applied sciences

Article

Integration of Multi-Level Wavelet Decomposition and CNN for Brain Tumor MRI Classification

Mahammad Ismayilov * and Dalia Čalnerytė

Department of Applied Informatics, Kaunas University of Technology, Studentų St. 50,
LT-51368 Kaunas, Lithuania; dalia.calneryte@ktu.lt
* Correspondence: mahammad.ismayilov@ktu.lt

Abstract

Magnetic resonance imaging (MRI) remains one of the most important tests for diagnosing and monitoring various diseases. In recent years, machine learning methods have been widely applied to automate MRI analysis. It supports decision-making by predicting disease and highlighting relevant regions. However, the proper use of feature extraction methods can improve the performance of the model. This paper proposes a WaveletFusion architecture that combines a two-dimensional Haar wavelet decomposition with a convolutional neural network (CNN) for classification. The approach was demonstrated on the Brain Tumor MRI dataset and further examined on the Br35H :: Brain Tumor Detection 2020 (Br35H). The model decomposes each MRI slice into approximation and directional detail subbands and fuses multi-scale wavelet features within the convolutional pipeline. To evaluate the effect of decomposition depth, WaveletFusion variants from one to eight levels were compared with a Baseline CNN model under the same training protocol. The results showed that performance improved progressively with increasing decomposition depth up to level 7, whereas the 8-level configuration consistently declined, indicating that excessive decomposition introduces information loss and over-compression in the deepest approximation pathway. The best-performing configuration, which outperformed both the Baseline CNN and the WaveletFusion variations in five independent runs, was the 7-level WaveletFusion model, achieving a test accuracy of 0.94 ± 0.01 and test macro-F1 of 0.93 ± 0.02. A similar tendency was observed on the Br35H dataset, where the 7-level model achieved a 0.97 ± 0.01 test accuracy and 0.97 ± 0.01 test macro-F1, while the 8-level configuration remained weaker on both datasets. These results show that multi-scale wavelet fusion can improve Brain Tumor MRI classification while maintaining a compact model size and a fair comparison setting, and that the decomposition depth must be selected carefully.

Keywords: convolutional neural network (CNN); Haar wavelet decomposition; image classification; multi-level wavelet fusion; explainable artificial intelligence (XAI)

Academic Editor: Vladislav Toronov

Received: 23 March 2026
Revised: 24 April 2026
Accepted: 30 April 2026
Published: 2 May 2026

1. Introduction

In medical image analysis, deep learning has become the leading approach due to the ability of neural network models to learn automatically hierarchical visualizations from image data, improving classification and segmentation performance [1–3]. In neurology, the integration of convolutional neural networks has demonstrated capability in identifying complex textural patterns within magnetic resonance imaging scans [4,5].

https://doi.org/10.3390/app16094482

However, several surveys have emphasized that deep learning models have unique challenges in magnetic resonance imaging (MRI), especially in brain tumor classification, such as limited annotated datasets, inter-scanner variability, and high intra-class variability among tumor types [3,6].

Brain tumor classification using MRI has been extensively explored with different CNN architectures trained either from scratch or through the machine learning technique called transfer learning [7]. Early studies demonstrated that conventional CNNs could achieve high classification accuracy for brain tumor types when applied to curated MRI datasets [8,9]. Following research included deeper deep learning models such as ResNet, DenseNet, and VGG-based models to increase feature representation capacity while mitigating vanishing gradient issues [10,11]. Additionally, transfer learning has become popular in medical image analysis due to limited datasets, which can pose challenges for the models. To adapt to MRI classification, transfer learning enables models pretrained on large natural image datasets with enhanced convergence stability and performance [12,13]. Despite the promising accuracy results, scientists have raised concerns about the dataset splitting strategies and potential data leakage [14]. The strict methodological reporting frameworks, including CLAIM and STARD-AI, also mention the importance of transparent data splitting and validation strategies to ensure trustworthy clinical AI development [15,16]. CNNs are a powerful tool for brain tumor classification because they can automatically learn discriminative features directly from MRI images, capturing complex spatial patterns such as tumor shape, texture, and intensity variations. This reduces reliance on handcrafted features and enables end-to-end learning with strong performance. However, they primarily operate in the spatial domain and may overlook subtle frequency-based or multi-scale information that is clinically important. In addition, CNNs are prone to overfitting, especially when trained on limited or imbalanced medical datasets, which can reduce generalization to new patients, scanners, or clinical settings [17].

Besides conventional CNN architectures, integration of classical signal processing methods with deep learning models has gained increasing attention nowadays [18]. For example, the discrete wavelet transform provides multi-resolution analysis by decomposing images into subbands representing different frequency components, enabling separation of coarse structural information from fine-grained textural details [19]. Several studies have applied wavelet-based preprocessing before CNN classification to reduce dimensionality while preserving diagnostically relevant features [20,21]. As an example, Haar wavelet decomposition has also been integrated due to its lower computational cost and suitability in such classification tasks [21]. The integration of discrete wavelet transforms with neural networks was applied to detect hepatocellular carcinoma from abdominal ultrasound images. This approach used a coarse-to-fine lesion detection strategy with wavelet-based pattern-augmented lesion images to enhance morphological feature discrimination [22]. The wavelet transformations were used to denoise, compress, and extract features of MRI in brain tumor classification task to extract texture and edge features efficiently [23].

To combine the strengths of frequency-domain, hybrid wavelet-CNN models have been proposed [24]. Wavelet coefficients are used as additional input channels to several CNN architectures, enabling the network to learn from both raw pixel intensities and frequency components simultaneously [25] as well as to better capture multi-scale texture and structural features [26]. Moreover, other studies integrate wavelet pooling layers into CNN architectures, while preserving spectral information and reducing loss during downsampling [27,28], demonstrating that it can retain high-frequency tumor boundary information more effectively [29]. Furthermore, multi-level wavelet decompositions have been introduced to directly model hierarchical frequency representations within deep neural networks, thereby increasing robustness to noise [30]. 2D-wavelet decomposition results together with texture features were used as engineered features in parallel to auto-learned features extracted using ResNet50. The fusion of these features resulted in capturing both local and global characteristics and ensuring the use of features that can be missed by automatic feature extraction methods [31]. A multi-stage wavelet-attention deep learning framework for brain MRI tumor classification built on a ResNet-50 transfer-learning backbone adapted for MRI images addressed class imbalance and demonstrated improved generalization [32]. Besides brain MRI classification, prior studies have demonstrated that incorporating wavelet-domain information into CNN pipelines can enrich multi-scale feature learning by combining spatial intensity patterns with localized frequency cues, which has been associated with improved classification performance in related biomedical image-analysis settings [33,34]. By integrating wavelet transforms, CNNs explicitly capture multi-scale frequency information such as edges, textures, and tumor boundaries, which helps them better represent subtle structural variations and typically leading to improved discrimination and robustness. However, hybrid models are more complex, requiring additional preprocessing, careful hyper-parameter tuning, and higher computational cost [26].

Explainable artificial intelligence techniques, such as Grad-CAM and class visualization, have become a critical component of clinical imaging research to increase trust in medicine and enhance transparency. These techniques have also been widely applied to brain tumor classification to highlight regions that contribute to model predictions [35]. Nevertheless, recent studies emphasize that explainability must extend beyond visual heatmaps to include clinical feasibility and robustness [36,37]. Furthermore, the latest research has also investigated computational efficiency and lightweight architectures to ease deployment in clinical environments. For instance, the lightweight CNN backbones integrated with efficient wavelet-based compression were developed to reduce computational complexity while maintaining classification performance at the same time [38]. To represent a power within compact architectures, parametric wavelet transforms and attention-based fusion mechanisms have further enhanced adaptability [39].

An integration of wavelet decomposition results to CNN as input to different layers is proposed. Wavelet transformation captures both global context and local structural details by decomposing an image into approximation and directional detail components. It also provides a spatial representation that highlights edges and textural variations.

2. Materials and Methods

The processing pipeline of the proposed Wavelet-CNN model for Brain Tumor MRI classification is summarized (Figure 1). Each pre-processed axial brain MRI slice of size 256 × 256 × 1 is decomposed using a two-dimensional Haar wavelet transform. At each level of decomposition, the latest approximation component is split into a new approximation subband and three detail subbands, which are vertical, horizontal, and diagonal higher-frequency information, respectively. Then, the approximation component is recursively decomposed to obtain a multi-level representation. At the same time, the corresponding triplets of detail subbands from each level are routed to a convolutional neural network module.

Each CNN branch processes the vertical, horizontal, and diagonal subbands from a specific scale, learns level-specific texture, and outputs a compact feature representation. Then, a deeper convolutional neural network (CNN) block combines high-frequency details cues with increasingly coarse contextual information from lower-resolution levels by fusing the features obtained across all scales. Finally, the fused representation is converted into a one-dimensional feature vector using global average pooling and passed to a feed-forward neural network classifier composed of fully connected layers. The final output layer produces the predicted class label, distinguishing among the image classes, and then the proposed architecture is evaluated through validation and independent test results, where classification accuracy and cross-entropy loss are used to track the training of the model. Final performance metrics are then reported on the independent test set using loss, accuracy, and macro-F1 with additional per-class precision, recall, and F1-score, supported by confusion matrix and classification analysis.

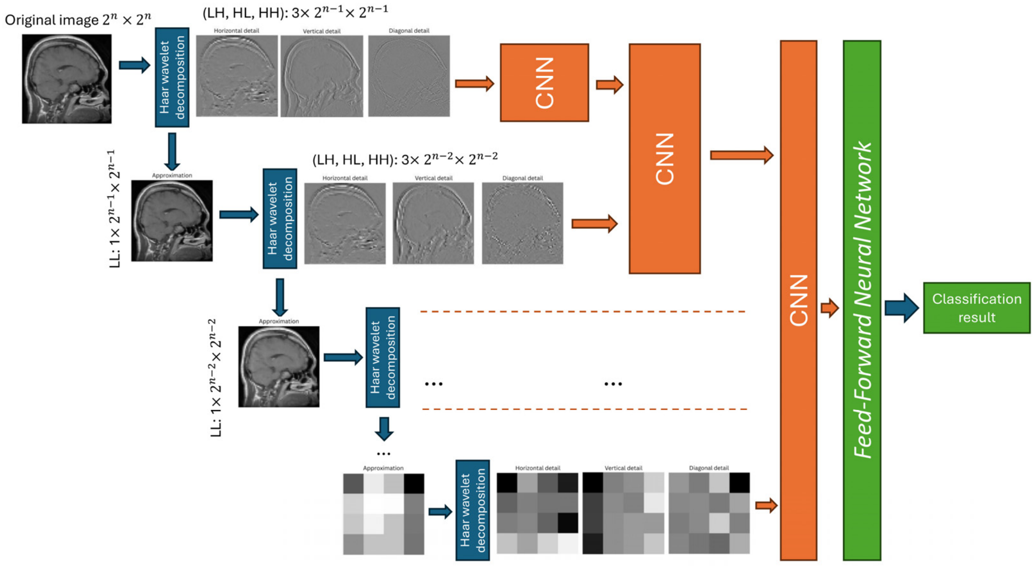

Figure 1. Schematic of the multi-scale Haar decomposition and CNN fusion model.

2.1. Dataset Description

The Brain Tumor MRI dataset is a publicly available collection of 7023 images from three clinical sources with 512 × 512 grayscale brain MRI slices, gathered from several sources and grouped into four diagnostic classes: meningioma, glioma, pituitary, and non-tumor. The examples of each class are shown in Figure 2 [40]. The dataset is split into training and test sets, resulting in 5712 images for training and 1311 for testing, where the training set contains 1595 no-tumor, 1339 meningioma, 1321 glioma, and 1457 pituitary images, while the test set includes 405 no-tumor, 306 meningioma, 300 glioma, and 300 pituitary images in these classes, correspondingly [41].

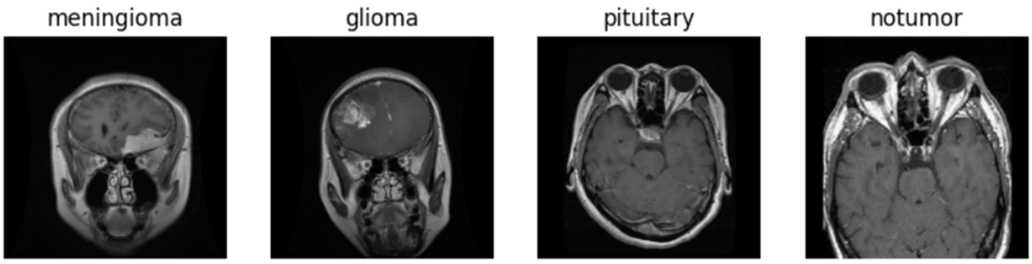

Figure 2. Brain Tumor MRI dataset.

