scopus Author ID: 36925155800
Галін Юрій Олександрович[1)]
ORCID: https://orcid.org/0009-0006-9629-9896; halin.yurii@chnu.edu.ua
[1)] Чернівецький національний університет ім. Ю. Федьковича, Коцюбинського, 2. Чернівці, 58002, Україна

АНОТАЦІЯ

Зростання поширеності пухлин головного мозку та складність інтерпретації магнітно резонансних зображень зумовлюють потребу у створенні автоматизованих систем підтримки прийняття рішень, здатних підвищити точність діагностики та зменшити навантаження на медичних фахівців. Сучасні методи глибокого навчання демонструють значний потенціал у задачах аналізу медичних зображень, однак вибір оптимальної архітектури моделей потребує порівняльного оцінювання з урахуванням точності, стійкості та обчислювальної ефективності. Метою дослідження є оцінювання ефективності сучасних архітектур глибокого навчання для багатокласової класифікації пухлин головного мозку за магнітно резонансними зображеннями та визначення їх придатності до практичного використання за різних обчислювальних умов. Дослідження базується на використанні перенавчання згорткових нейронних мереж. Набір даних було підготовлено із застосуванням процедур очищення, нормалізації та аугментації для підвищення здатності моделей до узагальнення та стійкості до змін вхідних даних. Оцінювання моделей здійснювалося за комплексом критеріїв якості класифікації, стійкості до спотворень даних та інтерпретованості результатів на основі візуалізації областей прийняття рішень. Додатково проаналізовано обчислювальну складність та ефективність роботи моделей для оцінювання можливості практичного впровадження. Порівняльне дослідження показало, що глибокі нейронні мережі здатні надійно розрізняти різні типи пухлин та нормальний стан мозку. Одна з архітектур продемонструвала найвищу точність класифікації та стійкість до зашумлених даних, тоді як інша забезпечила збалансоване поєднання обчислювальної ефективності та якості прогнозування, що є важливим для використання в умовах обмежених ресурсів. Візуалізаційний аналіз підтвердив, що моделі фокусуються на діагностично значущих ділянках зображень, що підвищує довіру до отриманих результатів. Проведене дослідження підтверджує ефективність моделей глибокого навчання для автоматизованої класифікації пухлин головного мозку та їх перспективність для інтеграції у інтелектуальні системи підтримки медичних рішень. Отримані результати мають практичну цінність для клінічного застосування, освітніх задач та подальших досліджень у галузі аналізу медичних зображень, особливо у випадках, де необхідна швидка та точна попередня діагностика.

Ключові слова: глибоке навчання; магнітно резонансна томографія; класифікація пухлин мозку; аналіз медичних зображень; згорткові нейронні мережі; системи підтримки прийняття рішень

ABOUT THE AUTHORS

Dmytro I. Uhryn - Doctor of Engineering Sciences, Professor, Computer Science Department. Yuriy Fedkovych Chernivtsi National University, 2, Kotsyubynsky Str. Chernivtsi, 58002, Ukraine.
ORCID: https://orcid.org/0000-0003-4858-4511; d.ugryn@chnu.edu.ua. Scopus Author ID: 57163746300
***Research field*:** Information technologies for decision support; swarm intelligence systems; branch geoinformation systems

Угрин Дмитро Ілліч - доктор технічних наук, професор, доцент кафедри Комп’ютерних наук. Чернівецький національний університет ім. Ю. Федьковича, Коцюбинського, 2. Чернівці, 58002, Україна

Yurii O. Ushenko - Doctor of Physical-Mathematical Sciences, Head of Computer Science Department. Yuriy Fedkovych Chernivtsi National University, 2, Kotsyubynsky Str. Chernivtsi, 58002, Ukraine.
ORCID: https://orcid.org/0000-0003-1767-1882; y.ushenko@chnu.edu.ua. Scopus Author ID: 6701840218
Research field: Information technologies and intelligence systems; digital signals and images processing

Ушенко Юрій Олександрович - доктор фізико-математичних наук, завідувач кафедри Комп’ютерних наук. Чернівецький національний університет ім. Ю. Федьковича, Коцюбинського, 2. Чернівці, 58002, Україна

Artem O. Karachevtsev - PhD, Assistant Professor, Computer Science Department, Yuriy Fedkovych Chernivtsi National University, 2, otsyubynsky Str, Chernivtsi, 58002, Ukraine.
ORCID: https://orcid.org/0009-0000-6226-6822; a.karachevtsev@chnu.edu.ua. Scopus Author ID: 36925155800
***Research field*:** Computer Science, Software Engineering, Web Development, Biomedical Optics

Карачевцев Артем Олегович - кандидат фізико-математичних наук, асистент кафедри Комп’ютерних наук, Чернівецький національний університет ім. Ю. Федьковича, Коцюбинського, 2, 58002, м. Чернівці, Україна

Yurii O. Halin - PhD student, Computer Science Department. Yuriy Fedkovych Chernivtsi National University, 2, Kotsyubynsky Str. Chernivtsi, 58002,Ukraine
ORCID: https://orcid.org/0009-0006-9629-9896; halin.yurii@chnu.edu.ua
Research field: Information technologies and intelligence systems

Галін Юрій Олександрович - аспірант кафедри Комп’ютерних наук. Чернівецький національний університет ім. Ю. Федьковича, Коцюбинського, 2. Чернівці, 58002,Україна