Kaunas University of Technology
Faculty of Informatics

Research of Image Classification Based on the Interaction of Wavelet Transforms and Artificial Neural Networks

Master's Final Degree Project

Mahammad Ismayilov
Project author

Prof. Dr. Dalia Čalnerytė
Supervisor

Kaunas, 2026

Kaunas University of Technology

Faculty of Informatics

Research of Image Classification Based on the Interaction of Wavelet Transforms and Artificial Neural Networks

Master's Final Degree Project

Artificial Intelligence in Computer Science (6211BX007)

Mahammad Ismayilov

Project author

Prof. Dr. Dalia Čalnerytė

Supervisor

Prof. Dr. Andrius Kriščiūnas

Reviewer

Kaunas, 2026

Kaunas University of Technology

Faculty of Informatics

Mahammad Ismayilov

Research of Image Classification Based on the Interaction of Wavelet Transforms and Artificial Neural Networks

Declaration of Academic Integrity

I confirm the following:

1. I have prepared the final degree project independently and honestly without any violations of the copyrights or other rights of others, following the provisions of the Law on Copyrights and Related Rights of the Republic of Lithuania, the Regulations on the Management and Transfer of Intellectual Property of Kaunas University of Technology (hereinafter – University) and the ethical requirements stipulated by the Code of Academic Ethics of the University;

2. All the data and research results provided in the final degree project are correct and obtained legally; none of the parts of this project are plagiarised from any printed or electronic sources; all the quotations and references provided in the text of the final degree project are indicated in the list of references;

3. I have not paid anyone any monetary funds for the final degree project or the parts thereof unless required by the law;

4. I understand that in the case of any discovery of the fact of dishonesty or violation of any rights of others, the academic penalties will be imposed on me under the procedure applied at the University; I will be expelled from the University and my final degree project can be submitted to the Office of the Ombudsperson for Academic Ethics and Procedures in the examination of a possible violation of academic ethics.

Mahammad Ismayilov

Confirmed electronically

Kaunas University of Technology

Faculty of Informatics

Task of the Master's final degree project

Topic of the project	Research of Image Classification Based on the Interaction of Wavelet Transforms and Artificial Neural Networks
Requirements and conditions	
Supervisor	
	(position, name, surname, signature of the supervisor) (date)

Ismayilov Mahammad. Research of Image Classification Based on the Interaction of Wavelet Transforms and Artificial Neural Networks. Master's Final Degree Project / supervisor prof. dr. Dalia Čalnerytė; Faculty of Informatics, Kaunas University of Technology.

Study field and area (study field group): Computer science, Informatics (B01).

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

Kaunas, 2026. 61 p.

Summary

In medical image analysis, image classification is an important task in the field of artificial intelligence, where small visual details can influence diagnostic decisions. Therefore, convolutional neural network models have shown strong performance in this field, but they may still be sensitive to noise and to the loss of important details during downsampling.

The main goal of this research is to develop and evaluate a proposed WaveletFusion architecture that integrates multi-level Haar wavelet decomposition with convolutional neural network-based feature extraction. The model uses approximation and directional detail subbands to preserve both structural and texture-related information from brain MRI images.

The WaveletFusion architecture was evaluated on two datasets that contain brain tumor MRI images: the multi-class Brain Tumor MRI and Br35H :: Brain Tumor Detection 2020 datasets. Several experimental and ablation analyses were performed, including a Grad-CAM explainability analysis to better understand the model‘s performance.

The results show that the WaveletFusion 7-level model achieved the best overall performance. It improved classification results, reaching 0.90 of test accuracy and macro-F1 on the Brain Tumor MRI dataset and 0.96 of test accuracy and macro-F1 on the Br35H :: Brain Tumor Detection 2020 dataset. The results demonstrate that combining multi-level wavelet decomposition with convolutional neural network learning can improve classification performance and support more informative feature representation for brain MRI image classification.

Ismayilov Mahammad. Bangelių transformacijų ir dirbtinių neuroninių tinklų sąveika paremto vaizdų klasifikavimo tyrimas. Magistro baigiamasis projektas / vadovė prof. dr. Dalia Čalnerytė; Kauno technologijos universitetas, Informatikos fakultetas.

Studijų kryptis ir sritis (studijų krypčių grupė): Informatikos mokslai, Informatika (B01).

Reikšminiai žodžiai: konvoliucinis neuroninis tinklas (CNN); Haro bangelių dekompozicija; vaizdų klasifikavimas, daugiapakopis bangelių sujungimas; paaiškinamasis dirbtinis intelektas (XAI).

Kaunas, 2026. 61 p.

Santrauka

Medicininės vaizdų analizės srityje vaizdų klasifikavimas yra svarbi dirbtinio intelekto užduotis, kur net mažiausios vizualinės detalės gali turėti įtakos diagnostiniams sprendimams. Todėl konvoliucinių neuroninių tinklų modeliai šioje srityje pasirodė esą labai veiksmingi, tačiau jie vis dar gali būti jautrūs triukšmui ir svarbių detalių praradimui mažinant skiriamąją gebą.

Pagrindinis šio tyrimo tikslas – sukurti ir įvertinti siūlomą „WaveletFusion“ architektūrą, kuri integruoja daugiapakopę Haro bangelių dekompoziciją su CNN pagrįstu požymių išskyrimu. Modelis naudoja aproksimacijos ir kryptingų detalių subjuostas, kad išsaugotų tiek struktūrinę, tiek su tekstūra susijusią informaciją iš smegenų MRT vaizdų.

„WaveletFusion“ architektūra buvo įvertinta naudojant du duomenų rinkinius, kuriuose yra smegenų auglių MRT vaizdai: tai yra daugiaklasis smegenų auglių MRT duomenų rinkinys ir binarinis „Br35H“ duomenų rinkinys. Buvo atliktos kelios eksperimentinės ir abliacinės analizės, įskaitant „Grad-CAM“ paaiškinamumo analizę, siekiant geriau suprasti modelio veikimą.

Rezultatai rodo, kad „WaveletFusion“ 7 lygių modelis pasiekė geriausią bendrą našumą. Jis pagerino klasifikavimo rezultatus, pasiekdamas 0,90 bandymo tikslumą ir makro-F1 smegenų auglių MRT duomenų rinkinyje bei 0,96 bandymo tikslumą ir makro-F1 „Br35H“ duomenų rinkinyje. Rezultatai rodo, kad daugialypio lygio bangų dekompozicijos ir CNN mokymosi derinys gali pagerinti klasifikavimo našumą ir užtikrinti informatyvesnį požymių atvaizdavimą smegenų MRT vaizdų klasifikavimui.

Table of contents

List of figures

List of tables

List of abbreviations and terms

Abbreviations:

AI – Artificial Intelligence;

ANN – Artificial Neural Network;

Br35H – Br35H :: Brain Tumor Detection 2020 dataset;

CAM – Class Activation Mapping;

CLAIM – Checklist for Artificial Intelligence in Medical Imaging;

CNN – Convolutional Neural Network;

CPU – Central Processing Unit;

CUDA – Compute Unified Device Architecture;

DWT – Discrete Wavelet Transform;

FLOPs – Floating Point Operations;

FPS – Frames per Second;

GPU – Graphics Processing Unit;

HOG – Histogram of Oriented Gradients;

kNN – k-Nearest Neighbor;

MRI – Magnetic Resonance Imaging;

RAM – Random Access Memory;

ReLU – Rectified Linear Unit;

SHAP – Shapley Additive exPlanations;

Grad-CAM – Gradient-weighted Class Activation Mapping;

SIFT – Scale-Invariant Feature Transform;

SVM – Support Vector Machine;

SWT – Stationary Wavelet Transform;

XAI – Explainable Artificial Intelligence.

Terms:

Ablation study – an experimental analysis where one model setting is changed at a time to evaluate its effect on performance.

Introduction

This master‘s final project thesis paper belongs to an Artificial Intelligence in Computer Science student majoring in the Data Analytics study modules.

The image analysis has become one of the technological achievements in recent years, especially with the approach of machine learning techniques. The application of neural network models, especially with convolutional neural network models, has become one of the crucial tools for investigating a wide range of complex images for tasks like object detection, classification, and anomaly detection across different study fields. However, despite their potential impact on such complex analysis, the performance of these neural network models may be limited when there is noise in the images provided and significant computational requirements.

To address these challenges, researchers have found that the neural network models and wavelet transform methods can be combined, making them more efficient at decomposing the images into different frequency components and allowing them to capture texture and structural information across multiple scales. Such a technique, like the isolation of key features, makes the wavelet transform one of the possible preprocessing steps because transforming the image in such a way can increase the neural network models‘ performance to identify important patterns even with noise. Nevertheless, existing methods still need to improve their performance in handling complex image resolution or quality variations.

Project novelty and relevance

The relevance of this work is related to the high importance of accurate image classification methods, especially in medical image analysis, where small visual details can be important for decision-making. This research focuses on integrating multi-level Haar wavelet decomposition with neural network feature extraction.

The novelty of the project is the proposed WaveletFusion architecture, which combines CNN learning with wavelet-based multi-scale image representation. Unlike the traditional convolutional neural network models, the proposed architecture uses both approximation and detail wavelet subbands from several decomposition levels, allowing the model to learn information from different spatial and frequency scales.

Aim and objectives

The main aim of this research project is to design, implement, and evaluate the proposed WaveletFusion architecture, which integrates multi-level Haar wavelet decomposition with CNN-based feature extraction for brain tumor MRI classification.

The following goals are carefully represented to accomplish this overarching goal:
1. Review existing analysis of wavelet transforms, artificial neural networks, and their combined use in image analysis.
2. Design a new image analysis method that integrates wavelet transforms with artificial neural networks based on the limitations identified in previous research.
3. Implement the proposed method and evaluate its performance on selected brain tumor MRI datasets.

4. Compare the obtained results with baseline and existing methods to determine classification performance.

Document structure

The thesis paper consists of three chapters. Each of which describes specific features of the research topic. Beginning with a literature analysis of previous works, the first chapter explores the challenges in image analysis, especially focusing on the interaction between wavelet transforms and artificial neural networks. The second chapter proceeds to the selected solution's specifications by describing the non-functional requirements, quality criteria, and selected libraries and specifications for the test environment. The experimental evaluation of the proposed architecture, including the experimental setup, training configuration, ablation studies, explainability analysis, and comparison of existing models, is presented in the third chapter. Lastly, the thesis paper summarizes the obtained results together with the objectives in the conclusions chapter.

Use of artificial intelligence tools

During the preparation of the master's thesis, language support tools were used only to improve the clarity and correctness of the text. Grammarly was used to check grammar and sentence structure. ChatGPT was used as an assistance tool to fix bugs related to code, organize the drafted material, improve the language, and clarify the content provided. These tools were not used to develop the main research idea, to propose the architecture, to generate experimental results, or to replace the author's own scientific work. The final content of the document remains the author's responsibility.

1. Analysis of Wavelet-Based CNN Methods for Image Classification

The chapter analyzes the main techniques and methodologies relevant to image analysis based on the interaction of artificial neural networks and wavelet transforms. It reviews traditional image processing techniques and deep learning approaches for image analysis. Then, it introduces the main principles of wavelet transforms, convolutional neural networks, and their limitations, and essential CNN architectures relevant to image classification. This also examines the integration of wavelet transforms into convolutional neural network models for image analysis tasks. In addition, explainable artificial intelligence methods for model interpretation are discussed, including their strengths, limitations, and practical relevance in medical image classification. Finally, it summarizes the analysis results that form the basis for the WaveletFusion architecture.

1.1. Traditional Image Processing Techniques and Approaches

The image processing techniques are mostly applied to refine the visual quality of the raw images for human interpretation and transform image data into representations that enable segmentation or classification [1]. For example, segmentation is frequently described as one of the difficult image analysis tasks because it maps continuous image evidence into discrete region hypotheses. A review of segmentation techniques highlights a different number of strategies across clustering methods, edge and region detection, emphasizing that each approach responds differently to noise and variability [2]. For example, edge-based methods are commonly anchored in principled edge detection criteria, with a canonical formulation describing objective goals for detection and localization [3]. In addition, thresholding is one of the simplest approaches, because it converts a grayscale image into a binary image using a threshold value, but it may not provide appropriate results in complex environments [1, 2]. In contrast, edge detection methods focus on detecting intensity discontinuities that may correspond to object boundaries [3]. To summarize the main families of traditional image processing techniques and show their role in the image classification tasks, Table 1 is provided below.

Table 1. Main Families of Traditional Image Processing Approaches

Method	Examples	Main Purpose
Image enhancement and preprocessing [1]	General image preprocessing operations.	Improve image quality and prepare images for further analysis.
Segmentation methods [2, 3]	Edge detection, watershed segmentation.	Separate meaningful regions or object boundaries from the image.
Texture feature extraction [4]	Gray-level spatial dependence features.	Describe texture patterns and intensity relationships.
Local and gradient-based descriptors [5, 6]	SIFT, HOG.	Extract local shape, edge, and gradient information.
Distance-based learning [7]	k-nearest neighbor	Classify samples based on similarity in feature space.
Margin-based learning [8]	Support vector machines.	Create decision boundaries for feature-based classification.
Ensemble learning [9]	Random forests.	Combine multiple decision trees to improve robustness.

Feature extraction is one of the important techniques in machine learning that converts raw images into quantitative descriptors. It is suitable for classification, and previous work formalized features of gray-level spatial dependence and demonstrated their utility across multiple image types [4]. Local invariant features and gradient-based descriptors, such as SIFT and HOG, were also crucial because they enabled robust matching and detection pipelines [5, 6]. Besides other feature types, spectral features serve as a fundamental pattern element in image analysis, capturing tonal variations that enable the creation of unique signatures for supervised classification [1]. Texture, local, and color descriptors, therefore, represent different ways of converting visual information into numerical features that can be used by classical machine learning models [4, 5, 6].

The distance-based classification is another machine learning approach in which decisions are determined by proximity in feature space. The nearest neighbor decision theory establishes key relationships between asymptotic nearest neighbor error and the Bayes rate, which results in motivation for careful feature scaling and metric choices when applying kNN in high-dimensional image descriptor spaces. In image classification, this means that kNN depends on the quality of the extracted features for effectiveness, because the class decision is based on the similarity between samples in the selected feature space [7].

Margin-based learning has become a standard approach for image descriptors because it provides strong generalization in high-dimensional feature spaces. The formal support vector network was extended via kernels, which is particularly relevant to feature-engineered image pipelines for HOG and linear SVM detection [6, 8]. Since handcrafted descriptors may form complex feature spaces, kernel-based support vector machines are useful when a simple linear boundary is not sufficient for separating image classes [8].

Ensemble learning techniques, such as random forests, integrate randomized decision trees and characterize the relationship between generalization error and the strength of the trees and their correlations. This offers robustness to techniques such as random forests across many image classification tasks [9].

1.2. Deep Learning Approaches for Image Analysis

Deep learning has changed the approach to image analysis by replacing manual descriptor design with hierarchical representation learning, particularly through convolutional neural networks. A canonical synthesis describes deep learning as multi-layer representation learning that has driven major improvements in visual recognition of objects and regions and detection of faces and text [10]. The results on large labeled datasets demonstrated that deeper CNNs can significantly improve accuracy relative to prior approaches, motivating a widespread shift toward end-to-end learning [11]. In image analysis, this shift is important because the same model can learn feature extraction and classification together, instead of requiring a separate handcrafted feature extraction stage before learning [10, 11].

For the previously mentioned segmentation, encoder-decoder architectures with skip connections are widely adopted because they integrate contextual abstraction with spatially precise localization. A canonical biomedical segmentation paper describes an architecture shown in Fig. 1. below, with a contracting path for context and a symmetric expanding path for localization, with strong reliance on data augmentation to train effectively with limited annotations, where feature maps from the contracting path are combined with upsampled features in the expanding path to obtain more precise segmentation [12].