This dataset was chosen for the following three reasons. Firstly, it is well known for using benchmarks for multi-class brain tumor classification, allowing a proper comparison with recent convolutional neural network models and transformer-based methods that report results on the same data [42]. Secondly, the variability in scanner settings, acquisition protocols, and patient populations in this dataset makes the classification task more challenging, as it is a collection of three independent clinical sources, thereby improving the robustness of the trained models [43]. Lastly, the inclusion of a specific no-tumor class reflects a realistic diagnostic scenario in which the models should recognize normal tissue and other brain tumor type classes.

To further evaluate the robustness of the proposed architecture, a publicly available dataset, Br35H :: Brain Tumor Detection 2020 dataset, was also used for supplementary evaluation. The images in this dataset contain 3060 brain MRI images (Figure 3), which were organized into 3 folders: yes, no, and pred. This additional dataset provides further evidence regarding the generalization ability of the proposed architecture across different publicly available MRI collections [44].

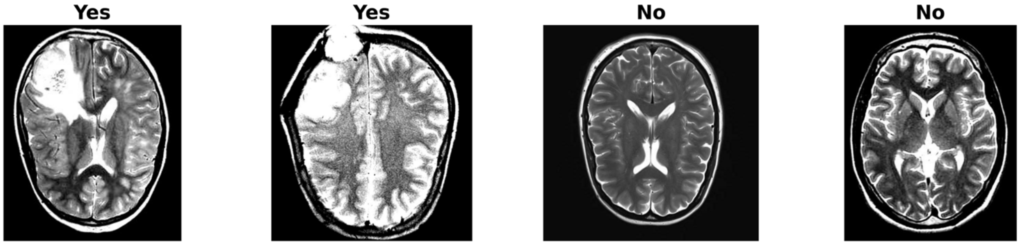

Figure 3. Br35H dataset.

2.2. Two-Dimensional Multi-Level Haar Wavelet Decomposition

To obtain a multi-scale representation of each axial brain MRI slice, the proposed architecture used the two-dimensional Haar wavelet transform. This is because Haar wavelets separate low-frequency detail from texture information with low computational cost. In this research, the input images are resized to 256 × 256 to enable twofold decomposition up to eight levels, since each level downsamples the spatial resolution by a factor of two. Despite the fact that the reduction in the input resolution removes some fine-pixel details, such resolution still preserves the overall shape of the tumor and lesion boundaries. Furthermore, the proposed architecture partially compensates for this reduction by separating the image into approximation and directional detail subbands and also enables the network to preserve significant edge and texture information across different scales. Therefore, resizing these images represents a trade-off between computational efficiency and information preservation, while maintaining sufficient textural content for reliable tumor classification.

In addition, the choice of 256 × 256 was made not only for computational convenience, but also to keep enough anatomical information for the classification task used in this study. At the same time, using the original 512 × 512 images or other larger resolutions would considerably increase memory usage, training time, and feature-map size throughout the network, and may increase the risk of overfitting if the model capacity and regularization are not adjusted accordingly. Therefore, 256 × 256 was selected in this study as a practical compromise between preserving diagnostically useful MRI information and keeping the model trainable, stable, and comparable across all experiments.

In image processing, the 2D Haar decomposition creates four subbands, where one is the approximation component (LL), and the remaining three subbands are detail components (LH, HL, and HH). In our proposed implementation, this operation is obtained as a fixed 2 × 2 strided convolution with same padding using Haar kernels that generate the four subbands directly. These outcomes match the standard 2D Haar DWT (discrete wavelet transform) interpretation, where low–low (LL) subband represents the approximation information and (LH, HL, HH) represent the directional detail responses.

To obtain a multi-level representation, the transform is applied recursively to the approximation subband.

$$\left(\mathrm{LL}^{(l)}, \mathrm{LH}^{(l)}, \mathrm{HL}^{(l)}, \mathrm{HH}^{(l)}\right) = \mathrm{DWT2}\left(\mathrm{LL}^{(l-1)}\right) \quad (1)$$

where l denotes the decomposition level and $LL^{(0)}$ is the input image; DWT2 means a one-level 2D discrete wavelet transform.

After l levels, the LL subband provides a progressively coarser approximation, while the detail subbands capture high-frequency information at the corresponding scale. In the proposed architecture, 2D Haar DWT is applied recursively to the LL subband to obtain multi-level representations, and the resulting detail subbands from each level are retained for multi-scale analysis in the subsequent network stages.

At the same time, each 2D Haar decomposition reduces the spatial resolution of the LL approximation subband by a factor of two in each dimension, so it is important to select the number of decomposition levels carefully. Therefore, for an input image of size 256 × 256, the LL representation undergoes a sequential reduction across 8 levels, reaching a 1 × 1 resolution at level 8. At this deepest level, the approximation branch no longer preserves explicit spatial localization of the tumor, because the full image is compressed into a single coefficient. For this reason, the 8-level decomposition was intentionally included as an upper-bound experimental condition to examine how far recursive decomposition can be extended before information loss outweighs the benefit of multi-scale sparsity. In contrast, the 7-level decomposition still retains a 2 × 2 LL approximation map, which is extremely coarse but still preserves minimal spatial organization of the image. Consequently, the 7-level decomposition can be interpreted as the deepest decomposition level that still preserves a spatially distributed approximation representation, whereas the 8-level decomposition crosses this limit by reducing the deepest LL approximation branch to a 1 × 1 representation, while the model still retains and fuses directional detail information from the current and earlier decomposition levels.

2.3. Proposed WaveletFusion Architecture

The proposed WaveletFusion model was built by combining multi-resolution Haar wavelet decomposition with a CNN backbone for classification (Figure 4). The key difference between the Baseline CNN is the wavelet decomposition levels included in the feature fusion pathway. After preprocessing, the input image is decomposed using a two-dimensional Haar wavelet transform into an approximation component and three detail components. At each decomposition stage, the three detail subbands are concatenated and passed through a convolutional feature extractor consisting of two 3 × 3 convolutional layers with batch normalization and ReLU activation. In parallel, the approximation subband is also processed by a small convolutional block before fusion. Afterward, the resulting approximation and detail subbands are fused and propagated through the CNN backbone through repeated convolutional blocks and downsampling, making higher-level representations. A fully connected layer, global average pooling, and softmax output are included in the classification head for the prediction of Brain MRI classes.

More precisely, the proposed WaveletFusion model processes the input MRI slice in a step-by-step manner. Firstly, the grayscale MRI slice of size 256 × 256 × 1 is pre-processed using the same pipeline as the Baseline CNN, including augmentation, rescaling, per-image standardization, and Gaussian noise injection during training. Afterward, the pre-processed image is passed through a one-level 2D Haar wavelet transform, which decomposes it into one LL subband and three directional detail subbands (LH, HL, and HH). Furthermore, these three detail subbands are then concatenated and processed by a small convolutional block, while the approximation subband is processed in parallel by another convolutional block. The extracted features are fused with the backbone representation from the previous stage, and the approximation subband is recursively decomposed, and the same procedure is repeated across the selected decomposition levels, allowing the model to learn multi-scale representations from fine to coarse resolutions. In the end, after the last fusion stage, the resulting deep representation is passed through global average pooling, a dense layer, and a softmax classifier to produce the final prediction.

By jointly operating the recursive decomposition pathway and the CNN feature fusion backbone, the proposed architecture does not rely only on spatial domain learning from the original MRI slice. It integrates approximation and directional detail information across multiple scales, allowing the model to combine coarse structural context with fine-grained edge and texture cues during classification.

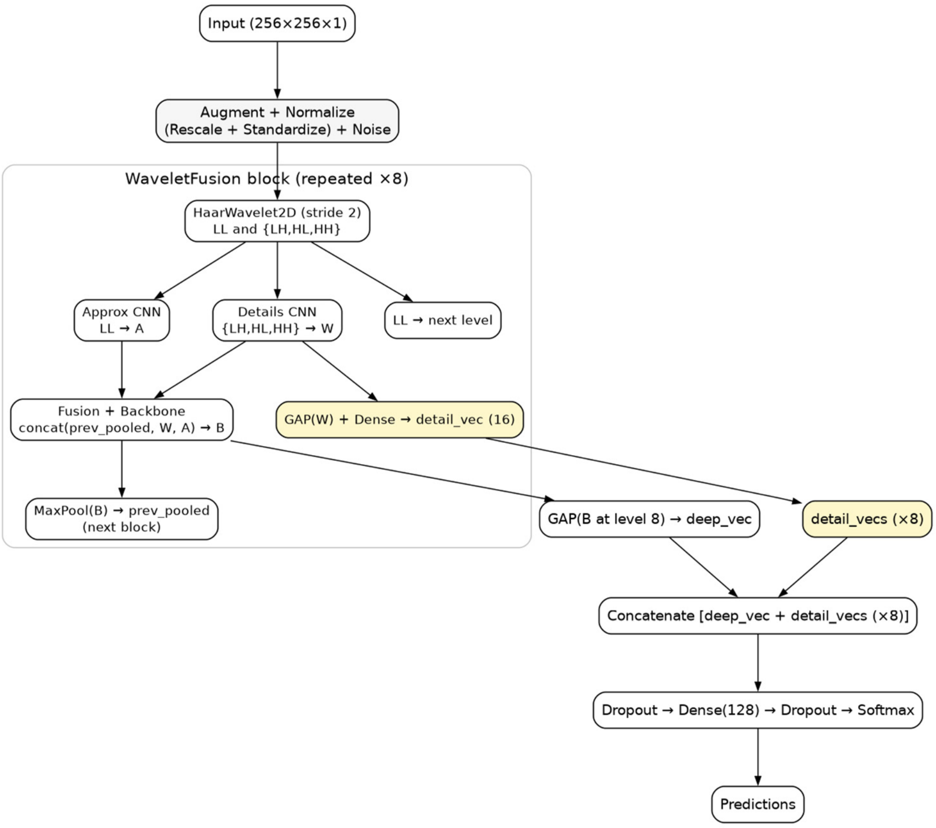

Figure 4. The architecture of the proposed WaveletFusion model.

In this research, to ensure a fair comparison, the Baseline CNN model was trained under the same preprocessing and optimization setup. The lightweight convolutional network takes a grayscale MRI slice of size 256 × 256 × 1 as input and applies the same augmentation pipeline, rescaling, per-image standardization, and Gaussian noise injection during training. The feature extractor consists of four convolutional stages with filter sizes 16, 32, 64, and 128, where each stage uses two 3 × 3 Conv2D layers followed by batch normalization and ReLU activation, and spatial dropout is applied within stages to reduce overfitting. A 2 × 2 max-pooling is used after the first three stages to downsample the feature maps (Figure 5).

After the final convolutional stage, a global average pooling layer is used to obtain a compact feature vector, followed by a fully connected layer with 128 units and dropout regularization, and a final softmax layer produces predictions for the four classes. The only difference between the two architectures is that the Baseline CNN does not include the multi-level Haar decomposition and fusion pathway.

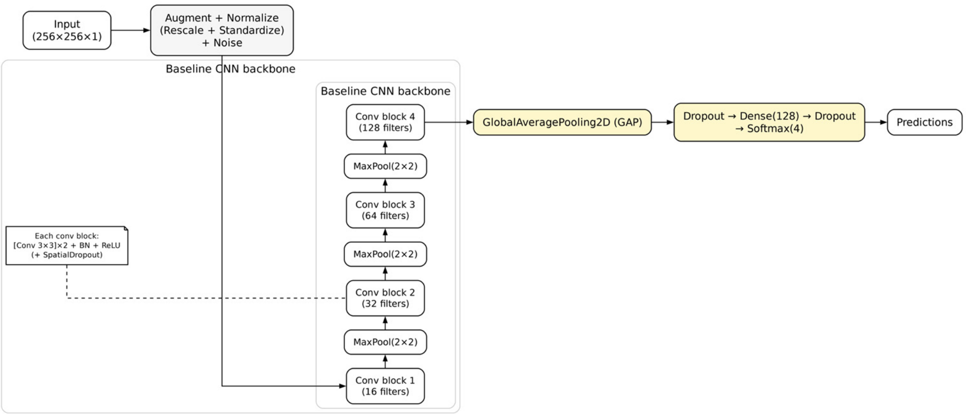

Figure 5. The architecture of the Baseline CNN.

2.4. Model Evaluation

Both WaveletFusion and Baseline CNN models share a similar reporting and CNN architecture, with the only difference being that the Baseline CNN learns spatial features directly from the input MRI slice using convolutional and pooling layers. In contrast, the WaveletFusion 8-level model first applies an 8-level two-dimensional decomposition and fuses multi-scale details and approximation features before classification. After training, benchmark results are provided to compare parameter count, accuracy, loss, and training time, allowing review and evaluation of the outcomes.

In multi-class classification, accuracy is the proportion of correctly predicted cases from the total number of predictions.

$$\mathrm{Accuracy} = \frac{1}{\mathrm{n_{samples}}} \sum_{\mathrm{i=0}}^{\mathrm{n_{samples}-1}} 1(\widehat{y}_\mathrm{i} = \mathrm{y_i}) \tag{2}$$

where $\mathrm{y^i}$ represents the true label of the ith sample in the dataset, $\widehat{y}_\mathrm{i}$ are the predicted labels produced by the model, and $\mathrm{n_{samples}}$ is the total number of samples in the dataset. The function $1(\widehat{y}_\mathrm{i} = \mathrm{y_i})$ returns 1 if the prediction matches the actual label and 0 otherwise.