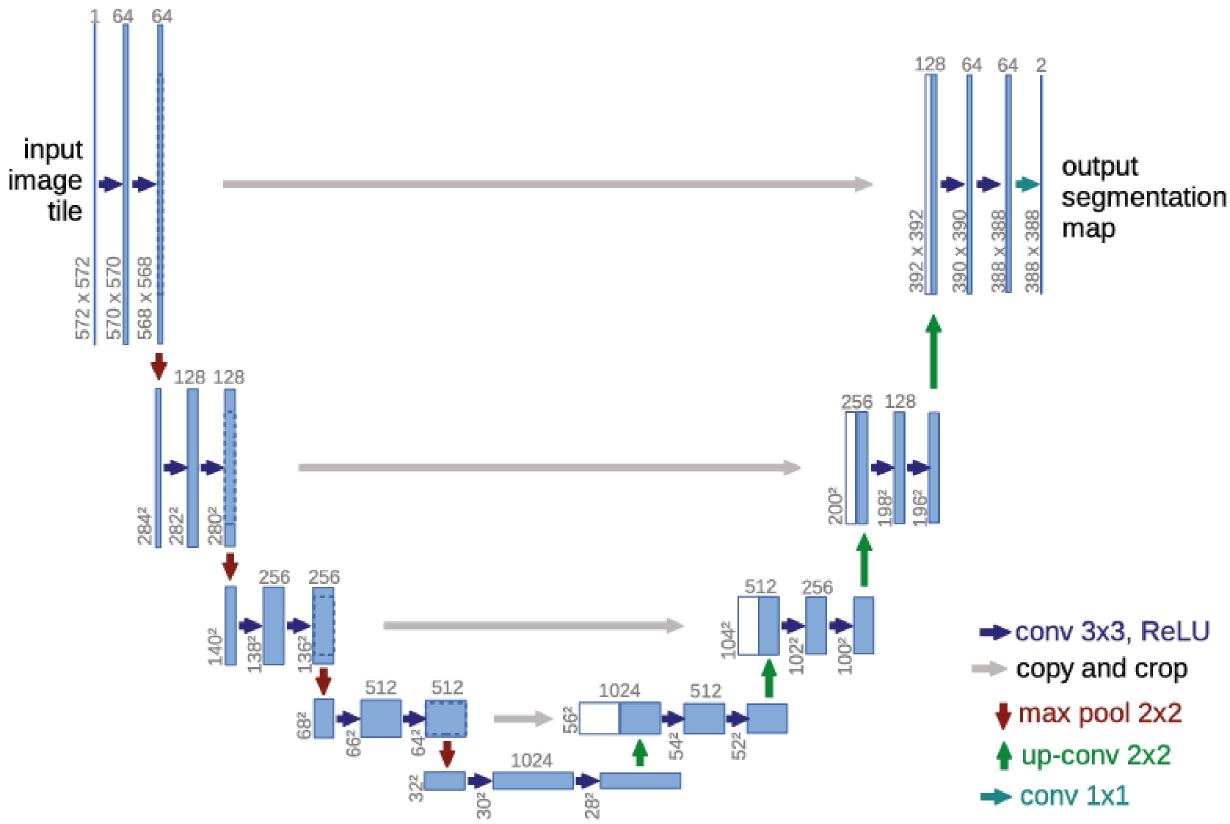

Fig. 1. U-net Architecture (Example for 32×32 Pixels in the Lowest Resolution) [12]

However, several studies have highlighted the unique challenges of this approach, especially in brain tumor classification, such as limited annotated datasets, interscanner variability, and high intra-class variability among tumor types. In MRI analysis, deep learning is not limited only to classification, because it has also been applied to image acquisition, reconstruction, restoration, segmentation, disease prediction, and disease detection [13].

Brain tumor classification using MRI images has been widely explored with different convolutional neural network architectures trained either through transfer learning or from scratch, which is also one of the machine learning techniques. Early studies demonstrated that conventional convolutional neural networks could achieve high accuracy for brain tumor types when applied to curated MRI datasets [14]. In brain tumor classification, CNN models usually learn discriminative patterns from MRI intensity structure, tumor texture, and spatial appearance, which can support automatic categorization of tumor classes [14, 15]. However, classification remains challenging because glioma, meningioma, and pituitary tumor images may show complex and sometimes overlapping visual characteristics [15].

The previously mentioned machine learning technique, called transfer learning, has become known due to limited medical image datasets, which can pose challenges for the models. To adapt to the classification of MRI images, transfer learning allows models pretrained on large natural image datasets with increased convergence stability and performance [15]. Although, despite the promising accuracy results, the researchers have expressed concerns regarding the dataset splitting strategies and possible data leakage [16]. CLAIM also mentions the significance of transparent data splitting and validation strategies to ensure reliable clinical AI development [17].

For reliable reporting, CLAIM recommends that medical imaging AI studies should describe the data source, preprocessing steps, data subsets, ground truth annotations, data partitions, model structure, training details, evaluation metrics, and validation procedure [17]. This is important for brain tumor classification because high accuracy alone does not guarantee that the model is clinically reliable or reproducible. A transparent description of preprocessing, model training, and evaluation makes it easier to compare different studies and understand whether the reported performance can generalize to new MRI data [16, 17].

1.3. Integration of Wavelet Transforms in Image Analysis

Besides traditional CNN architectures, integration of classical signal processing techniques with deep learning models has achieved increasing attention nowadays. In particular, the multi-resolution analysis, which is provided in the discrete wavelet transform, decomposes images into subbands that represent various frequency components, enabling the division of coarse structural information from fine-grained textural details [18]. Several studies have applied wavelets as preprocessing before convolutional neural network classification to reduce dimensionality while preserving diagnostically relevant features. As an example, the Haar wavelet decomposition method has also been integrated as it provides lower computational cost and suitability in such image classification tasks [19]. This type of decomposition is useful because the low-frequency approximation component preserves the main structural information, while the high-frequency components contain directional details such as horizontal, vertical, and diagonal variations. Therefore, DWT can be used not only as a preprocessing method, but also as a compact image representation before CNN-based classification [18, 19].

Additionally, the CNN-enhanced multi-level Haar wavelet features fusion network (CNN-MHWF2N) was proposed to alleviate the issue that ordinary CNNs learn mainly spatial characteristics and exclude the spectral features. The model combines four-level Haar wavelet decomposition features with convolutional features from different network layers and further refines them through a double-filtrating feature fusion module and inter-level information interaction [20].

The wavelet decomposition was used to replace early convolutional layers in the deep learning model design for medical image classification. It was demonstrated that a fixed feature extraction method reduced the number of parameters and reduced the training time. In addition, the models where wavelet decomposition was integrated were more stable than those using convolutions, which led to wider application possibilities in data-limited domains [21].

To improve classification accuracy and noise-robustness in the model, the common CNN operation, which is known as max-pooling, is replaced with a wavelet-based down-sampling layer, which is DWT_{ll}, because the max-pooling operation may lose important spatial information and can be noise-sensitive. This approach performs wavelet decomposition and passes only the low-frequency component to the next stage, as shown in Fig. 2. below. Also, the key point of integration is that it halves the spatial resolution while removing high-frequency components, so the output corresponds to the low-frequency part that keeps the main information of feature maps and reduces noise propagation through deeper layers [22]. Both approaches show that wavelet transforms can replace or modify standard CNN operations, such as early convolutions, pooling, or strided down-sampling to reduce trainable parameters, limit information loss, and improve robustness when training data are limited or noisy [21, 22].

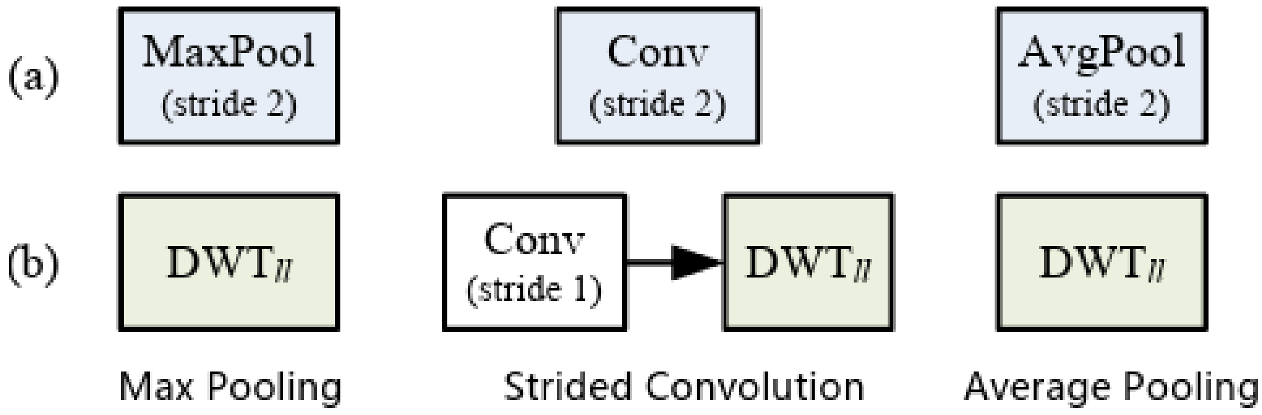

Fig. 2. (a) Baseline, the Down-sampling Operations in Deep Networks. (b) Wavelet Integrated Down-sampling in WaveCNets [22]

Several CNN architectures use wavelet coefficients as input channels to integrate the strengths of the frequency domain, allowing the network to learn from both raw pixel intensities and frequency components at the same time [23]. Moreover, other researchers integrate wavelet pooling layers into CNN architectures to preserve spectral information [24]. Furthermore, to increase robustness to noise, multi-level wavelet decompositions have been applied to directly model hierarchical frequency representations within deep neural networks [25]. Also, the main integration strategies are summarized in Table 2 to clarify how wavelets are used within convolutional neural network architectures.

Table 2. Main Ways of Integrating Wavelet Transforms with CNNs

Integration Approach	How Wavelets are Used	Main Purpose
Wavelet preprocessing [18, 19]	SWT decomposes the image into approximation and detail components before CNN classification.	Reduce image size and provide a compact frequency-based representation.
Multi-level feature fusion [20]	Haar wavelet features are combined with convolutional features from different CNN layers.	Use both spatial and frequency-related information.
Replacement of early CNN operations [21]	Fixed wavelet decomposition is used in place of certain learnable convolutional operations.	Reduce trainable parameters and improve training stability in small datasets.
Wavelet-based down-sampling [22]	DWT replaces max-pooling, average-pooling, or strided convolution by passing low-frequency components forward.	Reduce noise propagation and preserve the main object structure.
Wavelet-enhanced CNN layers [23, 25]	Wavelet coefficients or compressed wavelet representations are added to CNN layers.	Strengthen feature learning with multi-scale image information.

Integration Approach	How Wavelets are Used	Main Purpose
Multi-level wavelet CNN [24]	DWT and inverse wavelet transform are embedded into the CNN architectures.	Reduce information loss during down-sampling while efficiently enlarging the receptive field.

This kind of integration in medical image analysis is useful, especially because tumor regions may contain heterogeneous texture, weak boundaries, and information at different spatial scales. Wavelet coefficients can support the network by separating approximation information from detail information, while CNN layers can learn task-specific patterns from these representations. Therefore, such integration can combine the interpretability of frequency-based decomposition with the automatic feature learning capacity of CNNs [23, 24, 25].

1.4. Convolutional Neural Networks and Their Limitations

In deep learning, convolutional neural network models are widely recognized because of their ability to identify relevant features automatically. This approach effectively processes the images as it utilizes shared weights and local connections, which reduces the number of parameters needed, simplifying the training process and accelerating network performance over other neural network models [26, 27].

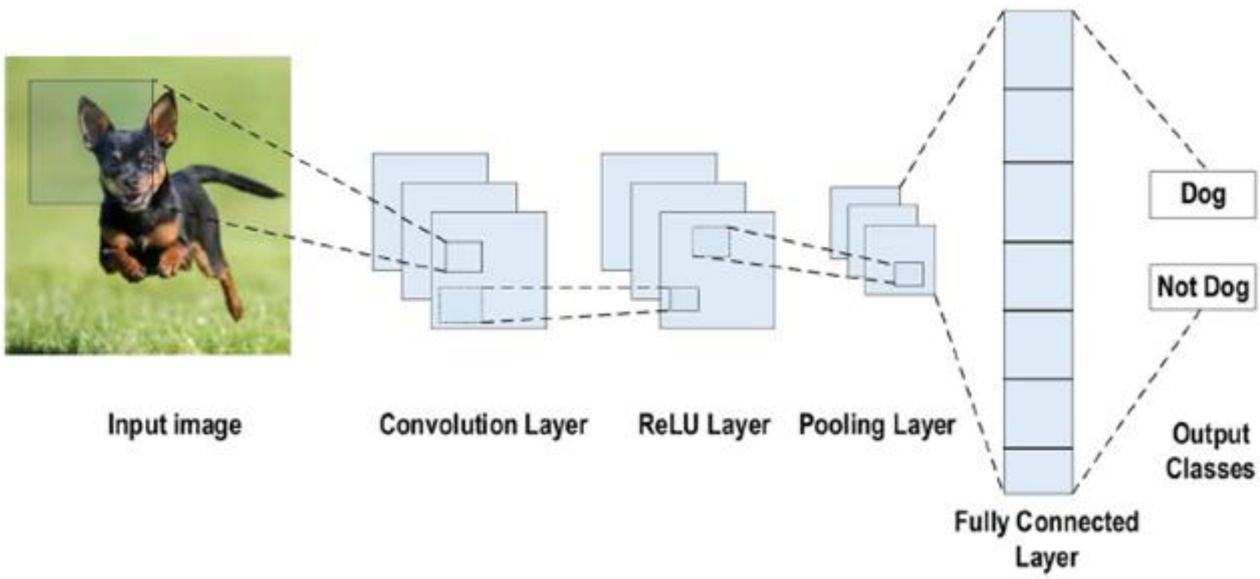

Fig. 3. An Example of CNN Architecture for Image Classification [26]

To better understand the structure of the convolutional neural network model, Fig. 3. above provides the general architecture of the CNN for image classification, where the process first starts with the input layer that receives the image. Then, a typical CNN skeleton for image identification contains multiple convolution layers followed by pooling layers, ending with fully connected layers. Afterward, the output of the convolution layer is passed through the ReLU layer, which helps learn more complex relationships and adds non-linearity to the network. Next, the pooling layer reduces the size of the feature maps to make the computation faster and help prevent overfitting, after which the data is sent to the fully connected layer. In the end, after combining the extracted features, it transfers to the output layer for the classification [26].

To improve and optimize the model for a specific dataset, multiple hyperparameters are used, which can be adjusted depending on the desired result [26, 27]. These essential hyperparameters are:

- *Dense layer*: performs high-level reasoning using extracted features and produces the final classification scores.
- Learning rate: establishes the efficiency and accuracy of model training.
- *Epochs*: represent how many times the model passes through the training data during learning.
- *Activation function*: introduces non-linearity into the network by applying an element-wise activation function to the output of the convolution layer, and it enables the model to learn complex patterns beyond linear relationships.
- *Optimizer*: plays a critical role in improving accuracy and training speed. Adam optimizer is the default one as it is a more reliable improvement.
- *Dropout*: used to select neurons randomly with a given probability that are not used during training, which makes the network less sensitive to the specific weights of neurons.
- *L2 regularization*: used because of its computational efficiency, produces a non-sparse solution, and does not perform feature selection.

Despite the promising results of the convolutional neural network models, there are several limitations for multi-scale image analysis, one of which is that the rescaling of an image can lead to drastically different outputs [28]. Another limitation is that as networks go deeper, the theoretical receptive field grows, but the effective receptive field tends to remain focused [29].

Another important issue is that CNNs are not always stable against small image transformations. Even very small shifts or rescaling of the input image can change the predicted class, although the change may be almost invisible to humans [28]. In addition, the receptive field of deep CNNs can be misleading because the theoretical receptive field may be large, but the effective receptive field usually covers only a smaller central part of it [29]. Therefore, CNN-based image classification should consider not only accuracy, but also robustness to small transformations and the actual amount of image context used by the model [28, 29].

1.5. Convolutional Neural Network with Emphasis on its Architectures

Over the past decade, a number of CNN architectures have been developed to increase CNN performance, with model architecture being crucial for improving application performance. One of the early architectures, named AlexNet, showed a strong potential in image classification as it includes five convolutional layers, three max-pooling layers, together with fully connected and normalization layers [30]. Its main advantage is that it helped CNNs become a strong solution for image recognition, but its large number of parameters can increase computational cost and overfitting risk when MRI datasets are limited [30, 31]. Due to these limitations, the ZefNet architecture was introduced to quantitatively visualize network performance by monitoring neuron activation. The key concept was to monitor the learning process during training, which helped identify performance issues within the model. This was experimentally validated on AlexNet, revealing that only certain neurons were active in the first two layers, while others were inactive [26].

Another architecture, ResNet, was developed to stabilize training in CNNs through residual connections. These connections help information pass through the network and are useful for brain tumor MRI classification. Because deeper layers can learn more complex tumor-related patterns, however, such deeper models may also require more computational cost, which was previously mentioned in the previous architecture and may not always improve the performance, if it is not carefully designed [30, 31].

DenseNet architecture, which uses dense connections, is another approach that improves feature propagation and can help the model preserve low-level and horizontal image information [32, 33]. This architecture is well known for its strong results in brain tumor MRI classification, as it can learn detailed texture information. Nevertheless, the deeper versions of this architecture, such as DenseNet169 and DenseNet201, require more computational resources [33, 34].

More lightweight CNN architectures, such as MobileNet and EfficientNet, are designed to resolve the main issue of the previously mentioned architectures, which is the computational cost, while keeping the good classification performance at the same time. These models can support faster training and deployment for medical image analysis, but very lightweight models may miss important features if the provided MRI images contain weak boundaries or low contrast [32, 35].

To improve CNN performance by emphasizing the most important image regions, like SE-ResNet50, which uses squeeze-and-excitation blocks to improve feature discrimination. As the tumor regions in the brain may contain small and subtle visual differences, this is very relevant, but attention modules also increase the architectural complexity and require careful validation [32, 33]. Hybrid and ensemble architectures also combined different CNN models to use their complementary strengths. For example, the previously mentioned architectures like ResNet, DenseNet, and EfficientNet showed that one model can provide high precision while another can improve the recall [35]. To compare these CNN architectures according to their main idea, together with advantages and limitations, Table 3 was provided.

Table 3. Comparative Overview of CNN for Image Classification

Architecture	**Main Idea**	**Advantage**	**Limitation**
ZefNet [26]	Understands internal CNN feature representations.	Helps identify weak architecture design choices by visualizing hidden-layer activations.	Does not directly reduce model complexity.
AlexNet [30]	Deeper early CNN with convolution, pooling, fully connected layers, and softmax.	Strong feature learning and important historical CNN model.	High parameter count and overfitting risk on small datasets.
GoogleNet [31]	Uses parallel filters of different sizes to capture multi-scale image features.	Processes objects and patterns at different scales with optimized computation.	More complex architecture than simple CNNs.
ResNet [30, 31]	Uses residual or shortcut connections.	Helps train deeper networks and reduces vanishing gradient problems.	More depth can increase complexity and does not always improve results.
SE-ResNet50 [32, 33]	Uses channel attention to emphasize important features.	Improves focus on informative tumor-related regions.	Adds architectural complexity.
MobileNet [32, 35]	Uses depth-wise separable convolutions.	Lightweight and efficient.	May miss subtle details in difficult MRI cases.
EfficientNet [32, 35]	Balances depth, width, and resolution.	High performance with controlled model size.	Still depends on careful tuning and dataset quality.

Architecture	Main Idea	Advantage	Limitation
Hybrid / Ensemble CNN [32, 35]	Combines several CNN architectures or predictions.	Improves robustness and balances precision / recall.	Higher computational cost and lower interpretability.
DenseNet [35]	Reuses features by connecting each layer with later layers.	Good feature propagation and strong texture learning.	Deeper versions require more computation.

Another architecture, GoogleNet, improves training accuracy by optimizing computational cost, allowing for better performance within the same time frame. The inception module in GoogleNet employs parallel convolutional filters of varying sizes (1 × 1, 3 × 3, 5 × 5) and max-pooling to effectively process objects at different scales, then concatenates the results for further processing [30].

Overall, these architectures show that CNN performance depends on the balance among feature-extraction capacity, computational cost, and robustness. Standard CNN architectures mainly learn spatial patterns, while brain MRI images may also contain important frequency and texture information. Therefore, wavelet decomposition can support CNN architectures by providing multi-scale approximation and detail information prior to classification [31, 32, 33].

1.6. Explainable Artificial Intelligence Methods

The explainability artificial intelligence subchapter introduces the main explainable artificial intelligence methods used in this work and explains their importance for model interpretation. It presents Shapley Additive exPlanations and Integrated Gradients, highlighting how these methods help explain the decision-making process of deep learning models. In addition, it discusses their main advantages, limitations, and practical use in image classification tasks. The methods are compared according to the type of explanation they provide, such as image-region localization, feature contribution analysis, and gradient-based attribution.

1.6.1. Shapley Additive exPlanations Method

An explainable artificial intelligence called SHapley Additive exPlanations method is used to interpret how individual input features contribute to a model's prediction. It is based on the idea of fairly distributing the model output among the features according to their contributions, so it shows which variables increase or decrease the prediction for a specific sample. SHAP can be applied to different machine learning models and emphasizes its additive property, in which the total of SHAP values equals the difference between the prediction and a baseline value. It is also useful because it provides both local explanations for individual predictions and global insight into overall feature importance [36].

The general explanation of the SHAP model can be written in Equation (1) [36, 37]:

$$g(z') = \varphi_0 + \sum_{i=1}^{M} \varphi_i z'_i \qquad (1)$$

where *g(z')* is the explanation model for a specific prediction, φ_0 is the baseline or expected model output, M is the number of input features, z'_i shows whether feature i is present or absent in the simplified input representation, and φ_i is the SHAP value of feature i. A positive φ_i means that the feature increases the prediction, while a negative φ_i means that the feature decreases the prediction.

Therefore, the final prediction is explained as the baseline value plus the sum of all feature contributions [36, 37].

In comparison with Integrated Gradients, SHAP gives a stronger feature-contribution interpretation based on cooperative game theory, while Integrated Gradients explains predictions by accumulating gradients from a baseline to the input. SHAP is more general because it can be applied to different types of models, including tree-based models, linear models, and neural networks. However, Integrated Gradients is usually easier to apply to deep neural networks because it only requires gradient calculations, while exact SHAP computation can be expensive for high-dimensional image data [36, 37, 38].

SHAP explains how individual features influence prediction. For brain MRI classification, SHAP can give more detailed contribution information, but pixel-level SHAP explanations for MRI images may be computationally expensive and harder to visualize when the input image contains many pixels [36, 37].

The main limitations of SHAP are computational cost, sensitivity to correlated features, and difficulty in high-dimensional data. Since exact SHAP values require evaluating many possible feature subsets, the computation becomes expensive when the number of features is large. This is especially important for MRI images because each image contains many pixels. In addition, if features are strongly correlated, the distribution of contribution values may become less clear, which can make interpretation more difficult [37].

1.6.2. Integrated Gradients Method

The integrated gradients method is a feature attribution technique that measures the contribution of each input feature to a model's prediction relative to a selected baseline. It is based on two axioms, which are sensitivity and implementation invariance, and computes gradients along the straight-line path between the input and baseline. In this way, it overcomes limitations of standard gradient-based explanations, particularly in situations where the gradient at the input is zero despite the feature influencing the prediction [38].

Equation (2) illustrates how integrated gradients are calculated [38]:

$$IntegratedGrads_i(x) ::= (x_i - x_i') \times \int_{\alpha=0}^{1} \frac{\partial F\big(x' + \alpha \times (x - x')\big)}{\partial x_i} d\alpha \tag{2}$$

where F is the function of the model, x is the original input sample for which the explanation is generated, x' is the baseline input that represents the absence of features, and i is the index of the feature [38].

Integrated Gradients is especially useful when a simple gradient explanation is not enough, because standard gradient attribution can fail when the model output becomes locally flat around the input. In this case, the gradient may be close to zero, even though the feature still strongly influenced the prediction. This explainability method reduces the problem by accumulating gradients along the path from a baseline to the real inputs, rather than using only the gradient at the final input [38, 39].

Integrated Gradients can be used to identify which pixels, or image regions, contribute most to the predicted class. For example, it has been applied to object recognition models to highlight class-specific regions and to diabetic retinopathy prediction, where the attribution maps focused on lesion-related areas in retinal fundus images [39].

However, Integrated Gradients also has limitations. The most important limitation is that the baseline should represent the absence of information, but this is not always easy to define for medical images. A black image is often used as a baseline in image models, but it may not always represent a realistic absence of information in MRI data [38, 39]. Other baseline choices, such as blurred images, noisy images, or uniformly generated images, can produce different explanations, which means that the attribution map may depend on the selected baseline [38].

1.7. Summary of Analytical Outcomes

The literature review shows that traditional image processing methods remain essential for image enhancement and feature extraction. These techniques provide a foundation for image classification by improving image quality and transforming visual information into measurable features. However, when there is noise in the provided images of the datasets, the performance of these image processing methods is limited, which is common in image classification tasks.

Furthermore, it also demonstrates that classical machine learning methods can also support when meaningful features are extracted beforehand. Nevertheless, these approaches depend heavily on handcrafted features and preprocessing quality. Therefore, their effectiveness may decrease when the image contains complex patterns that are challenging to represent manually.

It also illustrates that CNNs have significantly improved the accuracy results in image analysis. Nevertheless, despite their strong performance, these models still exhibit limited robustness to noise and scale sensitivity, which are critical in the classification of brain tumor MRI images.

The reviewed CNN architectures show that model design has a direct influence on classification performance, computational complexity, and robustness. The proposed architecture models introduce different solutions for improving feature extraction and reducing computational cost. However, none of them solves all the previously mentioned challenges. Although these models can improve representation capacity, they may also increase computational cost and overfitting risk, while lightweight models can be efficient but may miss subtle tumor-related details.

Wavelet transforms are suitable for image analysis because they provide a multi-resolution decomposition. In this circumstance, the discrete wavelet transform is applied in a multi-level manner because it is computationally efficient and appropriate for practical image decomposition before CNN-based feature extraction.

In addition, the reviewed studies show that combining wavelet transforms with CNN models can strengthen classification performance by enriching feature extraction with frequency-domain information. Wavelet-integrated models improve the representation of multi-scale image content and can preserve important detail that may otherwise be lost in standard convolution and pooling operations.

Lastly, explainable artificial intelligence methods are important for interpreting model predictions and improving transparency. The reviewed methods provide image-region localization, feature contribution analysis, and gradient-based attribution. Their inclusion is relevant because classification accuracy alone is not enough in medical imaging tasks, where understanding why a model makes a decision is also necessary.

As a result, it has been decided to develop a WaveletFusion architecture that integrates a wavelet decomposition technique together with a convolutional neural network.

2. Specification of the Proposed WaveletFusion Architecture

This chapter illustrates the methodology for the implementation and evaluation of the proposed WaveletFusion architecture. It defines the main requirements and outlines the research project plan. Then, it introduces the Brain Tumor MRI and Br35H :: Brain Tumor Detection datasets, including their preparation, preprocessing, and split strategy. It also includes the main methodological components used in the implementation, including Haar wavelet transforms, two-dimensional convolution, and CNN-based feature extraction. Finally, it explains the model variants explainability procedure, evaluation strategy, quality criteria, and fairness rules that form the basis for the experimental evaluation.

2.1. Requirements of the Proposed Solution

After reviewing existing approaches to integrating wavelet transforms into convolutional neural network models, several drawbacks have been observed. Wavelet transforms are more effective for noise reduction, and CNNs excel at hierarchical feature learning. However, many approaches use only approximation coefficients as CNN input, which discards high-frequency information that is important, especially in medical image analysis. Furthermore, architectures that pursue multi-level decomposition frequently address each decomposition layer independently, failing to propagate the hierarchical relationships between detail subbands and deeper convolutional representations.

To address these limitations, a WaveletFusion architecture was proposed that integrates multi-level Haar wavelet decomposition up to eight levels with a CNN backbone.

The main requirements of the proposed solution are as follows:

- Integrate multi-level Haar decomposition while preserving detail coefficients at each decomposition stage.
- Incorporate the hierarchical decomposition structure into CNN layers through a cross-scale fusion strategy inspired by the U-net architecture.
- Maintain identical preprocessing, training, and evaluation conditions to enable a fair comparison with the Baseline CNN model.
- Preserve computational efficiency while improving classification performance.
- Maintain reproducibility through unified experimental and evaluation standards.

This ensures that the final classifier receives structural information from all spatial scales and establishes a robust experimental foundation for evaluating the contribution of Haar wavelet components to enhance CNN in image classification.

Moreover, the following non-functional requirements are used as a constraints of the proposed solution and its experimental evaluation:

- Ensure the research process and results are transparent and reproducible.
- Ensure consistent preprocessing, training, and evaluation settings across all compared models.
- Ensure that all experiments are computationally feasible on standard hardware resources.
- Use established and reliable tools, libraries, and evaluation methods.
- Make the analysis accessible for other researchers to review or replicate.
- Keep computational complexity and parameter count within a controlled limit.

2.2. Research Project Plan

To investigate how wavelet-based image decompositions can enhance the performance of convolutional neural network models and reduce the number of parameters, this research was organized as a structured experimental study. The main purpose of the methodology is to design, implement, and evaluate a WaveletFusion architecture. The study focuses not only on the final classification performance, but also on how different decomposition depths, wavelet detail subbands, and level-group contributions influence model behavior.

This research unfolds the core questions:
1. How does the level of Haar wavelet decomposition affect CNN model complexity and classification performance?
2. Which combination of approximation and detail subbands achieves the most effective classification performance?
3. How robust is the proposed architecture hybrid across different decomposition and fusion configurations?

To answer these questions, the research follows a step-by-step methodological framework. First, two publicly available datasets were selected to evaluate the proposed solution in both multi-class and binary classification settings. Then, the MRI images were prepared using a unified preprocessing pipeline, including grayscale loading, resizing, normalization, per-image standardization, and augmentation. After that, the proposed solution was then developed by integrating recursive Haar wavelet decomposition with CNN feature extraction.

The experimental phase was organized in several stages. First, the proposed architecture with different decomposition depths was trained and compared to determine how the number of wavelet levels affects the performance. After identifying the strongest decomposition levels, additional ablation studies were made to analyze and understand the contribution of individual directional detail subbands and different level groups. Finally, the best-performing WaveletFusion variants were compared with the Baseline CNN and selected existing models using unified evaluation metrics.

For each decomposition level and coefficient configuration, the number of trainable parameters, validation and test losses, validation and test accuracies, macro-F1, training time, and per-class evaluation metrics were recorded. In addition, confusion matrix analysis was used to check the distribution of correct and incorrect predictions among the classes. Grad-CAM interpretability analysis was also applied to selected examples to observe which image regions influenced the model‘s predictions and to better understand the behavior of the proposed architecture.

Finally, the quantitative results were provided to discuss configuration-specific findings, highlighting scenarios where wavelet-based decomposition offers the greatest gains. It also provided a structured basis for comparing the proposed WaveletFusion variants with the Baseline CNN under a fair evaluation setting.

2.3. Dataset Description and Characteristics

This subchapter introduces two publicly available brain MRI datasets to evaluate the proposed WaveletFusion architecture in different medical image classification settings. Both datasets consist of grayscale brain MRI images and are suitable for image classification tasks. For multi-class classification, the Brain Tumor MRI dataset was used, where the model must distinguish between different tumor types and non-tumor cases. The second dataset, Br35H :: Brain Tumor Detection 2020, was used for binary classification.

2.3.1. Brain Tumor MRI Dataset

The Brain Tumor MRI dataset is a collection of human brain MRI scans used for early detection and classification research. This dataset provides 7200 high-quality images collected from different sources and contains four classes, which contain categorized images that are glioma, meningioma, pituitary, and no-tumor, as shown in Fig. 4. below. Therefore, it represents a multi-class classification problem, where the model must not only detect the tumor, but also identify the specific tumor type [40].

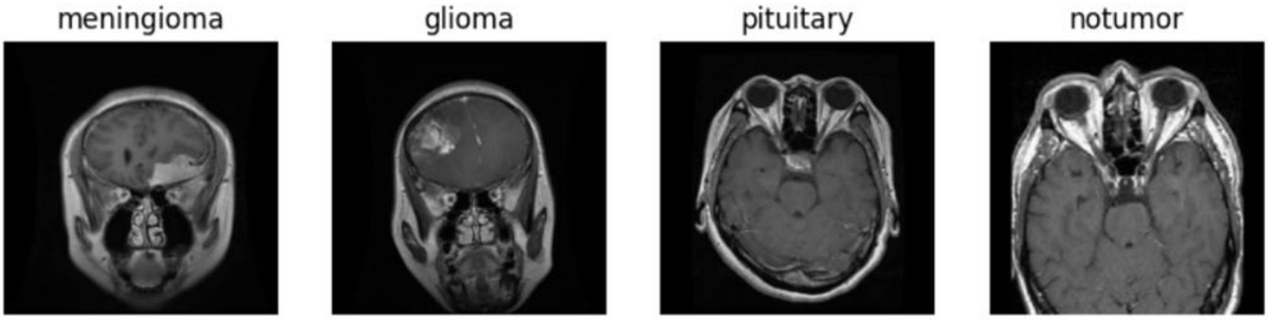

Fig. 4. Brain Tumor MRI Dataset

The dataset was structured into training and testing sets, where the training set contains 5600 images with 1400 images for each class, and the testing set contains 1600 images with 400 images for each class. To avoid data leakage, the updated dataset removed duplicated images and eliminated overlap between the training and test sets [40].

2.3.2. Br35H :: Brain Tumor Detection 2020 Dataset

The Br35H :: Brain Tumor Detection 2020 dataset is another publicly available dataset that contains 3060 brain MRI images designed for a binary classification task. The images in this dataset were organized into three folders: yes, no, and pred, where the yes (tumor cases) and no (no-tumor cases) folders contain 1500 MRI images each, and the pred folder is provided separately for prediction purposes. In this research, only the labeled yes and no classes were selected for model training and evaluation, and Fig. 5. illustrates a representative sample from this dataset. The dataset was selected because it provides clear binary classification and a balanced distribution of the two labeled classes, which is useful for evaluating convolutional neural network models [41].

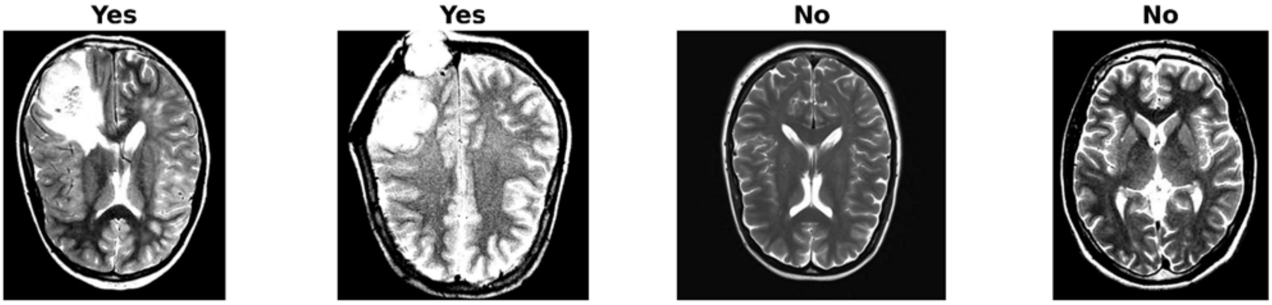

Fig. 5. Br35H Dataset

2.4. Dataset Preparation, Preprocessing, and Split Strategy

The proposed WaveletFusion architecture applies a Haar decomposition layer that expects single-channel input. MRI images are loaded in grayscale. During loading, each image is resized to a fixed resolution of 256 × 256 to ensure that the input tensor has a consistent spatial shape for both the Haar decomposition layers and the convolutional blocks. This resolution was also selected because it supports repeated two-dimensional Haar decomposition, where each decomposition level reduces the spatial resolution by half.