Categorical cross-entropy loss for a multi-class classification problem is used as a loss function to measure how well a model predicts the correct class.

$$\mathrm{Loss} = -\sum_{\mathrm{i=1}}^{\mathrm{C}} \mathrm{y_i} * \log(\widehat{y}_\mathrm{i}) \tag{3}$$

where $\mathrm{y^i}$ is the true label from the one-hot encoded target vector, $\widehat{y}_\mathrm{i}$ is the predicted probability for the class i and C is the number of classes.

To evaluate how well a model performs on a classification of Brain Tumor MRI classes, precision, recall, and F1-score are provided to measure how many samples predicted as a given class are actually correct, how many samples of that actual class are successfully detected, and provide a single balanced measure when the two differ.

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP + FP}}, \mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP + FN}}, \mathrm{F1} = \frac{2\ \times \mathrm{Precision}\ \times \mathrm{Recall}}{\mathrm{Precision + Recall}} \tag{4}$$

where TP (true positives) is the number of samples of the target class correctly predicted as that class, FP (false positives) are incorrectly predicted positive cases, and FN (false negatives) are incorrectly predicted negative cases.

Additionally, for predictive performance, training efficiency is characterized using the average epoch time and training FPS (frames per second), computed as the number of training images divided by the average epoch duration.

3. Results

The proposed WaveletFusion architecture illustrates classification performance by comparing an eight-level decomposition against the Baseline CNN model, where all configurations were trained on the same Brain Tumor MRI dataset and Br35H datasets, using the same preprocessing augmentation, optimization settings, and evaluation procedure to ensure a fair comparison. Both architectures were trained and evaluated in five independent trials.

3.1. Experimental Setup and Evaluation Protocol

All MRI images loaded into the models were resized to 256 × 256, rescaled to [0, 1], and standardized per image. For the WaveletFusion models, a two-dimensional Haar wavelet transform was applied up to eight decomposition levels. At each level, only horizontal, vertical, and diagonal components are concatenated and processed by a small convolution block consisting of two convolution layers. Additionally, the approximation component was applied to the next level for further decomposition and was also processed and fused with the detail features within the multi-level backbone.

A repeated hold-out validation was performed five times for the Brain Tumor MRI dataset by splitting the training dataset into training and validation subsets at 80% and 20%, respectively, with different run seeds. The test dataset was given in the dataset and was used as an independent test set to evaluate the model. Thus, the current data does not allow for reconstructing the patient-wise grouping or splitting the images at a patient level.

For the Br35H dataset, the supervised evaluation was performed only on the labeled yes (1500 images) and no (1500 images) classes, while the pred folder was removed because it does not provide ground-truth class labels. Since this dataset does not include a predefined test partition, a repeated hold-out protocol was also performed five times, first splitting the labeled yes/no images into training/validation and independent test subsets using a stratified 80%/20% split. Afterward, the training–validation subset was further split into training and validation subsets at 80% and 20%, respectively, again using stratification.

For both datasets, the splits were performed at an image level because patient identifiers were not available. Per-class precision, recall, and F1-score are computed for test-set predictions. The per-class tables show results from the report-seed run with the best validation accuracy, while averaged results across seeds are reported separately where indicated.

3.2. Ablation Study of Wavelet Decomposition Depth

To analyze the effect of wavelet decomposition depth on the proposed WaveletFusion architecture, it was trained with decomposition depths ranging from 1-level to 8-level (L1 to L8) under the same experimental protocol. For each decomposition level, the model was evaluated across five runs using identical training, validation, and test procedures (Table 1).

For the Brain Tumor MRI dataset, the test results showed a gradual increase in accuracy from 0.61 ± 0.02 at L1 to 0.94 ± 0.02 at L7, while macro-F1 improved from 0.55 ± 0.03 to 0.93 ± 0.02 over the same range. Similarly, the validation loss decreased substantially from 0.94 ± 0.04 at L1 to 0.49 ± 0.02 at L7, while validation accuracy and validation macro-F1 reached 0.94 ± 0.01 and 0.94 ± 0.01, respectively. These results illustrate that increasing the number of decomposition levels, specifically between L3 and L6, improves performance, with the best results at L7.

Furthermore, a similar trend was observed in the Br35H dataset. The test accuracy increased from 0.81 ± 0.02 at L1 to 0.97 ± 0.01 at L7. The validation loss also decreased from 0.48 ± 0.06 at L1 to 0.34 ± 0.03 at L7, while validation accuracy reached 0.97 ± 0.02. Although small fluctuations were observed at intermediate depths, the overall trend remained steady, and the best results were again obtained at L7 depth.

Table 1. Quantitative evaluation of WaveletFusion models with varying wavelet decomposition depths (L1–L8) across 5 runs (mean ± standard deviation) on the Brain Tumor MRI and Br35H datasets.

Dataset	Depth	Parameters	Validation Loss	Validation Accuracy	Validation Macro-F1	Test Loss	Test Accuracy	Test Macro-F1
Brain Tumor MRI	L1	19,708	0.94 ± 0.04	0.64 ± 0.03	0.59 ± 0.05	1.04 ± 0.04	0.61 ± 0.02	0.55 ± 0.03
	L2	58,816	0.79 ± 0.04	0.72 ± 0.02	0.64 ± 0.03	0.86 ± 0.06	0.70 ± 0.03	0.63 ± 0.04
	L3	101,812	0.70 ± 0.02	0.78 ± 0.02	0.73 ± 0.04	0.76 ± 0.03	0.75 ± 0.01	0.70 ± 0.03
	L4	143,862	0.59 ± 0.05	0.86 ± 0.03	0.84 ± 0.04	0.64 ± 0.04	0.83 ± 0.03	0.81 ± 0.04
	L5	185,912	0.57 ± 0.06	0.89 ± 0.03	0.88 ± 0.03	0.62 ± 0.06	0.86 ± 0.03	0.84 ± 0.04
	L6	226,888	0.49 ± 0.02	0.94 ± 0.01	0.93 ± 0.01	0.52 ± 0.03	0.92 ± 0.01	0.92 ± 0.01
	L7	267,864	0.49 ± 0.02	0.94 ± 0.01	0.94 ± 0.01	0.51 ± 0.03	0.94 ± 0.02	0.93 ± 0.02
	L8	308,840	0.55 ± 0.04	0.91 ± 0.02	0.91 ± 0.02	0.60 ± 0.05	0.90 ± 0.02	0.89 ± 0.03
Br35H	L1	19,450	0.48 ± 0.06	0.81 ± 0.04	0.81 ± 0.04	0.48 ± 0.03	0.81 ± 0.02	0.81 ± 0.02
	L2	58,358	0.47 ± 0.07	0.84 ± 0.04	0.84 ± 0.04	0.46 ± 0.03	0.85 ± 0.01	0.85 ± 0.01
	L3	101,554	0.45 ± 0.05	0.87 ± 0.03	0.86 ± 0.03	0.44 ± 0.01	0.88 ± 0.01	0.88 ± 0.01
	L4	143,604	0.47 ± 0.04	0.86 ± 0.03	0.86 ± 0.03	0.47 ± 0.01	0.86 ± 0.01	0.86 ± 0.01
	L5	185,654	0.47 ± 0.06	0.87 ± 0.04	0.87 ± 0.04	0.46 ± 0.04	0.88 ± 0.02	0.88 ± 0.02
	L6	226,630	0.36 ± 0.08	0.95 ± 0.03	0.95 ± 0.03	0.37 ± 0.08	0.94 ± 0.04	0.94 ± 0.04
	L7	267,606	0.34 ± 0.03	0.97 ± 0.02	0.97 ± 0.02	0.35 ± 0.02	0.97 ± 0.01	0.97 ± 0.01
	L8	308,582	0.43 ± 0.05	0.92 ± 0.03	0.92 ± 0.03	0.43 ± 0.04	0.92 ± 0.02	0.92 ± 0.02

However, the models' performance across decomposition levels (Figure 6) shows that excessive decomposition can remove spatial information. This is clearly visible where the 7-level (L7) decomposition achieved the best overall performance. However, the 8-level decomposition resulted in a decline, with test accuracy decreasing to 0.90 ± 0.02 and test macro-F1 to 0.89 ± 0.03. Correspondingly, for the Br35H dataset, the L8 configuration reduced test accuracy and test macro-F1 to 0.92 ± 0.02. Therefore, the L8 was included to evaluate the upper limit of decomposition depth and to analyze the trade-off between representation compactness and information loss. The results indicate that L7 provides a better balance between multi-scale sparsity and information loss, while L8 becomes less effective in this trade-off.

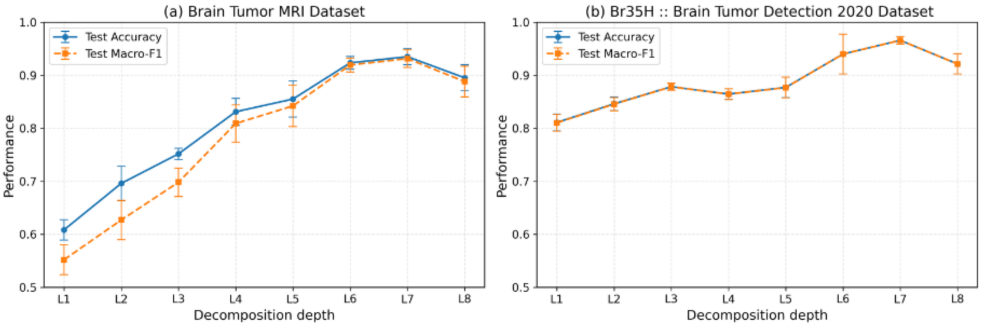

Figure 6. Test accuracy and macro-F1 of WaveletFusion across different wavelet decomposition depths for (**a**) Brain Tumor MRI and (**b**) Br35H datasets across 5 runs.

3.3. Ablation Study of Directional Detail Subbands

Since the depth-ablation results identified L7 as the best-performing configuration, it was chosen to analyze the contribution of directional wavelet detail subbands. To examine whether the contribution of directional wavelet detail subbands remains stable across the strongest decomposition depths, L6 and L8 were also chosen for the same subband ablation analysis. The same masking-based procedure was used for all three cases, in which selected detail subbands were retained at all decomposition levels, while the remaining detail components were suppressed. This procedure allowed the discriminative contributions of the LH, HL, and HH detail responses to be compared under the same training and evaluation protocol.

The comparison of test macro-F1 across wavelet detail subband configuration for the L6, L7 and L8 are provided (Figure 7 and Table 2), where the yellow line in each box plot represents the median, the dots indicate outlier seed results, and the blue dot shows the configuration with the highest median Test Macro-F1. It showed that the HH-only configuration remained the weakest setting, whereas LH and HL configurations consistently produced stronger results in both datasets. For the Brain Tumor Brain MRI dataset, in the L6 setting, the LH + HL configuration achieved the best overall performance, reaching 0.94 ± 0.01 test accuracy. The best single-subband result in the L8 setting was obtained with the HL component, achieving 0.91 ± 0.02 test accuracy, while the full three-subband model (LH + HL + HH) achieved 0.90 ± 0.02. However, in comparison with the L7 setting as in the previous analysis, it produced stronger performance across all subband configurations, especially in the LH + HL configuration, where it showed the best overall performance reaching 0.94 ± 0.01 test accuracy and 0.93 ± 0.01 test macro-F1, indicating that the jointly used horizontal and vertical directional details were sufficient to obtain the strongest predictive performance.