The datasets were separated into training, validation, and test subsets using fixed and reproducible strategies. For the Brain Tumor MRI dataset, the images in the predefined training directory were split into 80% for training and 20% for validation, whereas the separate testing directory was used as the final independent test set. For the Br35H dataset, the labeled images from the yes and no folders were first split as in the previously mentioned dataset to 80%/20%. Then, the training-validation pool was divided again into 80%/20%.

To align with the proposed pipeline, preprocessing was applied consistently across all models. The dataset preparation and preprocessing strategy can be summarized as follows:

- All MRI images were loaded as single-channel images.
- Each image was resized to 256 × 256 pixels to provide a fixed input size for all models and to enable repeated Haar wavelet decomposition.
- Pixel values were rescaled to the [0, 1] interval by dividing the original intensity values by 255.
- Each image was standardized individually to reduce contrast and intensity differences between MRI slices and to stabilize the optimization process during training.
- Data augmentation was applied only during model training to improve generalization and reduce overfitting.
- Gaussian noise was applied with a standard deviation of 0.02 during training to improve the robustness of the models.
- All experiments were repeated across five independent runs.

These preprocessing steps are important for stable normalization and consistent grayscale representation, supporting fair comparison between the Baseline CNN and WaveletFusion models.

2.5. Methodological Components of the Proposed Approach

This subchapter introduces the main methodological components used in the proposed WaveletFusion architecture. Since the proposed solution combines wavelet-based image decomposition with convolutional neural network feature extraction, two main components are described.

2.5.1. Haar Wavelet and Discrete Wavelet Transforms

The Haar wavelet transform decomposes a signal into two components, the approximation and detail coefficients, where the Haar wavelet function is essential, as it defines how the signal is split into these components. Equation (3) below illustrates the mathematical form of the Haar mother wavelet function, where $\psi(x)$ is the 1D Haar wavelet function and x is the independent variable [42].

$$\psi(x) = \begin{cases} 1, & 0 \leq x < \frac{1}{2} \\ -1, & \frac{1}{2} \leq x < 1 \\ 0, & otherwise \end{cases} \tag{3}$$

In image processing, the 1D Haar wavelet transform can be extended to two-dimensional (2D) image decomposition, which is widely used in image processing and image compression due to its simple implementation, memory efficiency, and low computational cost. By applying the one-dimensional Haar transform first to the rows and then to the columns of the image. As a result, one decomposition stage produces the four subbands: low-low approximation subband (LL), and three directional detail subbands (LH, HL, HH), while repeated decomposition of the LL subband provides a multi-level representation of the image shown, as shown in Fig. 6. below [42].

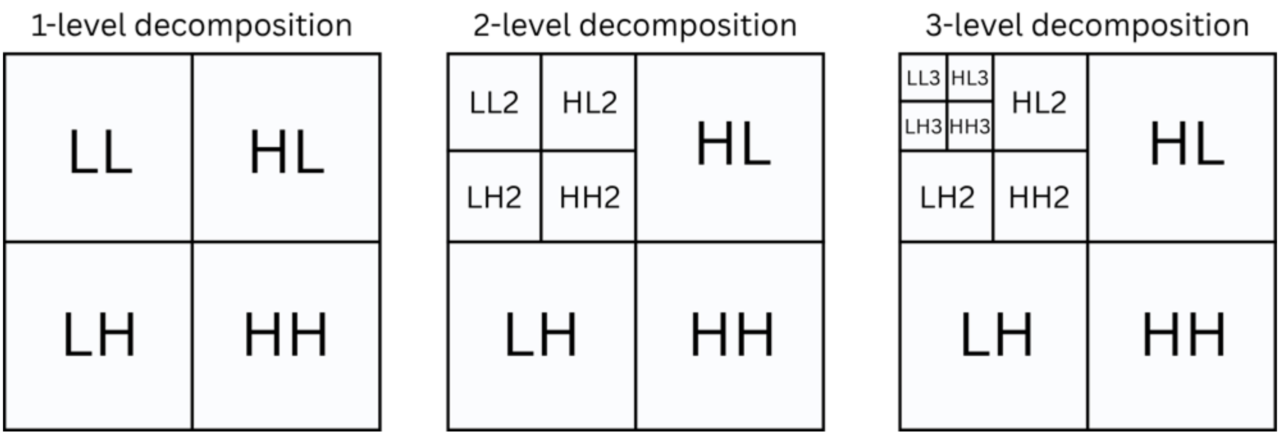

Fig. 6. Structure of Wavelet Decomposition

Discrete Wavelet Transform is a multiresolution technique for decomposing an image into components at different frequency ranges. It is widely used because it separates coarse image structure from fine details, which makes it suitable for compression, denoising, feature extraction, and segmentation. In the two-dimensional case, DWT is applied first to the rows and then to the columns of the image, producing four subbands shown in Fig. 7. below. The LL subband preserves the approximation of the image, while LH, HL, and HH contain vertical, horizontal, and diagonal detail information, respectively [42, 43].

The four subbands produced by the one-level 2-dimensional Haar wavelet decomposition are described below [43]:

- *Low-low (LL) subband*: preserves the approximation and main structural information of the image.
- *Low-high (LH) subband*: preserves directional detail information obtained from low-pass and high-pass filtering.
- *High-low (HL) subband*: preserves directional detail information obtained from high-pass and low-pass filtering.
- *High-high (HH) subband*: preserves diagonal detail information obtained from high-pass filtering in both directions.

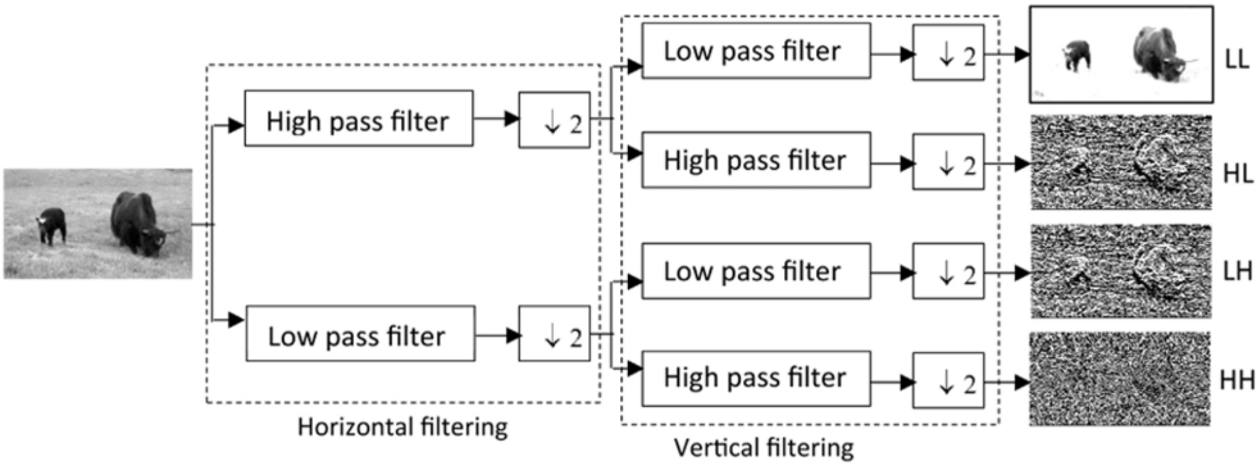

Fig. 7. Decomposition of Yak image Using Two Dimensional Discrete Wavelet Transform [43]

2.5.2. CNN-based Feature Extraction and Classification

Two-dimensional convolution extends the same overlap principle to two-dimensional signals such as images, where a small matrix called a kernel is moved across the image. At each position, the overlapping elements are multiplied and summed to produce one output value. This operation is shown in Equation (4), where x is the input image, h is the kernel, y is the output image, and i and j denote the output location [44].

$$y[i,j] = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} h[m,n] \cdot x[i-m, j-n] \tag{4}$$

In the proposed methodology, CNN-based feature extraction is used to transform MRI image information into learned feature representations that can be used for classification. In the Baseline CNN, convolutional layers operate directly on the preprocessed MRI image, allowing the model to learn spatial features from the original image domain. In the WaveletFusion model, convolutional layers are applied after the Haar wavelet decomposition, allowing the model to learn from both the approximation and the directional detail subbands.

Therefore, convolution has two roles in this research. First, it provides the main feature extraction mechanism for reference Baseline CNN. Second, it transforms wavelet-domain components into trainable feature maps inside the WaveletFusion model. The final class prediction is then produced from the feature representation generated by the corresponding architecture.

2.6. Proposed WaveletFusion Architecture

The main idea of the proposed WaveletFusion architecture is to combine the multi-resolution representation of the two-dimensional Haar wavelet transform with the hierarchical feature learning ability of convolutional neural networks. The processing pipeline of the model is summarized in Fig. 8 below. Each preprocessed axial brain MRI slice is first decomposed using a two-dimensional Haar wavelet transform. The approximation component is recursively passed to the next decomposition level, while the three detail components are immediately routed to a convolutional branch for feature extraction. This preserves the hierarchical structure of the wavelet decomposition and allows the use of information from multiple spatial scales of the classifier rather than relying on the final low-frequency approximation only.

The LL branch acts as the carrier of coarse contextual information and is repeatedly decomposed until the deepest level is reached. At the same time, the LH, HL, and HH detail subbands at each level preserve edge, boundary, and texture information that may otherwise be weakened or lost in standard convolution-pooling pipelines. Instead, it processes detailed information at every scale and progressively combines it with deeper features, which is conceptually similar to the skip-connection idea in U-Net-like architectures.

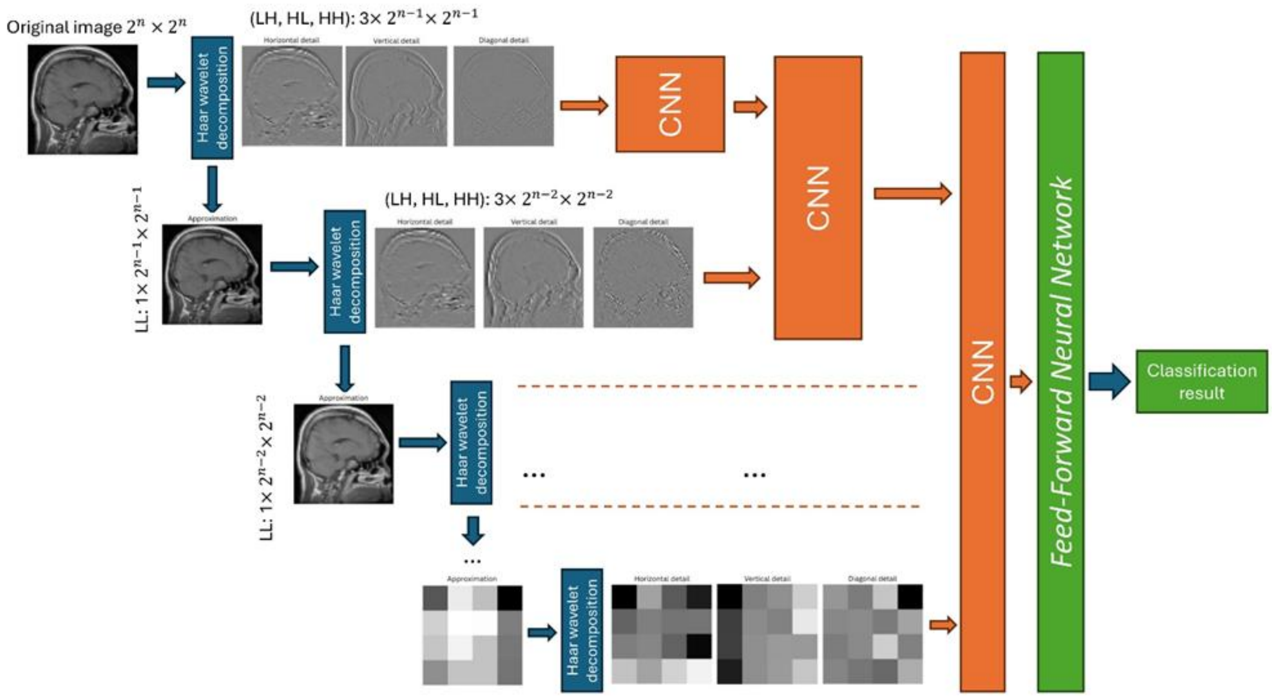

Fig. 8. Schematic of the Multi-Scale Haar Decomposition and CNN Fusion Model

After preprocessing, the wavelet decomposition is performed by a custom HaarWavelet2D layer with a stride of 2. The three detail subbands are concatenated and form a three-channel detail tensor, which is processed by a dedicated convolutional block composed of two 3 × 3 convolutional layers, batch normalization, ReLU activation, and spatial dropout. In parallel, the approximation subband LL is also processed by a smaller convolutional block so that the low-frequency approximation information is not only propagated to the next decomposition level, but also represented in a trainable feature form at the current level. A compact structural overview of the model components, output shapes, and parameter counts is described in Table 4.

Table 4. Structural Overview of the WaveletFusion 8-level Model for Brain Tumor MRI Dataset

Model	**Main Component**	**Output Shape**	**Parameters**
WaveletFusion 8-level	Input Image	(None, 256, 256, 1)	0
	Preprocessing and augmentation	(None, 256, 256, 1)	0
	8-level Haar wavelet decomposition	L1: 128 × 128 → L8 1 × 1	0
	Multilevel detail and approximation feature extraction	L1: (None, 128, 128, 28) → L8: (None, 1, 1, 44)	286,180
	Global average pooling and multi-level feature concatenation	(None, 172)	0
	Dense classification head	(None, 128)	22,144
	Softmax output layer	(None, 4)	516
Total:			308,840

After the separate processing of the approximation and detail branches, both outputs are fused with the previous backbone representation. At the first decomposition level, the fusion is performed between the detail features and the approximation features. From the second level onward, the fusion block also receives the pooled backbone representation from the previous level. The fused tensor is then passed through another convolutional block to form the backbone representation of that level. For all levels except the last one, this representation is downsampled by max-pooling and forwarded to the next stage of the architecture. In parallel, the LL approximation subband is also forwarded to the next Haar decomposition level.

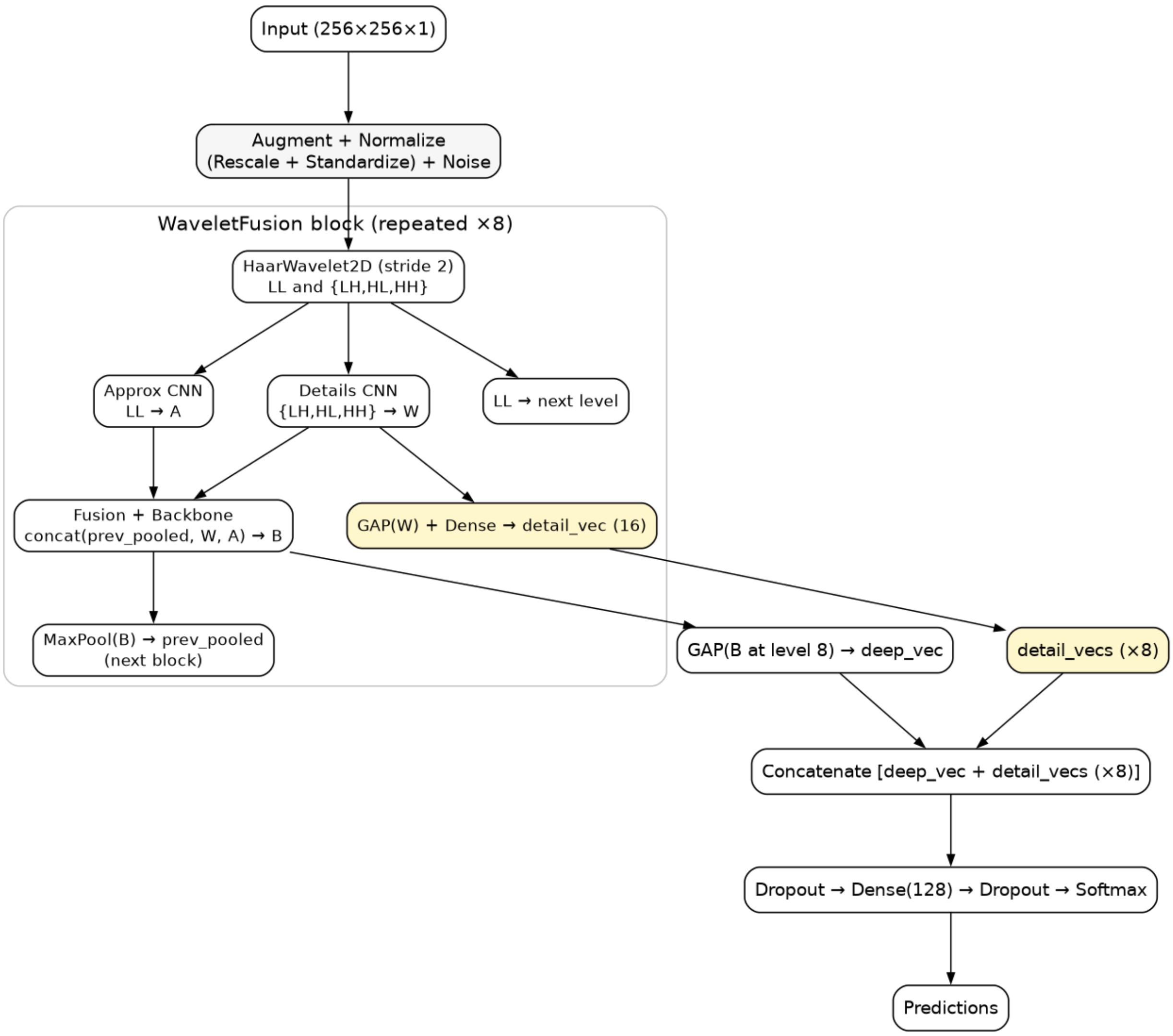

Fig. 9. The Proposed Architecture

In the final implementation used in this work, the architecture applies eight decomposition levels. At each level, the detail branch produces a compact level-specific descriptor using global average pooling followed by a small dense projection layer. At the deepest level, the final backbone feature map is also compacted using global average pooling. The classification head then concatenates the deepest backbone descriptor with the projected detail descriptors from all decomposition levels. This strategy allows the final classifier to combine deep global context with information extracted from every detail level of the wavelet decomposition. The structural logic of the proposed model is illustrated again in Fig. 9. provides a more detailed view of the architecture. It clarifies the two outputs accumulated during the process.

To provide a fair and interpretable reference for comparison, a lightweight Baseline CNN architecture was implemented alongside the proposed WaveletFusion model, as shown in Fig. 10. below. In contrast to WaveletFusion, however, Baseline CNN learns spatial representations directly from the input image and does not use any explicit wavelet decomposition or subband fusion mechanism.

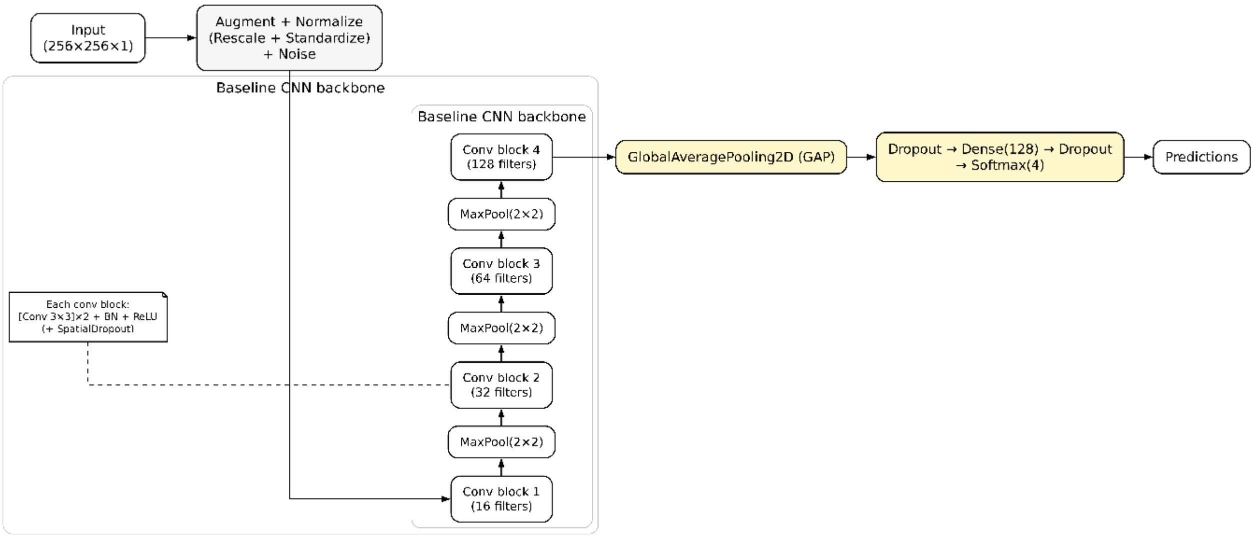

Fig. 10. The Architecture of Baseline CNN

The Baseline CNN feature extractor consists of four convolutional stages with 16, 32, 64, and 128 filters, respectively. Each stage contains two 3 × 3 convolutional layers, with batch normalization and ReLU activation applied afterward, while spatial dropout is incorporated into the convolutional stages to reduce overfitting. Max-pooling is applied after the first three stages in order to gradually reduce spatial resolution and enlarge the effective receptive field. After the final convolutional stage, global average pooling is used to convert the final feature maps into a compact vector. This vector is then forwarded through a dense layer with 128 neurons, dropout regularization, and a final softmax layer that produces the class probabilities.

The essential difference between the two models is that the Baseline CNN learns only from spatial image representations, whereas the proposed WaveletFusion model combines spatial convolutional learning with explicit multi-scale frequency-domain decomposition.

2.7. Explainability Analysis Procedure

In addition to quantitative evaluation, an explainability analysis was included to investigate how the WaveletFusion model forms its predictions. For this purpose, Gradient-weighted Class Activation Mapping (Grad-CAM) was used to demonstrate the spatial regions that contributed most strongly to the model's decision. It works by using the gradient information flowing into the last convolutional layer to estimate the importance of feature maps for a class of interest and then produces a class-discriminative localization map. Unlike the original Class Activation Mapping techniques, Grad-CAM does not require architectural modification or retraining and can be used with a wider range of CNN-based models, including models with fully connected layers and structured outputs. Equation (5) shows how the importance weight α_k^c is computed for each feature map k for class c [45].

$$\alpha_k^c = \frac{1}{Z}\sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k}, L_{Grad-CAM}^c = ReLU\left(\sum_k \alpha_k^c A^k\right) \tag{5}$$

Grad-CAM was selected for the experimental analysis because it is directly suitable for convolutional neural network models and produces visual heatmaps that can be compared with MRI image regions.

This is important for the proposed WaveletFusion model because the aim of the explainability analysis is to observe which spatial areas of the brain MRI image influence the classification decision. SHAP and Integrated Gradients can also explain model predictions, but they are more commonly used for feature-level or pixel-level attribution and may be more difficult to present clearly in layer-wise CNN visualization. Therefore, Grad-CAM was chosen because it provides a practical and interpretable way to analyze class-discriminative regions across the convolutional blocks of the WaveletFusion model.

The Grad-CAM analysis was performed in a layer-wise manner across successive convolutional blocks of the WaveletFusion backbone. Earlier blocks were expected to respond more strongly to general anatomical structures and edges, whereas deeper blocks were expected to produce more class-specific attention patterns. For correctly classified samples, the heatmaps were generated with respect to the predicted class to determine whether the model concentrates on clinically meaningful intracranial regions. For misclassified samples, two complementary views were considered: Grad-CAM targeted to the predicted class, and Grad-CAM targeted to the true class. This comparison provides an additional way to interpret misclassified samples by showing whether the relevant true-class evidence is still visible or whether it is outweighed by activation regions linked to the wrong prediction. The Grad-CAM settings used for these correctly and incorrectly classified cases are summarized in Table 5 below.

Table 5. Grad-CAM Configuration Used for Explainability Analysis

Component	**Applied Setting**
Model	WaveletFusion 8-level
Dataset	Brain Tumor MRI
Classes	Glioma, meningioma, pituitary, no tumor
Example types	Correctly and incorrectly classified examples
Target for correct samples	Predicted class
Target for incorrect samples	Predicted class and true class
Visualized layers	Backbone blocks B1-B5
Visualization mode	Input image with Grad-CAM overlay

The main purpose of this procedure is not only to provide visual confirmation of the model's behavior, but also to relate the decision-making process to the proposed multi-level fusion design. In this way, the explainability analysis complements the numerical evaluation by showing how the learned multi-scale wavelet features influence the localization and interpretation of class-discriminative evidence in brain MRI slices.

2.8. Evaluation Protocol and Experimental Fairness Rules

A strict evaluation protocol is used to ensure that differences in results are caused by the model design and not by uncontrolled changes in training or data preparation. The training procedure was repeated across five independent runs. For each run, the same train-validation split logic, preprocessing sequence, optimization settings, and evaluation pipeline were preserved. In addition, the experiments also recorded computational characteristics, including training time, average epoch time, and training throughput.

The fairness and reproducibility protocol followed these principles:
1. Each model was evaluated using the same dataset-specific split strategy.
2. The same image loading, resizing, normalization, standardization, augmentation, and noise-injection strategy was preserved.
3. All compared models used the same batch size, optimizer, learning rate, loss function, early stopping strategy, learning-rate reduction strategy, and maximum number of epochs.
4. The test set was not used during model training or validation. It was reserved only for the final evaluation.
5. Each model configuration was evaluated across five independent runs.
6. The final quantitative results were reported as mean ± standard deviation across the five runs.
7. When comparing the Baseline CNN with WaveletFusion models, the experimental setup was kept unchanged.
8. In the ablation experiments, only one design factor was changed at a time.
9. All models were evaluated using the same classification and computational criteria.
10. The experiments were executed in the same software and hardware environment.

This protocol supported a fair and reproducible comparison between the Baseline CNN and WaveletFusion models.

2.9. Evaluation Strategy and Quality Criteria

The quality of the proposed WaveletFusion model was evaluated using a suite of complementary metrics drawn from our experimental study.

Firstly, the accuracy was used as a main indicator of classification performance of the dataset, which is mathematically written in the Equation (6) below, where y_i represents the true label of the i^{th} sample in the dataset, $\widehat{y_i}$ are the predicted labels produced by the model and $n_{samples}$ is the total number of samples in the dataset. The function $1(\widehat{y_i} = y_i)$ returns *1* if the predicted label matches the actual label.

$$Accuracy = \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} 1(\widehat{y_i} = y_i) \tag{6}$$

To capture not just correctness but also the confidence of the model's probability outputs, Equation (7) shows that categorical cross-entropy loss was employed during both training and evaluation, where y_i is the true label from the one-hot encoded target vector, $\widehat{y_i}$ is the predicted probability for the class i and *C* is the number of classes.

$$Loss = -\sum_{i=1}^{C} y_i * log(\widehat{y_i}) \tag{7}$$

To understand where the models are making mistakes, how well they are performing in the classification tasks, and to compare the predictions. The confusion matrix breaks down the model's predictions into four categories:
- *True Positive (TP)*: instances where it correctly identifies a positive class when it is indeed positive

- *True Negative (TN)*: instances where it correctly identifies a negative class when it is indeed negative
- *False Positive (FP)*: instances where it incorrectly identifies a positive class when it should have been negative
- *False Negative (FN)*: instances where it incorrectly identifies a negative class when it should have been positive

To deal with the imbalanced datasets and track the performance, three other main metrics in Equation (8) were provided, such as the precision, F1-score, and recall, which are also calculated using the formulas provided in the equation below:

$$\begin{aligned} Precision &= \frac{TP}{TP + FP} \\ Recall &= \frac{TP}{TP + FN} \\ F1 - score &= \frac{2 \times Precision \times Recall}{Precision + Recall} \end{aligned} \tag{8}$$

Macro-F1 was used in Equation (9) to calculate the average F1-score across all classes by giving equal importance to each class. This is useful because it prevents the overall evaluation from being dominated by classes with stronger or easier classification results. Macro-F1 is calculated as follows, where F is the F1-score of class c, and *C* is the total number of classes.

$$Macro - F1 = \frac{1}{C}\sum_{i=1}^{C} F1_i \tag{9}$$

For each model, the number of trainable parameters was recorded to compare model size. FLOPs were calculated to estimate the number of floating-point operations required for one forward pass. Moreover, training efficiency was assessed using total training time to measure how long the full training process takes, average epoch time, and training throughput.

2.10. Specifications of Test Environment and Libraries

The implementation was carried out using Python 3.12.9 and TensorFlow 2.17.1 versions. TensorFlow was used as the main deep learning framework for building, training, and evaluating the Baseline CNN and WaveletFusion models.

The following libraries and software environments were applied throughout the implementation and testing phase:

- *Python* (3.12.9): it is the main programming language and framework for the development of neural network models. This programming language was selected due to its simplicity, broad accessibility, and many machine-learning and deep-learning libraries, which are supported by the community.
- *NumPy library* (1.26.4): this library is commonly used for performing advanced mathematical and scientific tasks in a Python environment, as it provides multidimensional arrays and matrices.

- *Matplotlib library* (3.10.0): this library is used for creating static visualizations, including heat maps, bar charts, line plots, and more. It can also embed plots in graphical user interface applications and export the visualization graphs in multiple file formats.
- *Pandas library* (2.2.3): it is a library that is designed for structured data analysis. It provides support for data alignment and handling missing data. Moreover, it provides time series analysis and can elaborate on the dataset.
- *Scikit-learn library* (1.6.1): it is a library for machine learning that integrates with a wide range of algorithms for supervised and unsupervised learning, including classification and clustering.
- *TensorFlow library* (2.17.1): this library is an open-source machine learning library and is designed for high-performance numerical computation. It represents computation as a graph of operations, enabling efficient deployment on CPUs and GPUs.

All experiments were executed in a modern Python environment, ensuring compatibility and performance.

3. Experimental Evaluation and Benchmarking of WaveletFusion Model

This chapter illustrates the experimental setup used to evaluate the proposed WaveletFusion architecture. It presents a testing and evaluation procedure, including dataset preparation, training configurations, and evaluation metrics. In addition, it presents ablation studies, reports overall benchmark results and explainability analysis, and provides a basis for evaluating the effectiveness of the proposed model.

3.1. Experimental Setup and Computing Environment

Since a convolutional neural network model requires significant computational resources to perform efficiently, especially on complex datasets, the following models in this research were trained on the Jupyter server environment at Kaunas University of Technology. According to the system configuration, the available graphical processing hardware consisted of NVIDIA H100 NVL GPUs with 95,830 MiB of memory each. The environment used NVIDIA driver version 565.57.01 and supported CUDA version 12.7. The computing environment was equipped with two Intel Xeon Gold 6438Y+ processors, providing a total of 64 logical CPUs, and approximately 1.0 TiB of system RAM.

3.2. Experimental Scenarios and Model Variants

The experimental study was organized in several stages in order to evaluate the proposed architecture systematically. First, the Baseline CNN was used as the reference model. Then, a family of WaveletFusion variants was implemented by changing only the depth of the multi-level Haar decomposition while keeping the remaining training and evaluation conditions fixed. In these variants, the decomposition depth ranged from one level to eight levels.

After the wavelet decomposition depth comparison, the best-performing configurations were further analyzed. One ablation experiment examined the contribution of directional detail subbands by testing different combinations while keeping the overall model structure unchanged. Another ablation experiment examined the contribution of level groups by retaining only selected decomposition ranges.

This experimental design answered several important questions, such as how much decomposition depth is beneficial, which directional detail subbands carry the most discriminative information, and which decomposition scales contribute most strongly to final classification performance. Because all variants were trained under the same protocol, the results provide a controlled and fair basis for comparing the role of wavelet-based information in the proposed framework.

3.3. Training Configuration and Optimization Strategy

To evaluate the duration of the training, all CNN models are trained for a maximum of 30 epochs using a batch size of 32, and the time to complete training is used as the final score, which is shown in Table 6. Label smoothing was applied during training of the models to improve generalization, and the Adam optimizer was used with 3×10^{-4} learning rate, as it provides stable convergence for convolutional neural network models.

Table 6. Training Configuration Used for All Compared Models

Setting	Value
Input size	256 × 256 × 1

Setting	Value
Batch size	32
Maximum epochs	30
Optimizer	Adam
Learning rate	3×10^{-4}
Loss function	Categorical cross-entropy
Label smoothing	0.05
Learning-rate reduction	Monitor validation loss, patience = 3, factor = 0.5
Regularization	L2 regularization, spatial dropout, dropout

In addition, a learning-rate reduction strategy was used so that the learning rate would decrease automatically when validation loss stopped improving for several epochs. Regularization was further strengthened through L2 weight regularization in convolutional and dense layers, spatial dropout inside feature extraction blocks, and dropout in the classifier head. These settings were kept fixed across the compared models so that the results would reflect differences in feature representation rather than differences in optimization strategy.

3.4. Ablation Study of Wavelet Decomposition Depth

To evaluate how the number of wavelet decomposition levels influences the proposed WaveletFusion architecture, it was tested with decomposition depths ranging from 1-level to 8-level (L1 to L8) on two selected datasets. The model was trained under the same experimental protocol, and at each decomposition level, it was evaluated across five runs [46].