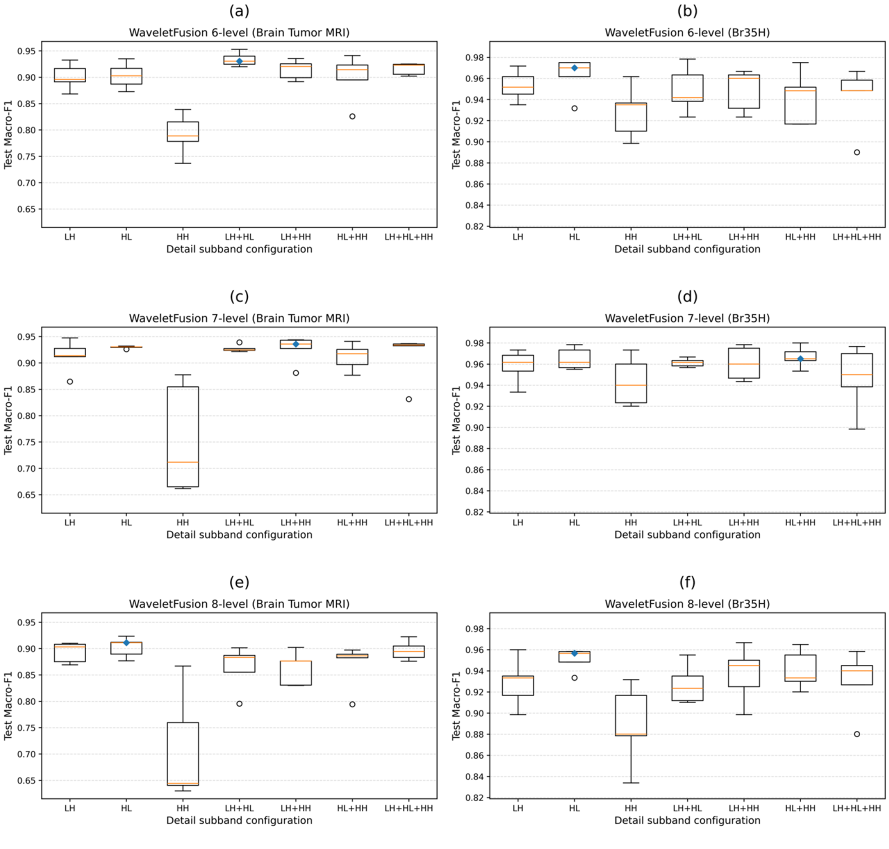

Figure 7. Effect of wavelet detail subband configuration on test macro-F1 across 5 runs for (**a**) WaveletFusion 6-level on the Brain Tumor MRI dataset, (**b**) WaveletFusion 6-level on the Br35H dataset,

(**c**) WaveletFusion 7-level on the Brain Tumor MRI dataset, (**d**) WaveletFusion 7-level on the Br35H dataset, (**e**) WaveletFusion 8-level on the Brain Tumor MRI dataset, and (**f**) WaveletFusion 8-level on the Br35H dataset.

A similar tendency was also observed for the Br35H dataset, although the overall performance level was higher across all subband configurations. In this case, the HL component produced the strongest single-subband result at all three decomposition depths, reaching 0.96 ± 0.02 test accuracy and 0.96 ± 0.02 test macro-F1 at L6, 0.97 ± 0.01 in both test accuracy and test macro-F1 at L7, and 0.95 ± 0.01 in both metrics at L8.

Table 2. Test-set comparison of WaveletFusion subband configurations at decomposition depths L6, L7, and L8 for Brain Tumor MRI and Br35H datasets across 5 runs (mean ± standard deviation).

Dataset	Subband Configuration	L6 Test Accuracy	L6 Test Macro-F1	L7 Test Accuracy	L7 Test Macro-F1	L8 Test Accuracy	L8 Test Macro-F1
Brain Tumor MRI	LH	0.91 ± 0.02	0.90 ± 0.02	0.93 ± 0.01	0.93 ± 0.02	0.90 ± 0.02	0.89 ± 0.02
	HL	0.91 ± 0.02	0.90 ± 0.02	0.93 ± 0.02	0.93 ± 0.02	0.91 ± 0.02	0.90 ± 0.02
	HH	0.82 ± 0.03	0.79 ± 0.04	0.82 ± 0.08	0.80 ± 0.09	0.77 ± 0.07	0.71 ± 0.10
	LH + HL	0.94 ± 0.01	0.94 ± 0.01	0.94 ± 0.01	0.93 ± 0.01	0.88 ± 0.03	0.86 ± 0.03
	LH + HH	0.92 ± 0.02	0.92 ± 0.02	0.92 ± 0.02	0.92 ± 0.02	0.87 ± 0.03	0.86 ± 0.03
	HL + HH	0.91 ± 0.04	0.91 ± 0.04	0.93 ± 0.01	0.92 ± 0.01	0.88 ± 0.04	0.87 ± 0.04
	LH + HL + HH	0.92 ± 0.01	0.92 ± 0.01	0.93 ± 0.01	0.92 ± 0.01	0.90 ± 0.02	0.90 ± 0.02
Br35H	LH	0.95 ± 0.01	0.95 ± 0.01	0.96 ± 0.02	0.96 ± 0.02	0.93 ± 0.02	0.93 ± 0.02
	HL	0.96 ± 0.02	0.96 ± 0.02	0.97 ± 0.01	0.97 ± 0.01	0.95 ± 0.01	0.95 ± 0.01
	HH	0.93 ± 0.02	0.93 ± 0.02	0.94 ± 0.02	0.94 ± 0.02	0.89 ± 0.04	0.89 ± 0.04
	LH + HL	0.95 ± 0.02	0.95 ± 0.02	0.96 ± 0.00	0.96 ± 0.00	0.94 ± 0.02	0.93 ± 0.02
	LH + HH	0.95 ± 0.02	0.95 ± 0.02	0.96 ± 0.02	0.96 ± 0.02	0.94 ± 0.03	0.94 ± 0.03
	HL + HH	0.94 ± 0.02	0.94 ± 0.02	0.97 ± 0.01	0.97 ± 0.01	0.94 ± 0.02	0.94 ± 0.02
	LH + HL + HH	0.94 ± 0.03	0.94 ± 0.03	0.95 ± 0.03	0.95 ± 0.03	0.93 ± 0.03	0.93 ± 0.03

Therefore, the results from both datasets support the stability of the relative contribution pattern of the detail subbands, where LH and HL components carry the most useful discriminative information, while HH contributes less when used alone. The L6 model follows the same general tendency. Therefore, the comparison between these three models not only supports the previously illustrated depth-ablation results but also suggests a subsequent interpretability analysis.

3.4. Ablation Study of Level-Group Contributions

To further analyze which parts of the multi-level wavelet representation contribute most, an additional ablation was carried out for the WaveletFusion 6-level, WaveletFusion 7-level, and WaveletFusion 8-level models. These three models were selected because they achieved the strongest results in the previous analysis, whereas lower decomposition levels showed clearly weaker overall classification performance. For the WaveletFusion 6-level model, the retained groups were defined as early-scale (L1–L2), intermediate-scale (L3–L4), coarse-scale (L5–L6), and full-scale (L1–L6). For the WaveletFusion 7-level model, the retained groups were defined as early-scale (L1–L2), intermediate-scale (L3–L5), coarse-scale (L6–L7), and full-scale (L1–L7). And lastly, for the WaveletFusion 8-level model, the retained groups were defined as early-scale (L1–L2), intermediate-scale (L3–L5), coarse-scale (L6–L8), and full-scale (L1–L8).

The results for the WaveletFusion 8-level model (Figure 8) showed that the early-scale fusion configuration achieved the strongest overall performance with a test accuracy of 0.89 ± 0.02 and a test macro-F1 of 0.88 ± 0.02. The full-scale fusion setting was slightly weaker, with 0.88 ± 0.02 test accuracy and 0.87 ± 0.03 test macro-F1. In comparison, the intermediate-scale fusion configuration showed a clear decrease to 0.84 ± 0.05 test accuracy and 0.82 ± 0.06 test macro-F1, whereas the coarse-scale fusion configuration produced the weakest results, with only 0.73 ± 0.01 test accuracy and 0.69 ± 0.01 test macro-F1. These findings indicate that the strongest discriminative contribution comes from the early, fine-resolution levels, whereas the deepest coarse levels are insufficient when used in isolation.

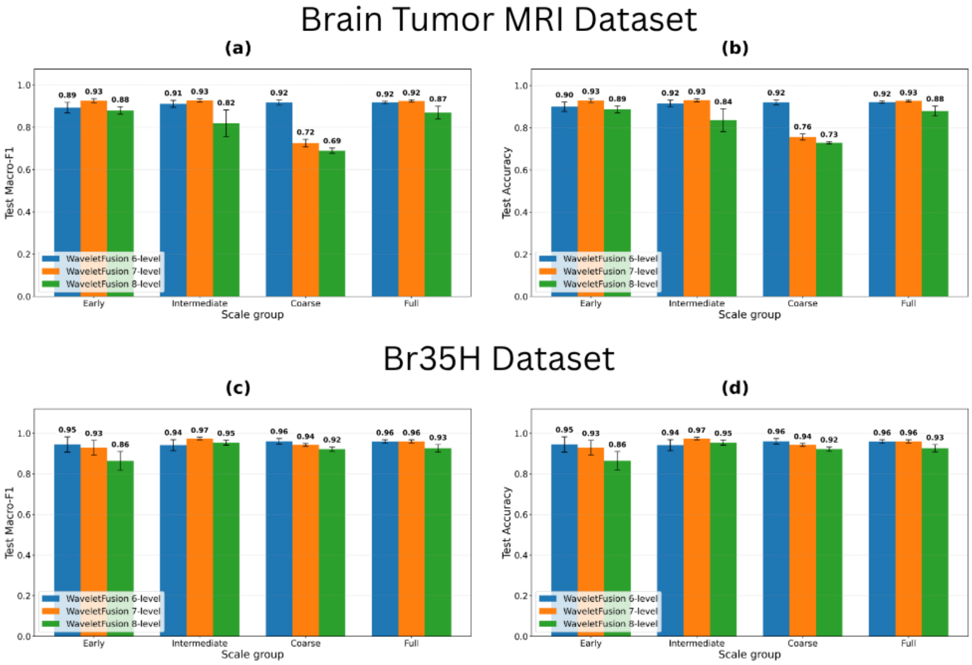

Figure 8. Test macro-F1 and accuracy for level-group ablations across 5 runs for (**a**) test macro-F1 on the Brain Tumor MRI dataset, (**b**) test accuracy on the Brain Tumor MRI dataset, (**c**) test macro-F1 on the Br35H dataset, and (**d**) test accuracy on the Br35H dataset, comparing the WaveletFusion 6-level, WaveletFusion 7-level, and WaveletFusion 8-level models.

Similarly, an even stronger tendency was observed for the WaveletFusion 7-level model, where the best test-set result was obtained by intermediate-scale fusion, which reached a test accuracy of 0.93 ± 0.01 and a test macro-F1 of 0.93 ± 0.01. The early-scale fusion and full-scale fusion settings followed very closely. In contrast, the coarse-scale fusion configuration again produced substantially lower performance, which again confirms that the coarse low-resolution information alone is not sufficient, while early and intermediate scales contain the most useful discriminative cues. For the WaveletFusion 6-level model, the intermediate-scale fusion configuration also produced the best result, reaching 0.91 ± 0.01 test accuracy and 0.91 ± 0.01 test macro-F1, while the coarse-scale fusion configuration remained weaker than the early-scale and full-scale settings.

For the Br35H dataset, the WaveletFusion 6-level and WaveletFusion 7-level models showed the strongest results in the intermediate-scale and full-scale settings, while the coarse-scale configuration again remained weaker than the early and full settings. For the WaveletFusion 8-level model, the intermediate-scale and full-scale settings also remained stronger than the coarse-scale configuration, confirming that the most useful discriminative information is concentrated in the early-to-intermediate decomposition levels, whereas the deepest coarse levels alone are not sufficient for the strongest predictive performance.

3.5. Explainable Analysis of the WaveletFusion 8-Level Model

To understand how the WaveletFusion 8-level model forms its predictions, a layer-wise Grad-CAM analysis was carried out on representative test-set examples from the four classes and computed with respect to the predicted class logit for five successive convolutional blocks in the fusion backbone, namely B1 (Block 1) to B5 (Block 5). To explain successful predictions, one correctly classified image was selected for each class, and representative incorrectly classified images were also examined to identify the reasons. The correctly classified glioma and meningioma cases (Figure 9) show a clear progression from broad low-level responses to more selective late-layer attention.

Across all classes, blocks 1 and 2 show that they mainly respond to general skull-based boundaries, while blocks 3 to 5 become more discriminative. The deeper overlays in glioma and meningioma classes progressively contract toward the lesion-bearing regions, showing that the later fusion blocks are responsible for refining tumor localization. On the contrary, for the no-tumor case, the later blocks do not converge to a single mass-like hotspot. Instead, the responses remain comparatively weak and distributed across normal intracranial structures, consistent with the absence of a focal lesion. In contrast, the later responses in the pituitary class emphasize a compact central or inferior region, indicating that the model can isolate characteristic pituitary evidence despite the broader anatomical responses seen in the earlier blocks.