Table 7. Quantitative Evaluation of WaveletFusion Models with Different Wavelet Decomposition Depths on the Brain Tumor MRI and Br35H Datasets Across 5 Runs (Mean ± Standard Deviation)

Dataset	Depth	Parameters	Test Accuracy	Test Loss	Test Macro-F1	Validation Accuracy	Validation Loss	Validation Macro-F1
Brain Tumor MRI	L1	19,708	0.61 ± 0.01	1.05 ± 0.01	0.55 ± 0.03	0.64 ± 0.02	0.95 ± 0.01	0.59 ± 0.03
	L2	58,616	0.67 ± 0.02	0.91 ± 0.04	0.61 ± 0.02	0.71 ± 0.03	0.82 ± 0.05	0.64 ± 0.03
	L3	101,812	0.72 ± 0.01	0.84 ± 0.02	0.68 ± 0.02	0.76 ± 0.03	0.72 ± 0.03	0.71 ± 0.05
	L4	143,862	0.77 ± 0.01	0.81 ± 0.02	0.75 ± 0.01	0.83 ± 0.02	0.65 ± 0.02	0.82 ± 0.02
	L5	185,912	0.83 ± 0.03	0.74 ± 0.06	0.82 ± 0.04	0.88 ± 0.02	0.57 ± 0.05	0.88 ± 0.03
	L6	226,888	0.88 ± 0.02	0.67 ± 0.05	0.87 ± 0.03	0.93 ± 0.02	0.49 ± 0.05	0.93 ± 0.02
	L7	267,864	0.89 ± 0.01	0.64 ± 0.02	0.89 ± 0.01	0.95 ± 0.01	0.48 ± 0.03	0.95 ± 0.01
	L8	308,840	0.85 ± 0.03	0.73 ± 0.07	0.85 ± 0.03	0.92 ± 0.02	0.55 ± 0.04	0.92 ± 0.02
Br35H	L1	19,450	0.81 ± 0.02	0.48 ± 0.03	0.81 ± 0.02	0.81 ± 0.05	0.48 ± 0.06	0.81 ± 0.05
	L2	58,358	0.85 ± 0.01	0.45 ± 0.03	0.85 ± 0.01	0.85 ± 0.04	0.46 ± 0.07	0.84 ± 0.04
	L3	101,554	0.88 ± 0.01	0.43 ± 0.01	0.88 ± 0.01	0.86 ± 0.03	0.45 ± 0.05	0.86 ± 0.03
	L4	143,604	0.87 ± 0.01	0.46 ± 0.02	0.87 ± 0.01	0.86 ± 0.03	0.47 ± 0.04	0.86 ± 0.03
	L5	185,654	0.89 ± 0.04	0.45 ± 0.08	0.89 ± 0.04	0.88 ± 0.06	0.45 ± 0.08	0.87 ± 0.06
	L6	226,630	0.96 ± 0.02	0.34 ± 0.03	0.96 ± 0.02	0.95 ± 0.03	0.34 ± 0.05	0.95 ± 0.03
	L7	267,606	0.97 ± 0.01	0.34 ± 0.02	0.97 ± 0.01	0.97 ± 0.02	0.34 ± 0.03	0.97 ± 0.02

Dataset	Depth	Parameters	Test Accuracy	Test Loss	Test Macro-F1	Validation Accuracy	Validation Loss	Validation Macro-F1
	L8	308,582	0.95 ± 0.02	0.38 ± 0.03	0.95 ± 0.02	0.94 ± 0.03	0.38 ± 0.05	0.94 ± 0.03

The quantitative results on both datasets show that increasing the decomposition depth improves the classification performance until a certain point, which is shown in Table 7. The test accuracy in the Brain Tumor MRI dataset increased from 0.61 ± 0.01 at L1 to 0.89 ± 0.01 at L7, and the test loss decreased from 1.05 ± 0.01 at L1 to 0.64 ± 0.02 at L7. However, performance dropped at L8, as test accuracy and test macro-F1 decreased to 0.85 ± 0.03, suggesting that excessive decomposition may weaken useful spatial information. The same tendency was followed with validation results, where the validation accuracy reached 0.95 ± 0.01, and the validation macro-F1 reached 0.95 ± 0.01 at L7 [46].

Similarly, the results in the Br35H dataset showed that the model, both in test accuracy and macro-F1, reached 0.97 ± 0.01 at L7, together with 0.34 ± 0.02 test loss. Validation results also peaked at L7 with validation accuracy and macro-F1 of 0.97 ± 0.02. As in the Brain Tumor MRI dataset, a small decrease was observed at L8, indicating again that the most effective configuration is achieved before the deepest decomposition level [46].

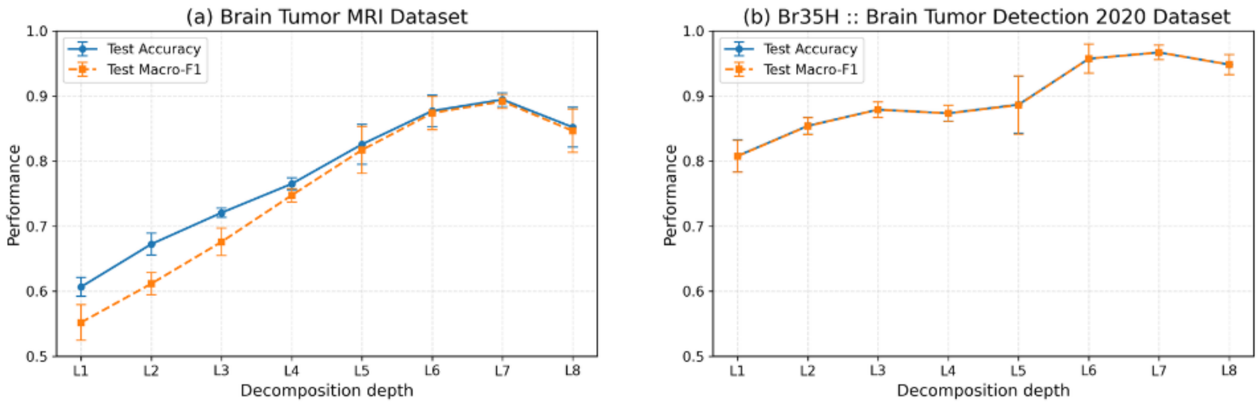

Fig. 11. Effect of Wavelet Decomposition Depth on Test Accuracy and Test Macro-F1 for (a) Brain Tumor MRI and (b) Br35H Datasets Across 5 Runs (Mean ± Standard Deviation)

The test accuracy and test macro-F1 curves in Fig. 11. follow a very similar pattern across both datasets. Both metrics in the Brain Tumor MRI and Br35H datasets increase gradually from L1 to L7. However, both curves show a decline at L8 and confirm again that L7 is the most suitable decomposition depth for both datasets.

3.5. Ablation Study of Directional Detail Subbands

After the wavelet decomposition depth ablation results, the WaveletFusion 6-level, 7-level, and 8-level models were selected for further analysis because they were the three best-performing decomposition depths on both datasets. To understand which directional components contribute most, different subband configurations were evaluated to compare the effect of using each detail direction separately and in combination.

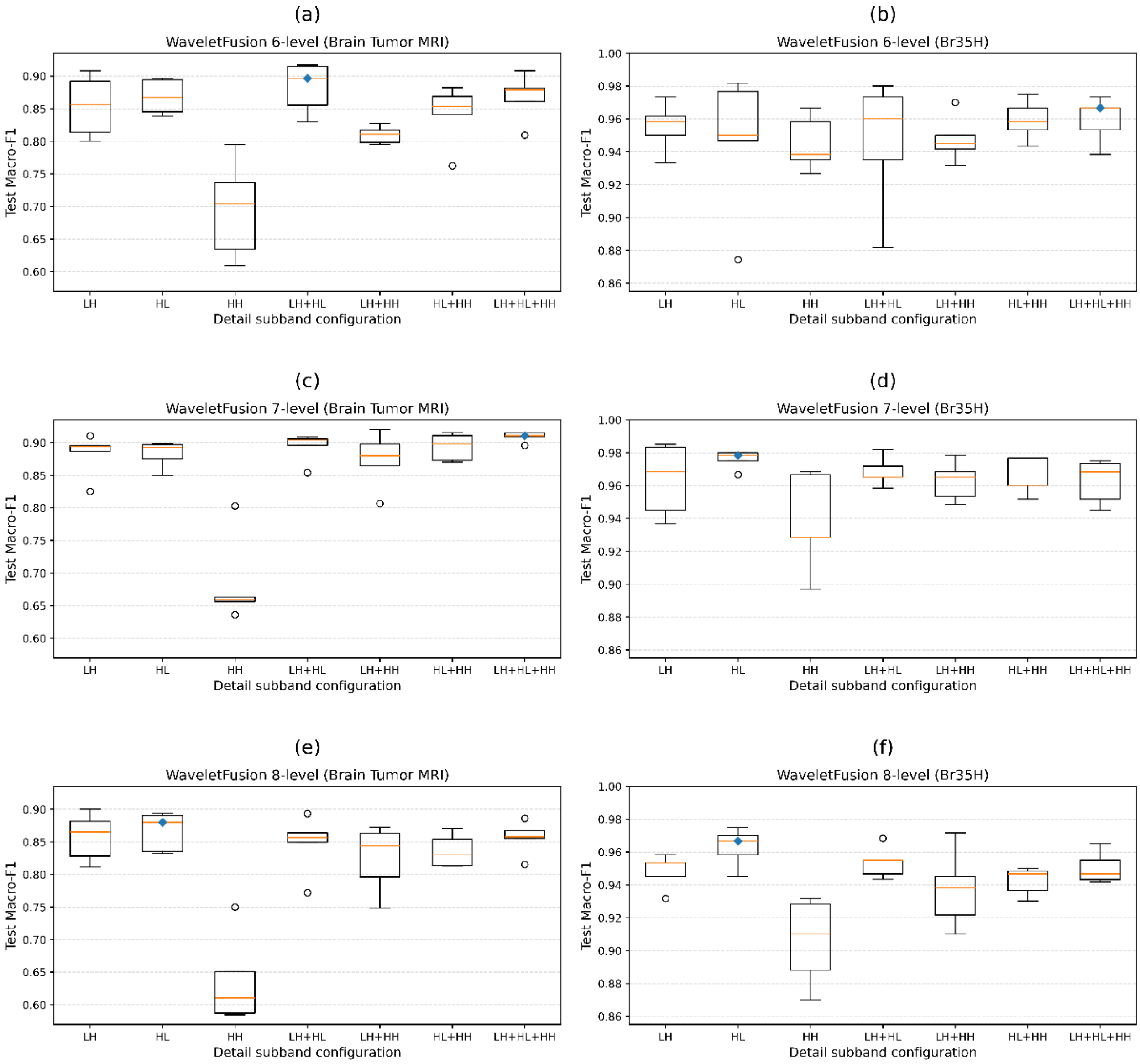

Fig. 12. Test Macro-F1 Distributions for Directional Detail Subband Configurations Across five runs: WaveletFusion 6-level on (a) Brain Tumor MRI and (b) Br35H, WaveletFusion 7-level on (c) Brain Tumor MRI and (d) Br35H, and WaveletFusion 8-level on (e) Brain Tumor MRI and (f) Br35H

For the Brain Tumor MRI dataset, the HH configuration yielded the weakest results across the three models, which is shown in Fig. 12. above, indicating that diagonal detail alone is insufficient. In contrast, LH + HL configurations produced stronger results, showing that they contain more useful discriminative features. The full LH + HL + HH configuration achieved the strongest and most stable result for the WaveletFusion 7-level model. The results for the Br35H dataset showed the same tendency, where the HH subband shows weaker results, especially in the WaveletFusion 8-level model, while the HL component and the rest of the combined configurations remain more reliable. The strongest result was achieved in the WaveletFusion 7-level model again with the HL configuration, reaching the highest test macro-F1 among the evaluated configurations.

The numerical results shown in Table 8 support the trends observed in the box plots for both datasets. For the Brain Tumor MRI dataset, the WaveletFusion 6-level model achieved its best test accuracy result with the LH + HL configuration (0.89 ± 0.04). The LH + HL + HH configuration in the

WaveletFusion 7-level model performed best, achieving 0.91 ± 0.01 of test accuracy and macro-F1 [46]. However, the LH + HL and HL + HH configurations also illustrated strong results at the 7-level. In contrast, the HH configuration produced much lower performance, especially at the 7-level and 8-level. Furthermore, the HL configuration in the WaveletFusion 7-level model achieved the highest test accuracy (0.98 ± 0.01) and test macro-F1 (0.98 ± 0.01) results. The combined LH + HL and HL + HH configurations also performed strongly at WaveletFusion 7-level, while the LH + HL + HH configuration did not outperform the HL subband configuration.

Table 8. Test Accuracy and Test Macro-F1 Results of Directional Detail Subband Configurations for L6, L7, and L8 WaveletFusion Models 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	0.86 ± 0.04	0.85 ± 0.05	0.89 ± 0.03	0.88 ± 0.03	0.86 ± 0.03	0.86 ± 0.04
	HL	0.87 ± 0.03	0.87 ± 0.03	0.89 ± 0.02	0.88 ± 0.02	0.87 ± 0.03	0.87 ± 0.03
	HH	0.74 ± 0.05	0.70 ± 0.08	0.71 ± 0.06	0.68 ± 0.07	0.69 ± 0.05	0.64 ± 0.07
	LH + HL	0.89 ± 0.04	0.88 ± 0.04	0.90 ± 0.02	0.89 ± 0.02	0.85 ± 0.04	0.85 ± 0.05
	LH + HH	0.82 ± 0.01	0.81 ± 0.01	0.88 ± 0.04	0.87 ± 0.04	0.83 ± 0.04	0.82 ± 0.05
	HL + HH	0.85 ± 0.04	0.84 ± 0.05	0.90 ± 0.02	0.89 ± 0.02	0.84 ± 0.02	0.84 ± 0.03
	LH + HL + HH	0.87 ± 0.03	0.87 ± 0.04	0.91 ± 0.01	0.91 ± 0.01	0.86 ± 0.02	0.86 ± 0.03
Br35H	LH	0.96 ± 0.01	0.96 ± 0.01	0.96 ± 0.02	0.96 ± 0.02	0.95 ± 0.01	0.95 ± 0.01
	HL	0.95 ± 0.04	0.95 ± 0.04	0.98 ± 0.01	0.98 ± 0.01	0.96 ± 0.01	0.96 ± 0.01
	HH	0.95 ± 0.02	0.95 ± 0.02	0.94 ± 0.03	0.94 ± 0.03	0.91 ± 0.03	0.91 ± 0.03
	LH + HL	0.95 ± 0.04	0.95 ± 0.04	0.97 ± 0.01	0.97 ± 0.01	0.95 ± 0.01	0.95 ± 0.01
	LH + HH	0.95 ± 0.01	0.95 ± 0.01	0.96 ± 0.01	0.96 ± 0.01	0.94 ± 0.02	0.94 ± 0.02
	HL + HH	0.96 ± 0.01	0.96 ± 0.01	0.97 ± 0.01	0.97 ± 0.01	0.94 ± 0.01	0.94 ± 0.01
	LH + HL + HH	0.96 ± 0.01	0.96 ± 0.01	0.96 ± 0.01	0.96 ± 0.01	0.95 ± 0.01	0.95 ± 0.01

Overall, the comparison between these three models supports the previous depth ablation results, where the best performance was again observed around the WaveletFusion 7-level configuration and can be further examined in a subsequent interpretability analysis.

3.6. Ablation Study of Level-group Contributions

After analyzing the effect of decomposition depth and directional detail subbands, an additional ablation study of level-group contributions was performed to evaluate which groups of wavelet decomposition levels contribute most to the final performance. The decomposition levels for the selected WaveletFusion 6-level, 7-level, and 8-level models were divided into four groups: Early, Intermediate, Coarse, and Full. The retained groups for WaveletFusion 6-level model were defined as L1-L2 (early-scale), L3-L4 (intermediate-scale), L5-L6 (coarse-scale), and L1-L6 (full scale). Similarly, the retained groups for the WaveletFusion 7-level model L1-L2 were defined as early-scale, L3-L5 as intermediate-scale, L6-L7 as coarse-scale, and L1-L7 as full-scale. Lastly, the retained groups for the WaveletFusion 8-level model were defined as early-scale (L1-L2), intermediate-scale (L3-L5), coarse-scale (L6-L8), and full-scale (L1-L8) [46].