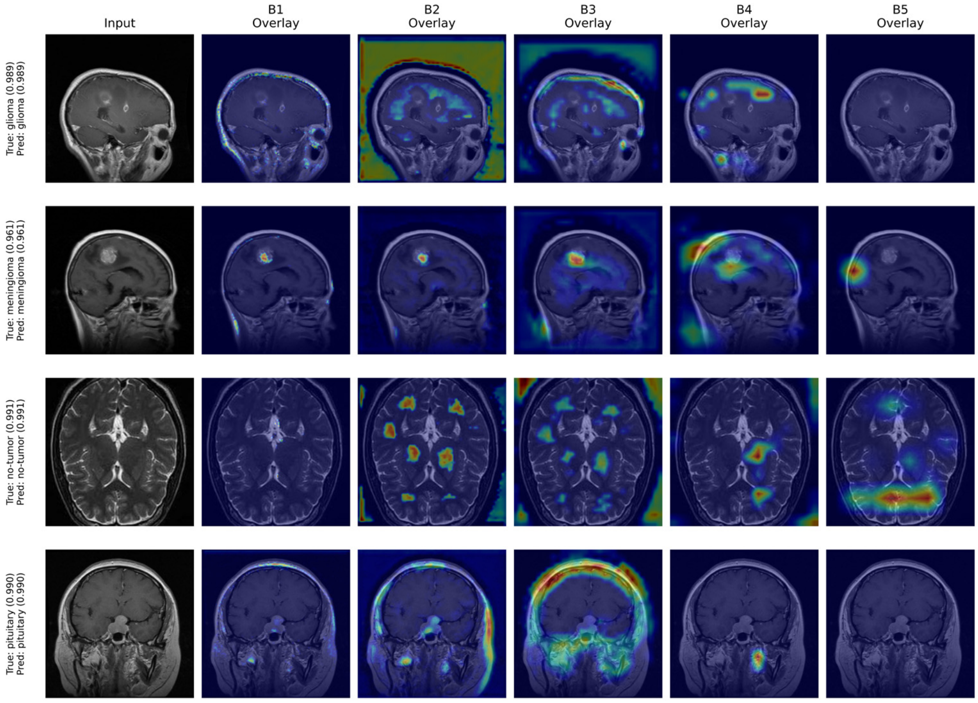

Figure 9. Correctly classified representative examples of the WaveletFusion 8-level model.

Representative error cases from all four classes (Figure 10), where Grad-CAM is targeted to the predicted class. These overlays illustrate why the model selected the incorrect label, especially in glioma and meningioma examples; the deeper blocks increasingly emphasize regions or surrounding anatomical patterns that resemble the competing class rather than isolating the full lesion area. In the no-tumor and pituitary cases, the deeper responses form compact hotspots over localized structures and strong intensity transitions, which may be interpreted by the model as tumor-like evidence.

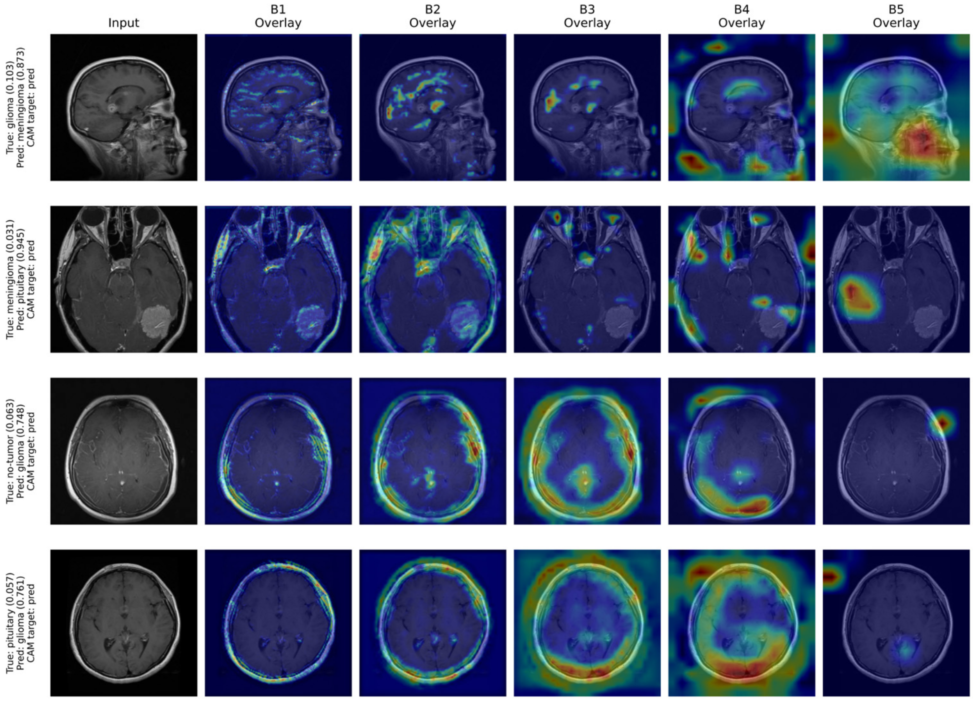

Figure 10. Incorrectly classified representative examples of the WaveletFusion 8-level model with Grad-CAM targeted to the predicted class.

The same representative error cases are shown in (Figure 11), but here Grad-CAM is targeted to the true class, and the corresponding true-class probability is reported for each example. Compared with Figure 10, the true-class responses are more fragmented or less spatially dominant in the deeper blocks, which means that the model still retains some evidence for the true class, but this evidence is insufficient to dominate the final decision. The lower true-class probability further supports this interpretation, showing that although the correct class is not entirely ignored, its contribution is outweighed by stronger activation patterns associated with the incorrect prediction.

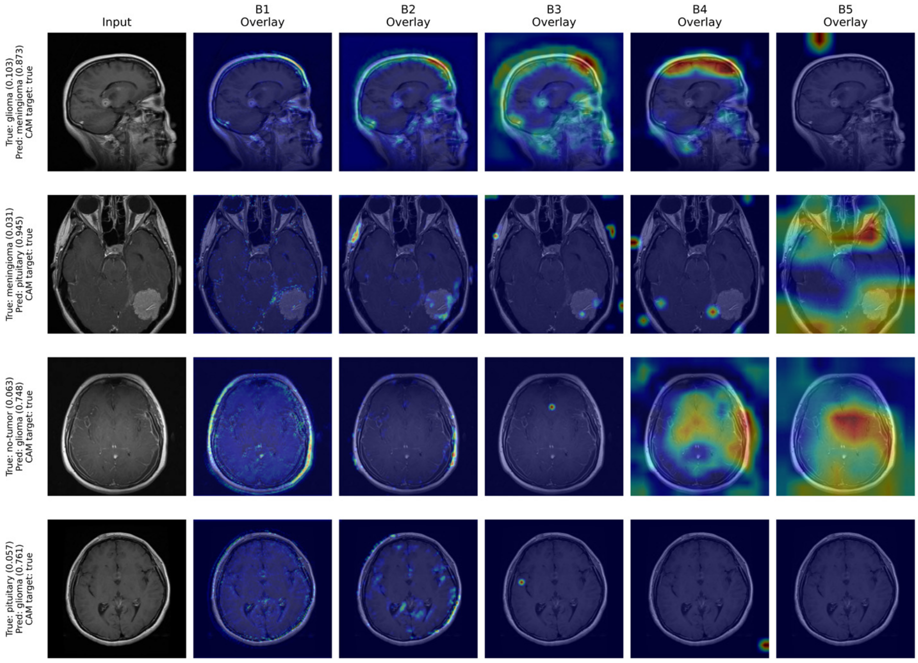

Figure 11. Incorrectly classified representative examples, but with Grad-CAM targeted to the true class.

https://doi.org/10.3390/app16094482

Overall, the Grad-CAM analysis confirms that the WaveletFusion 8-level model does not use all backbone blocks in the same way. Blocks 1 and 2 mainly respond to global structure and edge information, whereas blocks 4 and 5 are more directly linked to the final class decision. When these later blocks converge on misleading but discriminative regions, the model favors the wrong label even though some residual evidence for the true class remains present.

3.6. Comparative Performance Analysis of the Baseline CNN and WaveletFusion Models

Overall, integrating wavelet decompositions improved test accuracy compared to the Baseline CNN, and kept strong classification performance on the Brain Tumor MRI and Br35H datasets with slightly smaller parameter count (Table 3). However, the decomposition ablation further showed that the WaveletFusion 7-level configuration (267,864 parameters for Brain Tumor MRI and 267,606 parameters for Br35H) achieved the best overall performance, indicating that a larger model size did not cause the observed performance improvements. In terms of efficiency, the Baseline CNN model trains faster (average epoch time 6.78 s) than the WaveletFusion 7-level model (average epoch time 14.86 s), as expected due to the additional multi-level wavelet processing on the Brain Tumor MRI dataset. Similarly, the Baseline CNN also achieved lower inference latency (3.64 ms/img) and higher inference throughput (275.1 img/s) than the WaveletFusion 7-level and the WaveletFusion 8-level models. For the Br35H dataset, the same trend was observed, where the Baseline CNN again trained faster (average epoch time 3.32 s) than the WaveletFusion 8-level and WaveletFusion 7-level models, and also achieved lower inference latency (2.27 ms/img) and higher inference throughput (439.94 img/s). Although the WaveletFusion 7-level and WaveletFusion 8-level models showed lower FLOPs than the Baseline CNN (1690.28 M), indicating that FLOP-based complexity does not fully capture the execution overhead introduced by the wavelet decomposition.

Table 3. Computational characteristics of the Baseline CNN and WaveletFusion models.

Dataset	Model	Parameters	FLOPs (M)	Average Epoch Time	Train FPS (img/s)	Inference Latency (ms/img)	Inference Throughput (img/s)	Training Time (s)
Brain Tumor MRI	Baseline CNN	311,700	1690.28	6.78	674.5	3.64	275.1	169.5
	WaveletFusion 8-level	308,840	819.13	16.23	281.7	4.97	201.4	405.9
	WaveletFusion 7-level	267,864	819.05	14.86	307.6	4.70	212.7	371.7
Br35H	Baseline CNN	311,442	1690.28	3.32	577.52	2.27	439.94	46.64
	WaveletFusion 8-level	308,582	819.13	5.24	366.41	4.45	224.73	132.19
	WaveletFusion 7-level	267,606	819.05	5.63	341.19	4.73	211.31	140.94

The performance metric results of all the models are provided (Table 4). They showed that both WaveletFusion models outperform the Baseline CNN on the test set. The WaveletFusion 8-level model achieved a higher average test accuracy (0.89 ± 0.02) than the Baseline CNN (0.82 ± 0.02) across 5, while the test loss decreased from 0.68 ± 0.02 to 0.60 ± 0.03. Moreover, the WaveletFusion 8-level model improves the test macro-F1 results from 0.80 ± 0.03 to 0.89 ± 0.02. However, the WaveletFusion 7-level model showed the strongest results, reaching a test accuracy of 0.94 ± 0.01, while also achieving the lowest test loss (0.49 ± 0.02).

Table 4. Overall performance of the Baseline CNN and WaveletFusion models across 5 runs (mean ± standard deviation).

Dataset	Model	Test Loss	Test Accuracy	Test Macro-F1	Validation Loss	Validation Accuracy	Validation Macro-F1
Brain Tumor MRI	Baseline CNN	0.68 ± 0.02	0.82 ± 0.02	0.80 ± 0.03	0.62 ± 0.03	0.85 ± 0.02	0.84 ± 0.03
	WaveletFusion 8-level	0.60 ± 0.03	0.89 ± 0.02	0.89 ± 0.02	0.55 ± 0.03	0.92 ± 0.02	0.91 ± 0.02
	WaveletFusion 7-level	0.49 ± 0.02	0.94 ± 0.01	0.93 ± 0.02	0.49 ± 0.02	0.94 ± 0.01	0.94 ± 0.01
Br35H	Baseline CNN	0.54 ± 0.03	0.82 ± 0.03	0.81 ± 0.03	0.54 ± 0.04	0.81 ± 0.03	0.81 ± 0.03
	WaveletFusion 7-level	0.35 ± 0.02	0.97 ± 0.01	0.97 ± 0.01	0.35 ± 0.04	0.97 ± 0.02	0.97 ± 0.02
	WaveletFusion 8-level	0.41 ± 0.01	0.94 ± 0.01	0.94 ± 0.01	0.40 ± 0.03	0.94 ± 0.02	0.94 ± 0.02

A similar trend was observed in the Br35H dataset, where the Baseline CNN achieved test accuracy of 0.82 ± 0.03 and test macro-F1 of 0.81 ± 0.03. In contrast, the WaveletFusion 8-level model improved these results to 0.94 ± 0.01 for both test accuracy and test macro-F1. However, the WaveletFusion 7-level model again achieved the best overall performance, with a test accuracy of 0.97 ± 0.01 and the lowest test loss (0.35 ± 0.02).

Validation results follow the same trend, where WaveletFusion 8-level model validation loss decreases from 0.62 ± 0.03 to 0.55 ± 0.03, and validation accuracy increases from 0.85 ± 0.02 to 0.92 ± 0.02. The WaveletFusion 7-level model further improved these results, achieving the lowest validation loss (0.49 ± 0.02) together with the highest validation accuracy (0.94 ± 0.01) and validation macro-F1 (0.94 ± 0.01).

At the same time, the broader comparison analysis (Table 5) shows that the proposed WaveletFusion 7-level model also remained competitive with several previously published methods. It outperformed the results of Decision Tree, k-NN, ResNet50, and Custom CNN, but remained below stronger models such as Xception Base, DenseNet169 Base, and Fusion + CA + Chirplet Transform. For the Br35H dataset, the WaveletFusion 7-level model also showed competitive overall performance, outperforming ResNet50, DenseNet121, the proposed CNN, and InceptionV3, while remaining slightly below stronger Pre-BN CNN, Post-BN CNN, and Post-BN Res CNN models.

Table 5. Overall comparison of the proposed models with representative published methods on the Brain Tumor MRI and Br35H datasets.