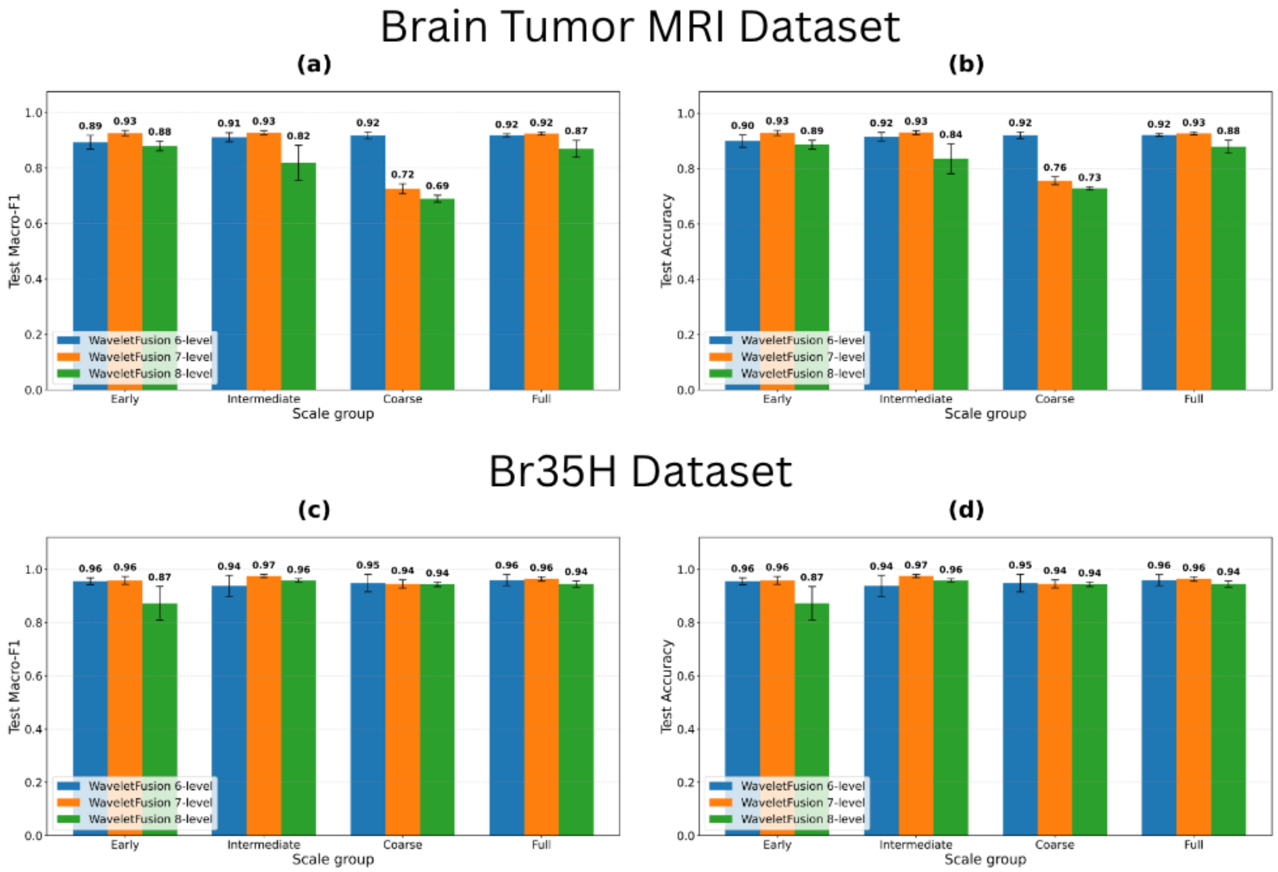

Fig. 13. Level-group Contribution Analysis Across Five Runs with (a) Test Macro-F1 on Brain Tumor MRI, (b) Test Accuracy on Brain Tumor MRI, (c) Test Macro-F1 on Br35H, and (d) Test Accuracy on Br35H

The WaveletFusion 7-level model in the Brain Tumor MRI dataset demonstrated that the early and intermediate-level groups provide the most useful information in these configurations, as shown in Fig. 13. above. In contrast, the coarse-scale group illustrated weaker results for the WaveletFusion 7-level and WaveletFusion 8-level models, indicating that deep low-resolution features alone are not sufficient for accurate tumor classification. At the same time, the best performance for the Br35H dataset was observed in the intermediate-scale and full-scale configurations. These results show that the full-scale configuration remains beneficial when the model combines information across multiple levels rather than relying solely on coarse features.

3.7. Explainable Analysis of the WaveletFusion 8-level Model

To better understand how the WaveletFusion 8-level model makes its decisions, a Grad-CAM explainability analysis was performed. This explainability analysis was carried out on representative test-set images from the four classes of the Brain Tumor MRI dataset and visualized across five successive convolution blocks (B1-B5). This made it possible to observe how the model's attention changes from early feature extraction layers to deeper decision-related layers.

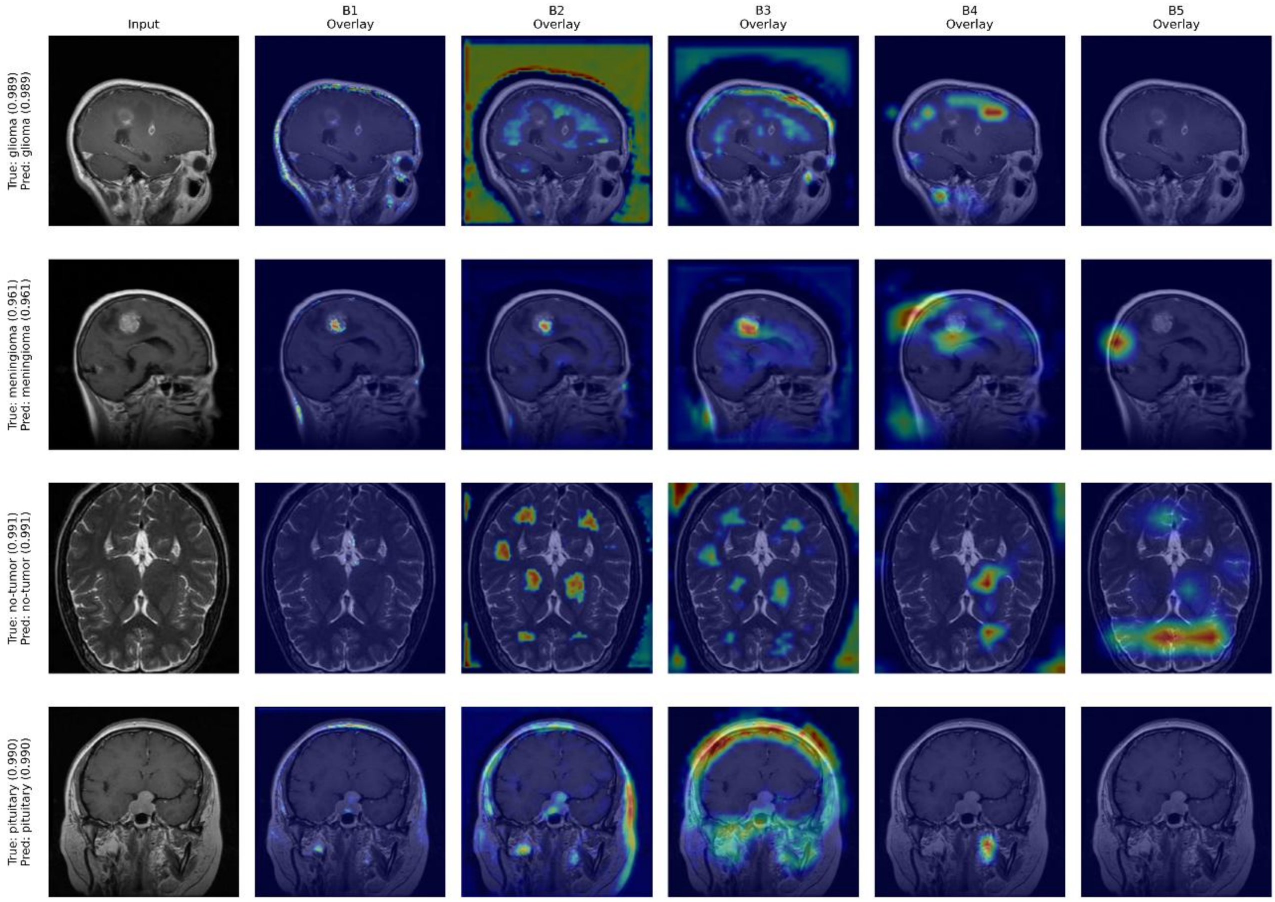

Fig. 14. Layer-wise Grad-CAM Visualization of Correctly Classified Examples for the WaveletFusion 8-level Model

The correctly classified examples in Fig. 14. show that in Block 1 (B1) and Block 2 (B2), the model mainly responds to general anatomical boundaries and edges. In the later blocks, like Block 4 (B4) and Block 5 (B5), the attention becomes more concentrated on class-relevant regions. For tumor classes, the deeper activations tend to move toward lesion-related areas, while for the no-tumor class, the attention remains more distributed. This indicates that the model gradually transforms low-level visual features into more class-specific representations.

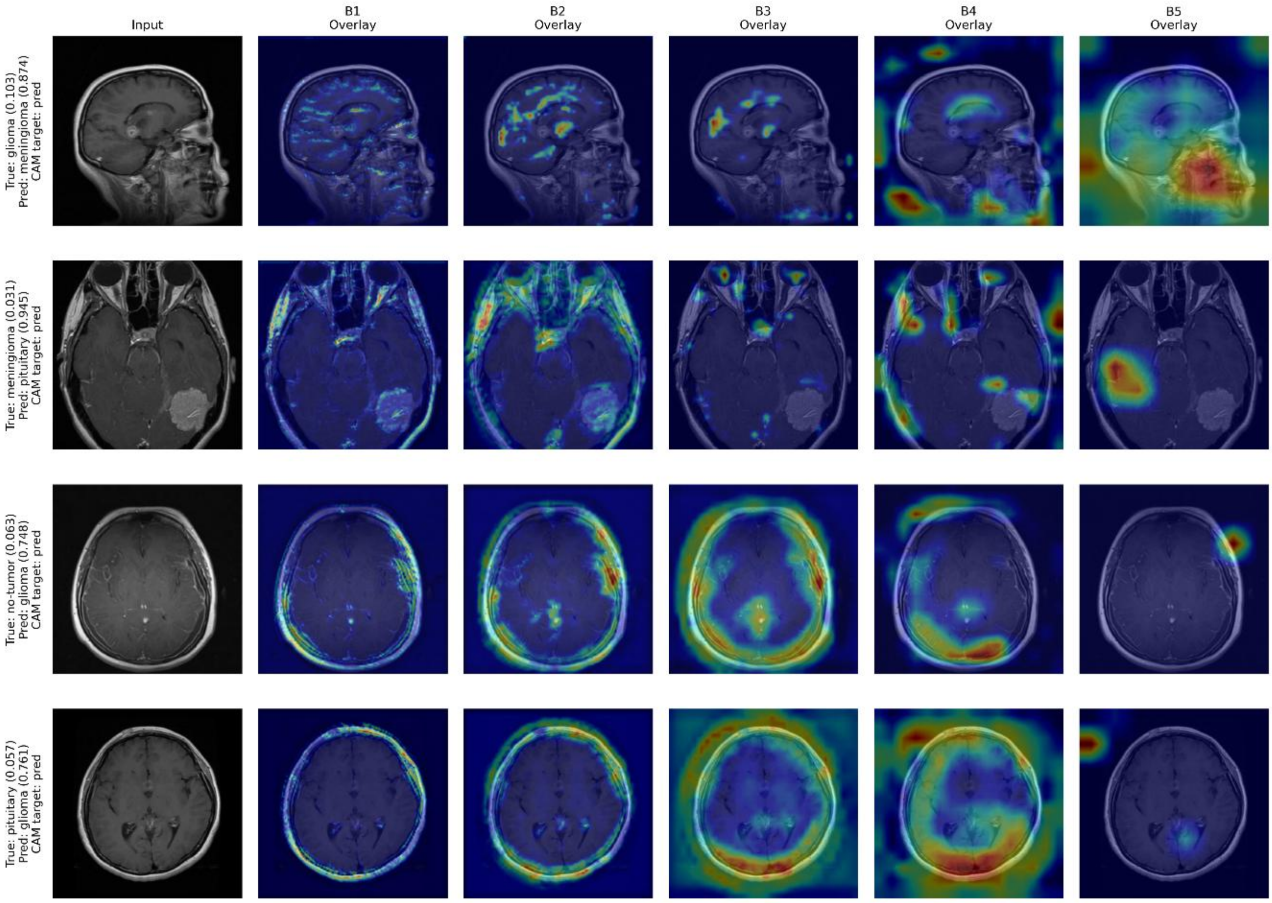

Fig. 15. Layer-wise Grad-CAM Visualization of Incorrectly Classified Examples for the WaveletFusion 8-level Model

The misclassified examples are shown in Fig. 15. above indicate that the deeper blocks in several cases produce strong activation over localized regions or anatomical structures that may visually resemble the incorrectly predicted class. For example, in B4 and B5, attention becomes more focused and decision oriented. Therefore, the misclassifications appear to occur when the model finds stronger discriminative evidence for the wrong class than for the correct one.

The same misclassified examples are illustrated in Fig. 16. below, but this time the Grad-CAM maps are generated with respect to the true class. Compared with the predicted-class visualizations, the true-class attention maps are generally weaker and more fragmented. This means that the model still contains some evidence for the correct class, but this evidence is not strong enough to control the final prediction.

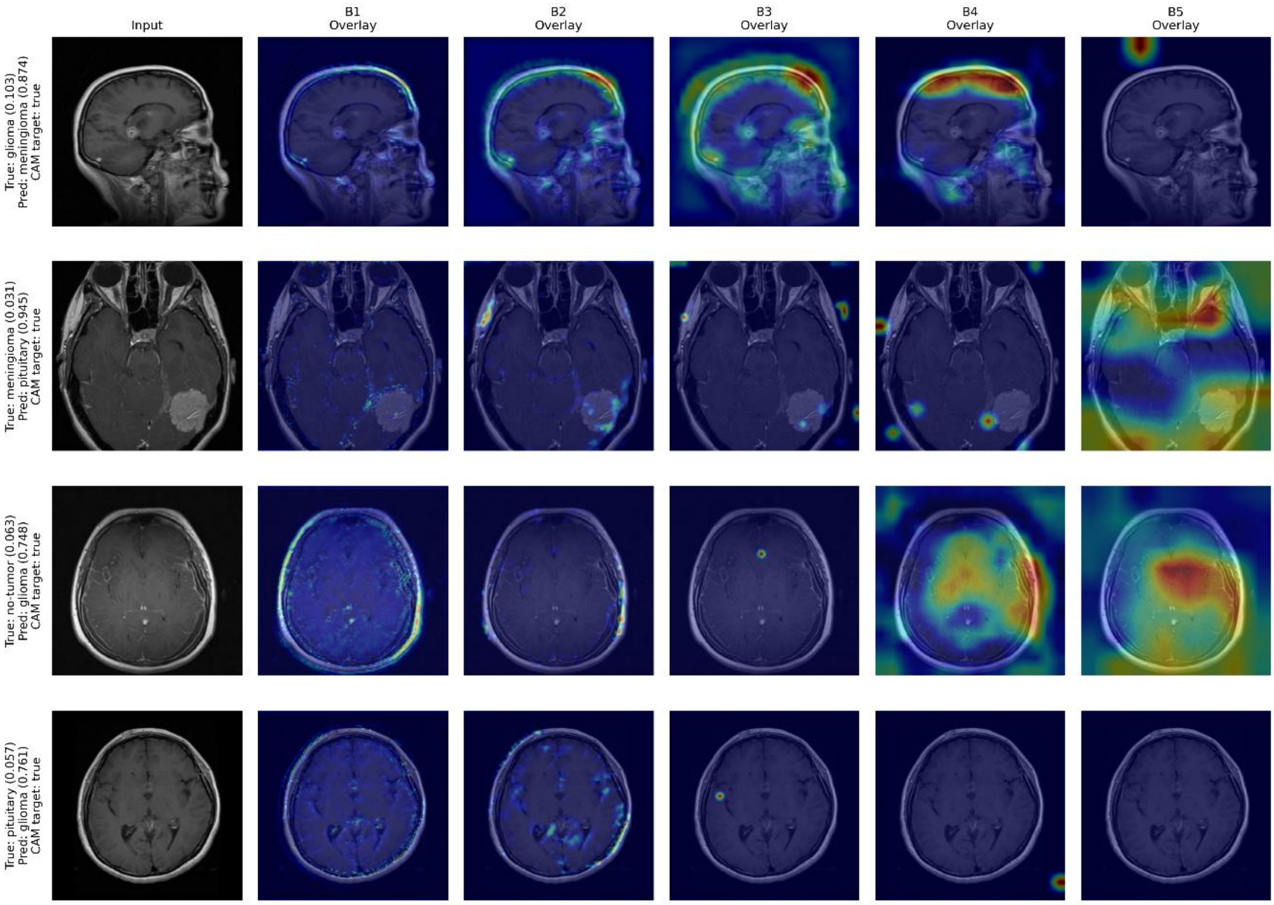

Fig. 16. Layer-wise Grad-CAM Visualization of the Incorrectly Classified Examples Targeted to the True Class for the WaveletFusion 8-level Model

In general, the explainability analysis shows that the WaveletFusion 8-level model uses different convolutional blocks in different ways. The results support the idea that the proposed architecture learns meaningful multi-scale representations. However, the misclassified cases show that visually similar tumor patterns or strong localized intensity changes can still mislead the model, especially when the deeper decision-related blocks focus on regions associated with an incorrect class.

3.8. Overall Benchmark Results and Comparison of the Models

Overall, the benchmark results show that the integration of multi-level decomposition improved the classification performance of the convolutional neural network models on both datasets, with a slightly smaller parameter count. The comparison in Table 9 includes Baseline CNN, WaveletFusion 7-level, and WaveletFusion 8-level models, because the ablation results showed these two WaveletFusion configurations provided the best metric results. However, further ablation analysis showed that the best overall performance was achieved by the WaveletFusion 7-level model (267,864 parameters for Brain Tumor MRI and 267,606 parameters for Br35H) than the WaveletFusion 8-level model, indicating that the larger model size did not lead to improved performance. Although the Baseline CNN for both datasets in terms of efficiency trains faster (average epoch time 6.03 for Brain Tumor MRI and average epoch time 2.68 for Br35H) than the WaveletFusion 7-level model because of additional multi-level wavelet processing. Similarly, the Baseline CNN model achieved lower inference latency (2.69 ms/img for Brain Tumor MRI and 2.49 ms/img for Br35H) and higher inference throughput (372.45 img/s for Brain Tumor MRI and 402.94 for Br35H) than the WaveletFusion models. Although the WaveletFusion variants showed lower FLOPs than the Baseline CNN, this indicates that such complexity does not fully capture the execution overhead.