Dataset	Model	Reported Accuracy	Reported F1-Score
Brain Tumor MRI	Baseline CNN	0.84	0.83
	WaveletFusion 8-level	0.91	0.91
	WaveletFusion 7-level	0.95	0.95
	Decision Tree [45]	0.83	0.83 *
	k-NN [45]	0.91	0.92 *
	ResNet50 [45]	0.91	0.92 *
	CNN Model [45]	0.90	0.95 *
	Xception Base [46]	0.96	0.95
	DenseNet169 Base [46]	0.98	0.98
	Fusion + CA + Chirplet Transform [46]	0.99	0.99
Br35H	Baseline CNN	0.85	0.84
	WaveletFusion 8-level	0.95	0.95
	WaveletFusion 7-level	0.98	0.98
	ResNet50 [47]	0.94	0.94
	ResNet101 [47]	0.96	0.96
	DenseNet121 [47]	0.93	0.93

	Proposed CNN [48]	0.94	0.94
	InceptionV3 [48]	0.77	0.69
	Pre-BN CNN [49]	0.94	0.98
	Post-BN CNN [49]	0.99	0.99
	Post-BN Res CNN [49]	0.99	0.99

* weighted F1-score reported.

The per-class metric analysis (Table 6) shows that the WaveletFusion models provide more consistent improvements across all classes. For example, both WaveletFusion configurations in the pituitary class maintained very high recall, indicating that the main features were preserved in both decomposition settings. However, in comparison with other classes, the meningioma class results remain lower than for the other classes, which can be considered the most challenging class. While the Baseline CNN shows weaker class balance, the WaveletFusion 7-level model provided the strongest overall performance and substantially reduced the gap in the meningioma class. Moreover, WaveletFusion models still achieved more balanced precision and recall results in the pituitary and no-tumor cases, despite the fact that the Baseline CNN model also classified these cases reliably. However, compared with the Dense121 Base and Dense169 Base models, the WaveletFusion 7-level model still shows lower class-wise scores in several categories, especially in the no-tumor and pituitary categories. The Fusion + CA + Chirplet Transform model reached near-perfect metrics across all classes.

Table 6. Per-class test metrics results of the Baseline CNN and WaveletFusion models for Brain Tumor MRI dataset.

Class	Metric	Baseline CNN	WaveletFusion 7-Level	WaveletFusion 8-Level	Dense121 Base [46]	Dense169 Base [46]	Fusion + CA + Chirplet Transform [46]
Glioma	Precision	0.8507	0.9617	0.9225	0.9900	0.9800	1.0000
	Recall	0.8167	0.9200	0.8733	0.9400	0.9400	0.9700
	F1-score	0.8333	0.9404	0.8973	0.9600	0.9600	0.9800
Meningioma	Precision	0.7546	0.9210	0.8764	0.9400	0.9300	0.9700
	Recall	0.5425	0.8758	0.7876	0.9700	0.9700	0.9900
	F1-score	0.6312	0.8978	0.8296	0.9600	0.9500	0.9800
No-tumor	Precision	0.8710	0.9591	0.9297	1.0000	1.0000	1.0000
	Recall	0.9457	0.9852	0.9803	1.0000	1.0000	1.0000
	F1-score	0.9605	0.9720	0.9543	1.0000	1.0000	1.0000
Pituitary	Precision	0.8182	0.9401	0.9139	0.9800	1.0000	1.0000
	Recall	0.9900	0.9933	0.9900	0.9900	0.9900	1.0000
	F1-score	0.8959	0.9660	0.9504	0.9900	0.9900	1.0000

The per-class metric analysis for the Br35H dataset (Table 7) also shows that the WaveletFusion models provide clear improvements over the Baseline CNN in both classes. For the no class, the WaveletFusion 7-level model increased precision, recall, and F1-score, respectively, compared with the Baseline CNN model. A similar improvement was observed for the yes class, with the WaveletFusion 7-level model achieving stronger, more balanced classification than the Baseline CNN model. Although the WaveletFusion 8-level model also improved the results over the Baseline CNN, it remained consistently weaker than the 7-level configuration in both classes. Compared with the published models, the WaveletFusion 7-level model remained competitive, but the Post-BN CNN and Post-BN Res CNN models still achieved the highest per-class scores on this dataset.

Table 7. Overall per-class test metrics results of the Baseline CNN and WaveletFusion models for Br35H dataset.

Class	Metric	Baseline CNN	WaveletFusion 7-Level	WaveletFusion 8-Level	Pre-BN CNN [49]	Post-BN CNN [49]	Post-BN Res CNN [49]
no	Precision	0.9203	0.9707	0.9243	0.9896	0.9966	0.9966
	Recall	0.7700	0.9933	0.9767	0.9661	0.9820	0.9856
	F1-score	0.8385	0.9819	0.9498	0.9777	0.9915	0.9915
yes	Precision	0.8029	0.9932	0.9754	0.9679	0.9870	0.9870
	Recall	0.9336	0.9701	0.9202	0.9902	0.9967	0.9967
	F1-score	0.8633	0.9815	0.9470	0.9789	0.9918	0.9918

The confusion matrix results (Figure 12) show that the Baseline CNN produces more misclassifications than the proposed WaveletFusion models. This is significantly visible in the meningioma class, where misclassification into other tumor categories becomes noticeably less frequent. Improvements are also observed in glioma, where the WaveletFusion models produce fewer errors in neighboring classes. For the no-tumor and pituitary cases, all models already perform strongly, but the WaveletFusion variants, 7-level configurations, show fewer misclassifications.

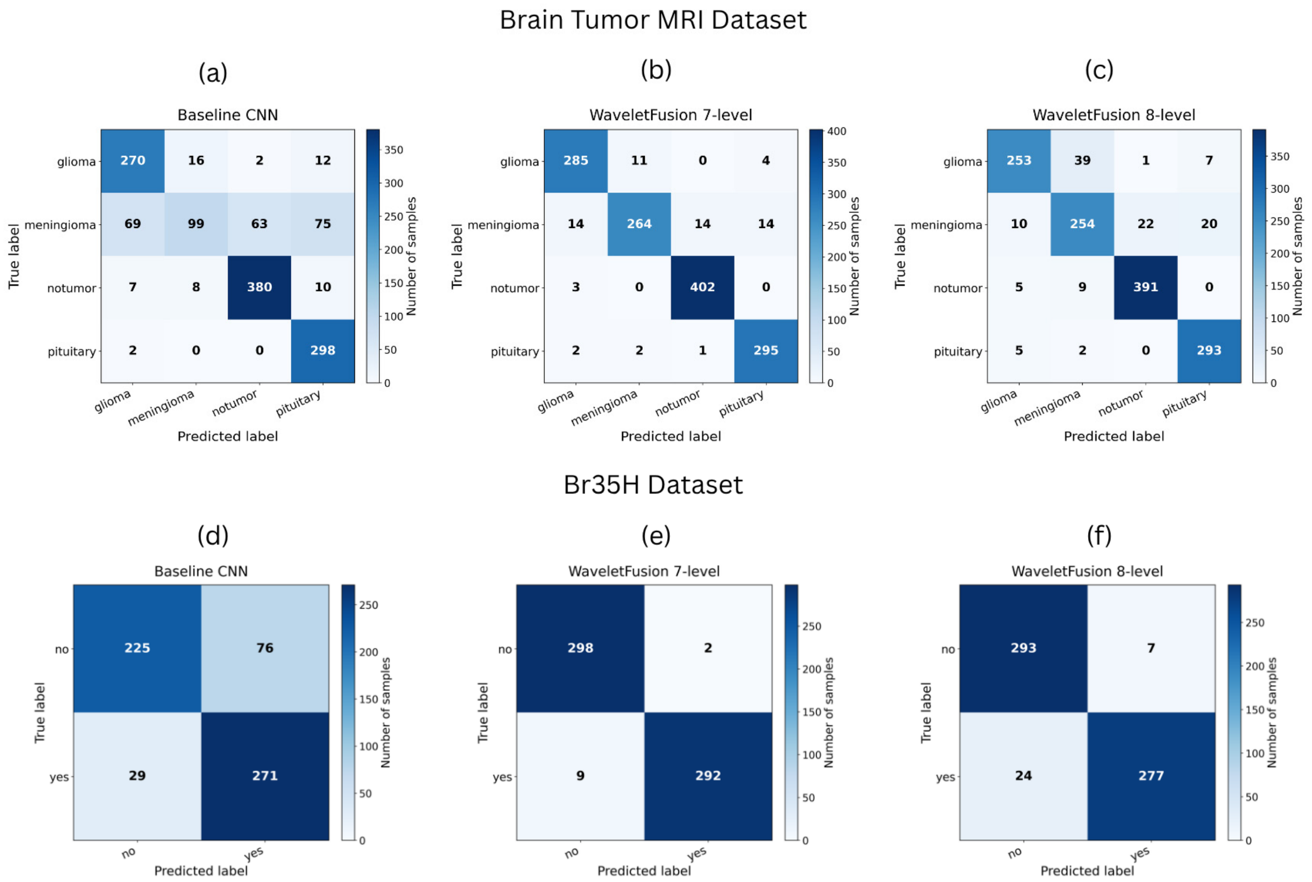

Figure 12. Test-set confusion matrices for the Brain Tumor MRI dataset: (**a**) Baseline CNN, (**b**) WaveletFusion 7-level, and (**c**) WaveletFusion 8-level; and for the Br35H dataset: (**d**) Baseline CNN, (**e**) WaveletFusion 7-level, and (**f**) WaveletFusion 8-level.

The unified training accuracy and training loss (Figure 13) show that WaveletFusion models improve more rapidly during the first epochs than the Baseline CNN model, indicating faster feature extraction. This trend is visible in both the Brain Tumor MRI and Br35H datasets. The relatively narrow variability bands after the early epochs further suggest that the WaveletFusion models learn more reproducibly across runs, which indicates that embedding multi-scale wavelet features improves optimization stability and helps the network converge toward stronger discriminative representations earlier in training. The corresponding training loss curves also show the same tendency, where the WaveletFusion models reduce loss more rapidly and follow a smoother trajectory than the Baseline CNN.

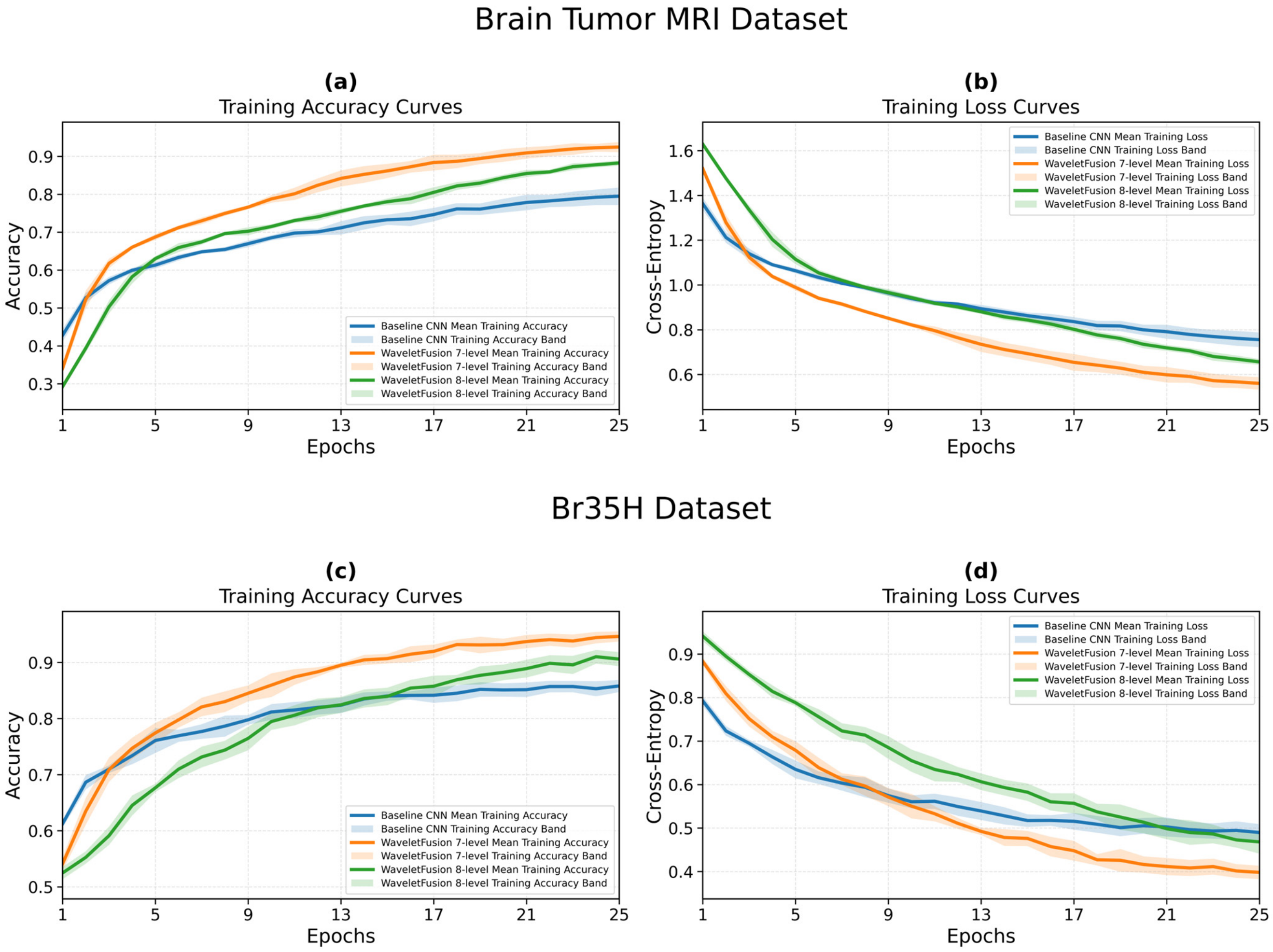

Figure 13. Unified training accuracy and training cross-entropy loss curves across 5 runs for the Baseline CNN, WaveletFusion 7-level, and WaveletFusion 8-level models on the Brain Tumor MRI dataset: (**a**) training accuracy, (**b**) training loss; and on the Br35H dataset: (**c**) training accuracy, (**d**) training loss.

Similarly, the unified validation accuracy and validation cross-entropy loss curves (Figure 14) show the same trend, where the Baseline CNN exhibits slower convergence and higher instability during the early epochs. Both WaveletFusion models reduce loss more rapidly and then follow a smoother trajectory, while the Baseline CNN model starts with a higher loss and shows a slower decline. However, despite the effectiveness of both WaveletFusion models, the WaveletFusion 7-level model achieves the lowest final loss, which is compatible with its stronger test and validation performance. This same overall tendency is observed in both the Brain Tumor MRI and Br35H datasets. The reduced fluctuation of the WaveletFusion curves also suggests that wavelet-based multi-scale representations improve gradient flow and help the model fit relevant MRI structure without the same degree of instability seen in the baseline.

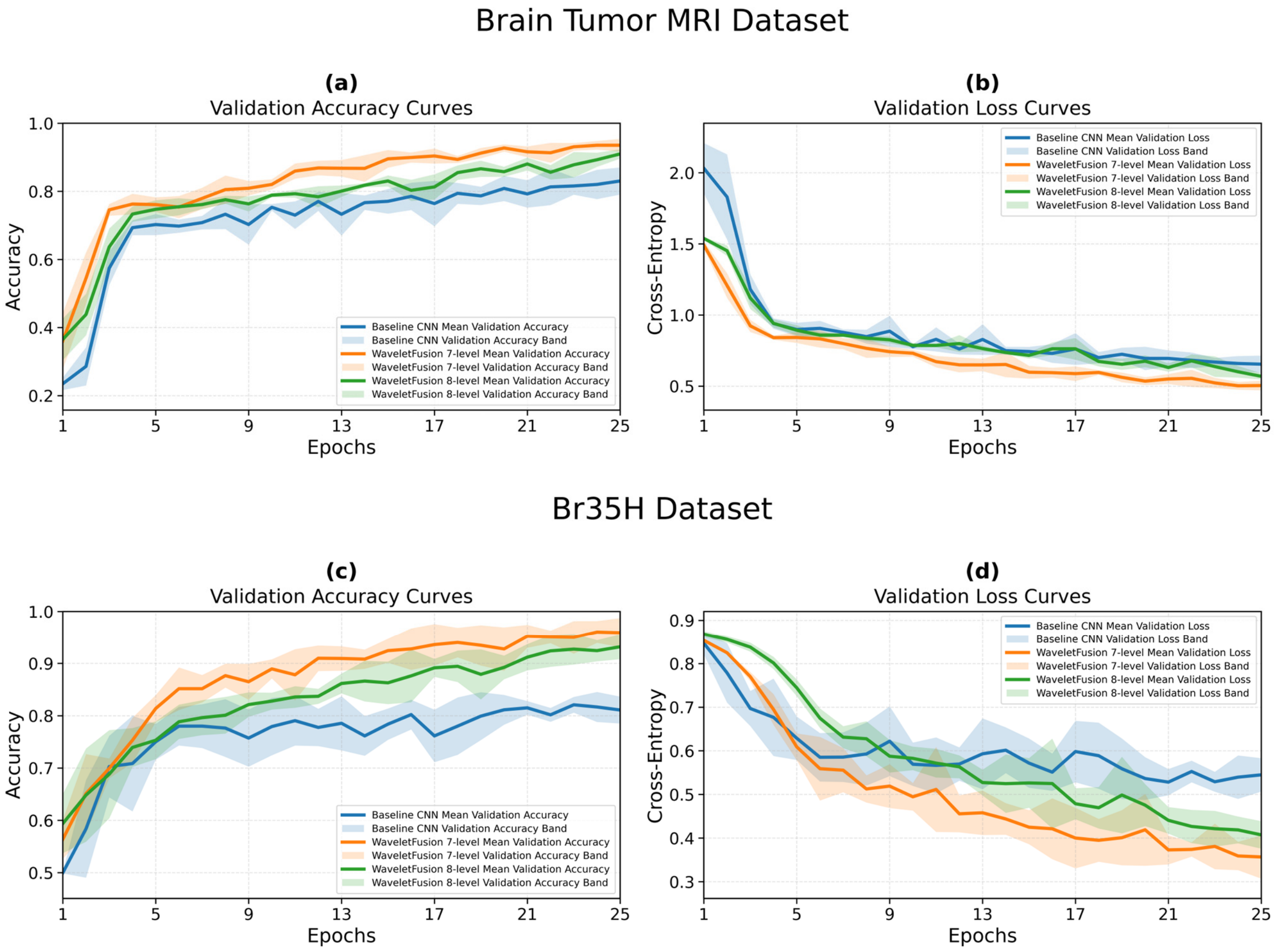

Figure 14. Unified validation accuracy and validation cross-entropy loss curves across 5 runs for the Baseline CNN, WaveletFusion 7-level, and WaveletFusion 8-level models on the Brain Tumor MRI dataset: (**a**) validation accuracy, (**b**) validation loss; and on the Br35H dataset: (**c**) validation accuracy, (**d**) validation loss.

4. Discussion

This research shows that, without increasing model size, integrating a multi-level Haar wavelet decomposition into the convolutional neural network improves classification results on Brain Tumor MRI by preserving fine-detail cues that spatial-only networks might lose compared to the standard Baseline CNN [33,34]. Notably, the proposed model achieved these results with a lower parameter count than the Baseline CNN, suggesting that the accuracy improvements stem from richer feature encoding. This is also supported by the decomposition-depth analysis, where the WaveletFusion 7-level configuration achieved the best overall results compared to the WaveletFusion 8-level, indicating that excessive decomposition can remove useful spatial information.

One of the advantages of the proposed approach is that it embeds frequency-domain cues in a computationally efficient way, increasing generalization under noise or limited training data. The datasets used in the research are characterized by the fact that the relevant information in the grayscale image is contained in a single area surrounded by a black background. Thus, using Haar wavelet decomposition aims to detect the patterns inside the information-rich area rather than capturing the boundaries of the distinct objects. Instead of learning all the image features during the training, the model processes meaningful features representing the vertical, horizontal, and diagonal details of different frequency. It leads to capturing the relationship between repeating patterns at different decomposition levels and is particularly useful with limited training data. Our findings also show that the wavelet-fusion network improved performance across all tumor types. This was supported by previous research, which showed that integrating wavelet transforms with CNNs can reduce overfitting and make models more robust to inter-scanner variability [19,30,33]. In addition, the subband ablation showed that LH and HL components contributed more strongly than HH, suggesting that horizontal and vertical directional details carry more useful information.

The per-class analysis also showed that the WaveletFusion 7-level model produced the best results across all four classes. Among them, meningioma remained the most challenging category. However, it also showed one of the clearest improvements in the 7-level configuration, suggesting that the selection of decomposition depth is especially important for classes with more uncertain textural borders and stronger overlap with adjacent tumor types.

On the other hand, wavelet decomposition incurs extra computation that increases training time, as shown in the results. This trade-off is common in multi-branch networks [24] and suggests further work on the model's architecture. Another weakness, as in many MRI studies, is that the evaluation was limited to 2D image-level analysis on publicly available datasets, which may not fully capture patient-level variability. Extending wavelet fusion to multi-sequence MRI could further reduce inter-class confusion, and an explicit study of scale depth may help identify the minimal decomposition level that retains most of the benefit while reducing computational overhead.

The article demonstrates an approach applied to brain tumor classification. Combined with the explanations generated using XAI, the results can be used by radiologists as a recommendation tool in making diagnosis. Although this research was limited to brain datasets, the approach can be applied to other medical problems as well.

Future studies should further validate the proposed architecture by using data from multiple clinics. In addition, multi-sequence MRI may help the model exploit textural information not fully captured by single-slice analysis. Moreover, using higher order Daubechies wavelet could improve the feature extraction stage and therefore reduce the number of decompositions needed to achieve high accuracy. Lastly, future development should explore lightweight fusion strategies and adaptive decomposition schemes to balance classification performance and computational efficiency.

5. Conclusions

This paper introduced a hybrid WaveletFusion model that combines a multi-level Haar wavelet decomposition with convolutional neural networks for Brain Tumor MRI classification. The proposed approach was designed to improve the representation of MRI data by explicitly integrating approximation and directional detail information across multiple spatial scales. In this way, the model captures both global structural context and fine-grained textural variations that are important for differentiating among brain tumor classes.

The experimental results showed that the WaveletFusion model provides a clear advantage over the standard Baseline CNN under the same training and evaluation protocol. Among the evaluated configurations, the WaveletFusion 7-level model achieved the best overall performance, reaching the strongest test accuracy and macro-F1, while also using fewer parameters than the baseline model. The 8-level model performed worse than the 7-level model, indicating that excessive decomposition can reduce useful spatial information.

The additional ablation and explainability analyses further clarified how the proposed architecture reaches its predictions, and it showed that the LH and HL directional detail subbands contributed more strongly than HH, and the most useful information was concentrated in the early-to-intermediate decomposition scales, while deeper levels provided important contextual support for the final decision. Grad-CAM analysis further indicates that the model generally focuses on meaningful intracranial regions and uses deeper decomposition features in a structured way, which supports the idea that integrating wavelet-domain information directly into the CNN pipeline enhances both predictive performance and interpretability.

The findings indicate that the proposed architecture remains a promising direction in neuroimaging analysis. However, more validation on larger datasets is necessary, and future extensions toward multi-sequence MRI integration may further enhance its applicability in this area.

Author Contributions: Conceptualization, D.Č.; methodology, M.I. and D.Č.; software, M.I.; validation, M.I.; formal analysis, M.I. and D.Č.; investigation, M.I.; writing—original draft preparation, M.I.; writing—review and editing, D.Č.; visualization, M.I. and D.Č.; supervision, D.Č. All authors have read and agreed to the published version of the manuscript.

Funding: This Research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The experiments in this study were conducted using the publicly available datasets: Brain Tumor MRI Dataset can be accessed via Kaggle at the following link: https://doi.org/10.34740/KAGGLE/DSV/14832123 (accessed on 15 January 2026) and Br35H :: Brain Tumor Detection 2020 can be accessed via Kaggle at the following link: https://www.kaggle.com/datasets/ahmedhamada0/brain-tumor-detection/data?select=Br35H-Mask-RCNN (accessed on 16 March 2026). No additional datasets were generated or analyzed during the current study.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial intelligence
CLAIM	Checklist for Artificial Intelligence in Medical Imaging
CNN	Convolutional neural network
DWT	Discrete wavelet transform
FPS	Frames per second
Grad-CAM	Gradient-weighted Class Activation Mapping
HH	High–high detail subband
HL	High–low detail subband
LH	Low–high detail subband
LL	Low–low approximation subband
MRI	Magnetic resonance imaging
ReLU	Rectified linear unit
STARD-AI	Standards for Reporting Diagnostic Accuracy Studies–Artificial Intelligence
XAI	Explainable Artificial Intelligence
Br35H	Br35H :: Brain Tumor Detection 2020

References

1. Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; van der Laak, J.A.W.M.; van Ginneken, B.; Sánchez, C.I. A survey on deep learning in medical image analysis. *Med. Image Anal.* **2017**, *42*, 60–88. https://doi.org/10.1016/j.media.2017.07.005.
2. Shen, D.; Wu, G.; Suk, H.-I. Deep learning in medical image analysis. *Annu. Rev. Biomed. Eng.* **2017**, *19*, 221–248. https://doi.org/10.1146/annurev-bioeng-071516-044442.

3. Lundervold, A.S.; Lundervold, A. An overview of deep learning in medical imaging focusing on MRI. *Z. Med. Phys.* **2019**, *29*, 102–127. https://doi.org/10.1016/j.zemedi.2018.11.002.
4. Aggarwal, R.; Sounderajah, V.; Martin, G.; Ting, D.S.W.; Karthikesalingam, A.; King, D.; Ashrafian, H.; Darzi, A. Diagnostic accuracy of deep learning in medical imaging: A systematic review and meta-analysis. *npj Digit. Med.* **2021**, *4*, 65. https://doi.org/10.1038/s41746-021-00438-z.
5. Mienye, I.D.; Swart, T.G.; Obaido, G.; Jordan, M.; Ilono, P. Deep convolutional neural networks in medical image analysis: A review. *Information* **2025**, *16*, 195. https://doi.org/10.3390/info16030195.
6. Wesseling, P. INSP-02. WHO 2021 classification of CNS tumors. *Neuro-Oncology* **2022**, *24*, i186. https://doi.org/10.1093/neuonc/noac079.698.
7. Amarnath, A.; Al Bataineh, A.; Hansen, J.A. Transfer-learning approach for enhanced brain tumor classification in MRI imaging. *BioMedInformatics* **2024**, *4*, 1745–1756. https://doi.org/10.3390/biomedinformatics4030095.
8. Khan, H.A.; Jue, W.; Mushtaq, M.; Mushtaq, M.U. Brain tumor classification in MRI image using convolutional neural network. *Math. Biosci. Eng.* **2020**, *17*, 6203–6216. https://doi.org/10.3934/mbe.2020328.
9. Rizwan, M.; Shabbir, A.; Javed, A.R.; Shabbir, M.; Baker, T.; Obe, D.A.-J. Brain tumor and glioma grade classification using gaussian convolutional neural network. *IEEE Access* **2022**, *10*, 29731–29740. https://doi.org/10.1109/ACCESS.2022.3153108.
10. Li, C.; Zhang, F.; Du, Y.; Li, H. Classification of brain tumor types through MRIs using parallel CNNs and firefly optimization. *Sci. Rep.* **2024**, *14*, 15057. https://doi.org/10.1038/s41598-024-65714-w.
11. Srinivasan, S.; Francis, D.; Mathivanan, S.K.; Rajadurai, H.; Shivahare, B.D.; Shah, M.A. A hybrid deep CNN model for brain tumor image multi-classification. *BMC Med. Imaging* **2024**, *24*, 21. https://doi.org/10.1186/s12880-024-01195-7.
12. Disci, R.; Gurcan, F.; Soylu, A. Advanced brain tumor classification in MR images using transfer learning and pre-trained deep CNN models. *Cancers* **2025**, *17*, 121. https://doi.org/10.3390/cancers17010121.
13. Gorenshtein, A.; Liba, T.; Goren, A. Lightweight transfer learning models for multi-class brain tumor classification: Glioma, meningioma, pituitary tumors, and no tumor MRI screening. *J. Imaging Inform. Med.* 2025, *in press*. https://doi.org/10.1007/s10278-025-01686-1.
14. Yagis, E.; Atnafu, S.W.; de Herrera, A.G.S.; Marzi, C.; Scheda, R.; Giannelli, M.; Tessa, C.; Citi, L.; Diciotti, S. Effect of data leakage in brain MRI classification using 2D convolutional neural networks. *Sci. Rep.* **2021**, *11*, 22544. https://doi.org/10.1038/s41598-021-01681-w.
15. Mongan, J.; Moy, L.; Kahn, C.E. Checklist for artificial intelligence in medical imaging (CLAIM): A guide for authors and reviewers. *Radiol. Artif. Intell.* **2020**, *2*, e200029. https://doi.org/10.1148/ryai.2020200029.
16. Sounderajah, V.; Guni, A.; Liu, X.; Collins, G.S.; Karthikesalingam, A.; Markar, S.R.; Golub, R.M.; Denniston, A.K.; Shetty, S.; Moher, D.; et al. The STARD-AI reporting guideline for diagnostic accuracy studies using artificial intelligence. *Nat. Med.* 2025, *in press*. https://doi.org/10.1038/s41591-025-03953-8.
17. Huda, N.; Ku-Mahamud, K.R. CNN-based image segmentation approach in brain tumor classification: A review. *Eng. Proc.* **2025**, *84*, 66. https://doi.org/10.3390/engproc2025084066.
18. Ottoni, M.; Kasperczuk, A.; Tavora, L.M.N. Machine learning in MRI brain imaging: A review of methods, challenges, and future directions. *Diagnostics* **2025**, *15*, 2692. https://doi.org/10.3390/diagnostics15212692.
19. Mallat, S. *A Wavelet Tour of Signal Processing*, 3rd ed.; Elsevier: Burlington, MA, USA, 2009.
20. Kutlu, H.; Avcı, E. A novel method for classifying liver and brain tumors using convolutional neural networks, discrete wavelet transform and long short-term memory networks. *Sensors* **2019**, *19*, 1992. https://doi.org/10.3390/s19091992.
21. Fayaz, M.; Torokeldiev, N.; Turdumamatov, S.; Qureshi, M.S.; Qureshi, M.B.; Gwak, J. An efficient methodology for brain MRI classification based on DWT and convolutional neural network. *Sensors* **2021**, *21*, 7480. https://doi.org/10.3390/s21227480.
22. Rhyou, S.-Y.; Yoo, J.-C. Automated ultrasonography of hepatocellular carcinoma using discrete wavelet transform based deep-learning neural network. *Med. Image Anal.* **2025**, *101*, 103453. https://doi.org/10.1016/j.media.2025.103453.
23. Pandey, A.; Pandey, V.K. Wavelet based classification using meta-heuristic algorithm with deep transfer learning technique. *SN Comput. Sci.* **2025**, *6*, 208. https://doi.org/10.1007/s42979-025-03683-1.
24. Wu, J.; Li, J.; Yang, J.; Mei, S. Wavelet-integrated deep neural networks: A systematic review of applications and synergistic architectures. *Neurocomputing* **2025**, *657*, 131648. https://doi.org/10.1016/j.neucom.2025.131648.
25. El Kader, I.A.; Xu, G.; Shuai, Z.; Saminu, S.; Javaid, I.; Ahmad, I.S.; Kamhi, S. Brain tumor detection and classification on MR images by a deep wavelet auto-encoder model. *Diagnostics* **2021**, *11*, 1589. https://doi.org/10.3390/diagnostics11091589.

26. Aboulmira, A.; Hrimech, H.; Lachgar, M.; Hanine, M.; Garcia, C.O.; Mezquita, G.M.; Ashraf, I. Hybrid model with wavelet decomposition and EfficientNet for accurate skin cancer classification. *J. Cancer* **2025**, *16*, 506–520. https://doi.org/10.7150/jca.101574.
27. Liu, P.; Zhang, H.; Lian, W.; Zuo, W. Multi-level wavelet convolutional neural networks. *IEEE Access* **2019**, *7*, 74973–74985. https://doi.org/10.1109/ACCESS.2019.2921451.
28. Zhao, X.; Huang, P.; Shu, X. Wavelet-attention CNN for image classification. *Multimed. Syst.* 2022, *in press*. https://doi.org/10.1007/s00530-022-00889-8.
29. Brito, A.d.S.; Vieira, M.B.; de Andrade, M.L.S.C.; Feitosa, R.Q.; Giraldi, G.A. Combining max-pooling and wavelet pooling strategies for semantic image segmentation. *Expert Syst. Appl.* **2021**, *183*, 115403. https://doi.org/10.1016/j.eswa.2021.115403.
30. Savareh, B.A.; Emami, H.; Hajiabadi, M.; Azimi, S.M.; Ghafoori, M. Wavelet-enhanced convolutional neural network: A new idea in a deep learning paradigm. *Biomed. Eng./Biomed. Tech.* **2019**, *64*, 195–205. https://doi.org/10.1515/bmt-2017-0178.
31. Shati, A.; Hassan, G.M.; Datta, A. A comprehensive fusion model for improved pneumonia prediction based on KNN-wavelet-GLCM and a residual network. *Intell. Syst. Appl.* **2025**, *26*, 200492. https://doi.org/10.1016/j.iswa.2025.200492.
32. Tran, T.K.; Kim, S.-H.; Yang, H.-J.; Lee, M. Multi-stage wavelet-attention deep learning for brain MRI classification. *Multimed. Syst.* **2026**, *32*, 176. https://doi.org/10.1007/s00530-026-02237-6.
33. Mewada, H. 2D-wavelet encoded deep CNN for image-based ECG classification. *Multimed. Tools Appl.* 2023, *in press*. https://doi.org/10.1007/s11042-022-14302-z.
34. Mewada, H.K.; Patel, A.V.; Chaudhari, J.; Mahant, K.; Vala, A. Wavelet features embedded convolutional neural network for multiscale ear recognition. *J. Electron. Imaging* **2020**, *29*, 043029. https://doi.org/10.1117/1.JEI.29.4.043029.
35. Mohamed Musthafa, M.; Mahesh, T.R.; Vinoth Kumar, V.; Guluwadi, S. Enhancing brain tumor detection in MRI images through explainable AI using Grad-CAM with Resnet 50. *BMC Med. Imaging* **2024**, *24*. https://doi.org/10.1186/s12880-024-01292-7.
36. van der Velden, B.H.; Kuijf, H.J.; Gilhuijs, K.G.; Viergever, M.A. Explainable artificial intelligence (XAI) in deep learning-based medical image analysis. *Med. Image Anal.* **2022**, *79*, 102470. https://doi.org/10.1016/j.media.2022.102470.
37. Jin, W.; Li, X.; Fatehi, M.; Hamarneh, G. Guidelines and evaluation of clinical explainable AI in medical image analysis. *Med. Image Anal.* **2022**, *84*, 102684. https://doi.org/10.1016/j.media.2022.102684.
38. Neena, K.A.; Kumar, A. Haar-initialized parametric wavelet compression with attention-driven lightweight CNN for brain tumor classification on edge devices. *Biomed. Phys. Eng. Express* 2026, *in press*. https://doi.org/10.1088/2057-1976/ae3760.
39. Çınarer, G.; Emiroğlu, B.G.; Yurttakal, A.H. Prediction of glioma grades using deep learning with wavelet radiomic features. *Appl. Sci.* **2020**, *10*, 6296. https://doi.org/10.3390/app10186296.
40. Nickparvar, M. *Brain Tumor MRI Dataset*; Kaggle: San Francisco, CA, USA, 2026. https://doi.org/10.34740/KAGGLE/DSV/14832123.
41. Krolik, J.; Lynn, J.; Rudden, J.H.; Vremenko, D. MRI brain tumor detection with computer vision. *arXiv* **2025**, arXiv:2510.10250. Available online: https://arxiv.org/abs/2510.10250 (accessed on 15 January 2026).
42. Gómez-Guzmán, M.A.; Jiménez-Beristaín, L.; García-Guerrero, E.E.; López-Bonilla, O.R.; Tamayo-Perez, U.J.; Esqueda-Elizondo, J.J.; Palomino-Vizcaino, K.; Inzunza-González, E. Classifying brain tumors on magnetic resonance imaging by using convolutional neural networks. *Electronics* **2023**, *12*, 955. https://doi.org/10.3390/electronics12040955.
43. Al-Adwan, A. Evaluating the effectiveness of brain tumor image generation using generative adversarial network with adam optimizer. *Int. J. Adv. Comput. Sci. Appl.* **2024**, *15, 512–521*. https://doi.org/10.14569/IJACSA.2024.0150653.
44. Hamada, A. *Br35H :: Brain Tumor Detection 2020*; Kaggle: San Francisco, CA, USA, 2021. Available online: https://www.kaggle.com/datasets/ahmedhamada0/brain-tumor-detection/data?select=Br35H-Mask-RCNN (accessed on 16 March 2026).
45. Dhingra, M.; Dhingra, V. Assessing machine learning model efficacy for brain tumor MRI classification: A multi-model approach. *J. Emerg. Investig.* **2026**, *9*. https://doi.org/10.59720/25-127.
46. Hekmat, A.; Zhang, Z.; Khan, S.U.R.; Bilal, O. Brain tumor diagnosis redefined: Leveraging image fusion for MRI enhancement classification. *Biomed. Signal Process. Control* **2025**, *109*, 108040. https://doi.org/10.1016/j.bspc.2025.108040.
47. Eleyan, G.; Al-Barakeh, Z.; Ghandour, R.; Neji, B.; Eleyan, A. Brain tumor detection via ensemble CNN-based deep learning models. In *Proceedings of the 2025 6th International Conference on Bio-engineering for Smart Technologies (BioSMART)*; IEEE: New York, NY, USA, 2025. https://doi.org/10.1109/BIOSMART66413.2025.11046093.

48. Nizamli, Y.; Filatov, A.; Fadel, W.; Shichkina, Y.; Mreish, K. A lightweight CNN architecture for efficient brain tumor detection in MRI scans. *Int. J. Electr. Electron. Res.* **2025**, *13*, 296–305. https://doi.org/10.37391/ijeer.130213.
49. Ullah, A.; Akhtar, N.; Hameed, S.; Sajid, H.U.; Noreen, H.; Hasnain, M. An explainable deep learning framework for brain tumor detection using MRI images. *VFAST Trans. Softw. Eng.* **2026**, *14*, 278–300. https://doi.org/10.21015/vtse.v14i1.2335.