Table 9. Computational Characteristics of the Baseline CNN and WaveletFusion Models on the Brain Tumor MRI and Br35H Datasets

Dataset	Model	Parameters	FLOPs (M)	Average Epoch Time	Train FPS (img/s)	Inference Latency (ms/img)	Inference Throughput (img/s)	Training Time
Brain Tumor MRI	Baseline CNN	311,700	1690.28	6.03	743.39	2.69	372.45	186.1
	WaveletFusion 7-level	267,864	819.05	9.76	461.30	4.24	235.88	276.2
	WaveletFusion 8-level	308,840	819.13	10.51	426.46	4.95	202.04	313.7
Br35H	Baseline CNN	311,442	1690.28	2.68	716.37	2.48	402.94	85.2
	WaveletFusion 7-level	267,606	819.05	4.73	406.19	4.70	212.64	141.2
	WaveletFusion 8-level	308,582	819.13	5.04	381.02	4.88	204.83	149.8

The performance metric results in Table 10 below illustrate that the WaveletFusion 7-level model achieved the best overall performance. The test accuracy and macro-F1 results on the Brain Tumor MRI dataset showed that the Baseline CNN reached 0.78 ± 0.02 test accuracy and 0.76 ± 0.03 test macro-F1. At the same time, the WaveletFusion 7-level model improved to 0.90 ± 0.01 for both test accuracy and test macro-F1. Moreover, the WaveletFusion 8-level model also outperformed the Baseline CNN but did not surpass the WaveletFusion 7-level model in both test accuracy and macro-F1 (0.85 ± 0.03). A similar trend was identified in the Br35H dataset, where WaveletFusion 7-level achieved 0.96 ± 0.01 in test accuracy and macro-F1, compared with 0.82 ± 0.03 for the Baseline CNN model [46].

Table 10. Overall Performance Comparison of the Baseline CNN and WaveletFusion Models

Dataset	Model	Test Accuracy	Test Loss	Test Macro-F1	Validation Accuracy	Validation Loss	Validation Macro-F1
Brain Tumor MRI	Baseline CNN	0.78 ± 0.02	0.78 ± 0.03	0.76 ± 0.03	0.85 ± 0.03	0.62 ± 0.05	0.84 ± 0.04
	WaveletFusion 7-level	0.90 ± 0.01	0.63 ± 0.02	0.90 ± 0.01	0.95 ± 0.01	0.47 ± 0.01	0.95 ± 0.01
	WaveletFusion 8-level	0.85 ± 0.03	0.73 ± 0.05	0.85 ± 0.03	0.92 ± 0.03	0.55 ± 0.06	0.91 ± 0.03
Br35H	Baseline CNN	0.82 ± 0.03	0.53 ± 0.03	0.82 ± 0.03	0.81 ± 0.02	0.54 ± 0.04	0.82 ± 0.04
	WaveletFusion 7-level	0.96 ± 0.01	0.35 ± 0.02	0.96 ± 0.01	0.96 ± 0.02	0.35 ± 0.03	0.96 ± 0.02
	WaveletFusion 8-level	0.95 ± 0.01	0.38 ± 0.02	0.95 ± 0.01	0.95 ± 0.01	0.37 ± 0.03	0.95 ± 0.01

To compare the existing approaches, Table 11 presents a comparison alongside the Baseline CNN and WaveletFusion variants. On the Brain Tumor MRI dataset, the WaveletFusion 7-level model outperformed traditional approaches such as VGG16, MobileNetV2, kNN Classifier, and ResNet50, achieving competitive results. However, DenseNet 169 and Fusion + CA + Chirplet Transform reported higher performance on this dataset. Similarly, the WaveletFusion 7-level model achieved strong performance in the Br35H dataset and performed competitively with ResNet101 and MobileNetV2 models. Although Pre-BN CNN and Post-BN Res CNN reported slightly higher results, the WaveletFusion 7-level model still demonstrates strong classification performance while using a wavelet-based multi-scale feature representation.

Table 11. Comparison of the Best-Run Results of the Proposed Models with Selected Existing Approaches on the Brain Tumor MRI and Br35H Datasets

Dataset	**Model**	**Reported Accuracy**	**Reported Macro-F1**
Brain Tumor MRI	Baseline CNN	0.74	0.71
	WaveletFusion 7-level	0.92	0.91
	WaveletFusion 8-level	0.88	0.88
	VGG 16 [32]	0.90	0.89[1]
	MobileNetV2 [32]	0.89	0.88[1]
	DenseNet169 [33]	0.98	0.96
	Fusion + CA + Chirplet Transform [33]	0.99	0.99
	kNN Classifier [34]	0.91	0.92[2]
	ResNet50 model [34]	0.91	0.92[2]
Br35H	Baseline CNN	0.79	0.78
	WaveletFusion 7-level	0.97	0.97
	WaveletFusion 8-level	0.95	0.95
	Pre-BN CNN [31]	0.98	0.98
	Post-BN Res CNN [31]	0.99	0.99
	ResNet101 [35]	0.96	0.96
	MobileNetV2 [35]	0.95	0.96
	Proposed CNN [47]	0.94	0.94
	InceptionV3 [47]	0.77	0.69

The per-class performance results in Table 12 show that Baseline CNN again performed worse on the Brain Tumor MRI dataset. This is especially visible in the meningioma class, indicating that Baseline CNN had difficulty distinguishing meningioma from other tumor classes. The WaveletFusion 7-level model significantly improved the results in all classes. The WaveletFusion 8-level model also provided significantly better results in all classes, but its performance was slightly lower than the 7-level configuration. Compared with DenseNet121, DenseNet169, and Fusion + CA + Chirplet Transform, the proposed WaveletFusion models achieved lower per-class results.

[1] Brain Tumor MRI 2026

[2] Weighted Macro-F1

Table 12. Best-run per-class Test Performance of the Baseline CNN and WaveletFusion Models, and Existing Approaches on the Brain Tumor MRI Dataset

Class	Metric	Baseline CNN	Wavelet Fusion 7-level	Wavelet Fusion 8-level	DenseNet121 [33]	DenseNet169 [33]	Fusion + CA + Chirplet Transform [33]
Glioma	Precision	0.6905	0.9496	0.9534	0.9900	0.9800	1.0000
	Recall	0.7250	0.8000	0.7675	0.9400	0.9400	0.9700
	F1-score	0.7073	0.8684	0.8504	0.9600	0.9600	0.9800
Meningio ma	Precision	0.7898	0.9164	0.8864	0.9400	0.9300	0.9700
	Recall	0.3100	0.8775	0.7800	0.9700	0.9700	0.9900
	F1-score	0.4452	0.8966	0.8298	0.9600	0.9500	0.9800
No-tumor	Precision	0.7322	0.8886	0.8411	1.0000	1.0000	1.0000
	Recall	0.9775	0.9975	0.9925	1.0000	1.0000	1.0000
	F1-score	0.8373	0.9399	0.9106	1.0000	1.0000	1.0000
Pituitary	Precision	0.7812	0.9188	0.8744	0.9800	1.0000	1.0000
	Recall	0.9550	0.9900	0.9925	0.9900	0.9900	1.0000
	F1-score	0.8594	0.9531	0.9297	0.9900	0.9900	1.0000

On the Br35H dataset, the per-class performance is shown in Table 13 below, where the Baseline CNN model achieved high recall for the "yes" class, but its precision was much lower at 0.7160. This means that the model detected many tumor cases correctly, but it also produced many false positive predictions. For the "no" class, the opposite behavior can be seen, where precision was high at 0.9439, but recall was only 0.6167. This shows that Baseline CNN did not classify the two classes in a balanced way. The WaveletFusion 7-level model produced much more balanced results for both classes. The WaveletFusion 8-level model also performed strongly, but its results were slightly lower than the 7-level model. Compared with the Pre-BN CNN, the WaveletFusion 7-level model achieved very similar results, while the Post-BN Res CNN remained the strongest model in this comparison. Overall, the results show that the WaveletFusion 7-level model greatly improves class balance compared with the Baseline CNN and remains competitive with existing CNN-based approaches [46].

Table 13. Best-run per-class Test Performance of the Baseline CNN and WaveletFusion models, and Existing Models on the Br35H Dataset

Class	Metric	Baseline CNN	WaveletFusion 7-level	WaveletFusion 8-level	Pre-BN CNN [31]	Post-BN Res CNN [31]
yes	Precision	0.7160	0.9702	0.9538	0.9679	0.9870
	Recall	0.9635	0.9701	0.9633	0.9902	0.9967
	F1-score	0.8215	0.9734	0.9535	0.9789	0.9918
no	Precision	0.9439	0.9766	0.9631	0.9896	0.9966
	Recall	0.6167	0.9701	0.9535	0.9661	0.9856
	F1-score	0.7460	0.9733	0.9583	0.9777	0.9915

The confusion matrix results in Fig. 17. were visualized to see where the models have more misclassifications. Overall, the Baseline CNN model produced more misclassifications than the WaveletFusion models. The largest improvement is especially visible in the meningioma class, where the Baseline CNN model misclassified most of the examples as glioma or no-tumor classes. Additionally, the improvement can be seen in the glioma class, where the WaveletFusion models produce fewer errors in neighboring classes. Correspondingly, Baseline CNN misclassified a larger number of no-tumor images as tumor cases in the Br35H dataset, while the WaveletFusion models produced a more balanced prediction distribution. Nevertheless, the confusion matrices again confirm that the WaveletFusion 7-level model provides the most stable classification behavior across both datasets [46].

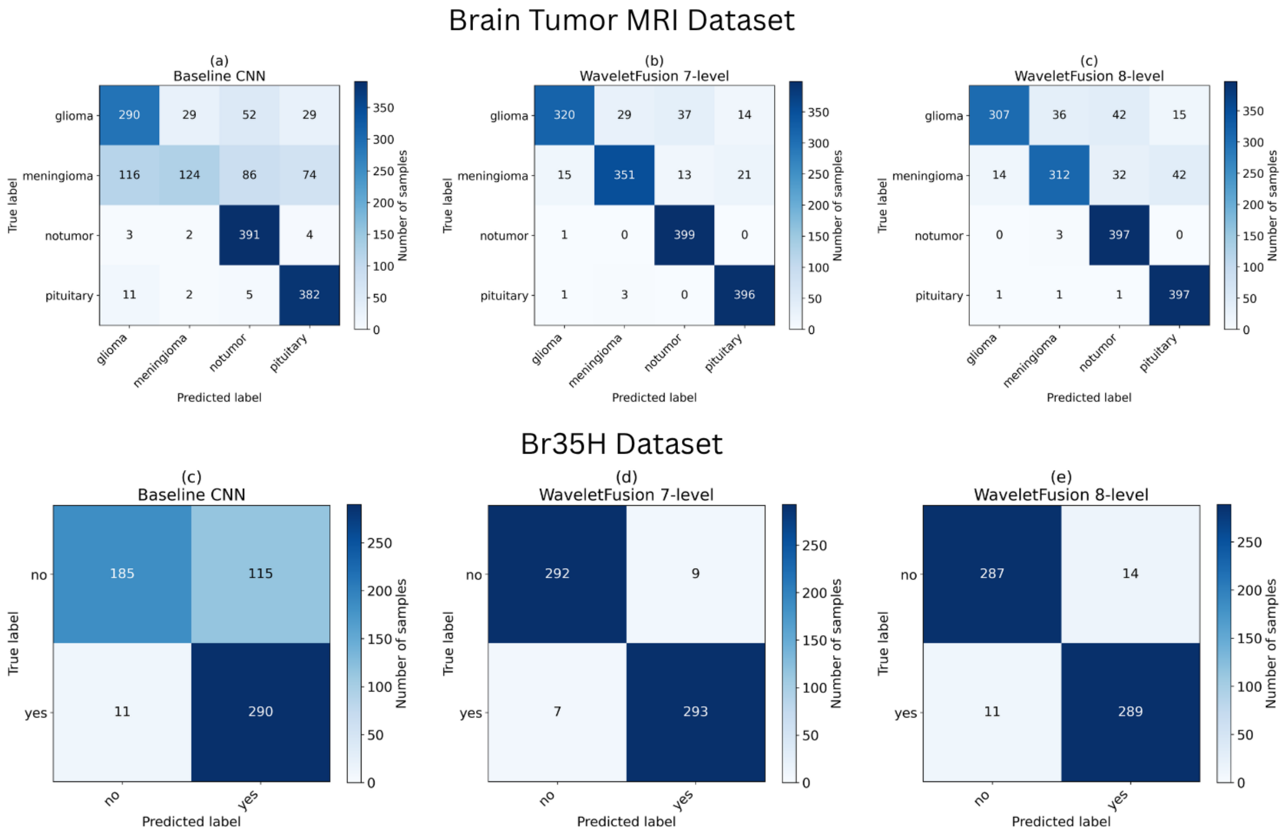

Fig. 17. Confusion Matrices of the Baseline CNN, WaveletFusion 7-level, and WaveletFusion 8-level Models for the Brain Tumor MRI and Br35H Datasets

The unified training accuracy and loss curves in Fig. 18. below shows the training behavior of three models. The WaveletFusion 7-level model on both datasets achieved faster and stronger learning compared with Baseline CNN. The training accuracy in the WaveletFusion 8-level model also increased more rapidly than the Baseline CNN, but it was less effective than the 7-level configuration. The corresponding training loss curves exhibit a similar tendency, with the WaveletFusion models exhibiting a smoother trajectory and faster loss reduction than the Baseline CNN.

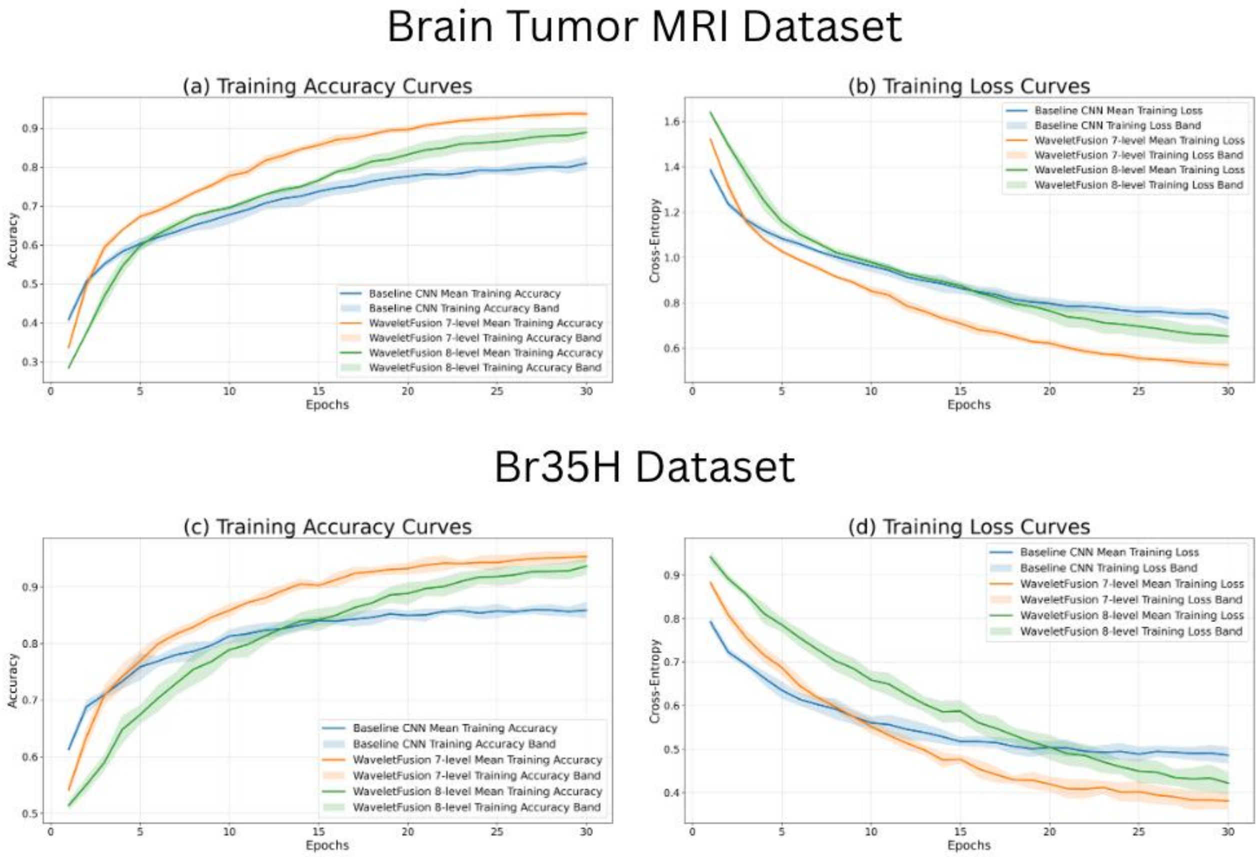

Fig. 18. Training Accuracy and Training Loss Curves for the Baseline CNN, WaveletFusion 7-level, and WaveletFusion 8-level Models on the Brain Tumor MRI and Br35H Datasets Across 5 Runs

Furthermore, unified validation accuracy and loss curves show that the WaveletFusion 7-level model generalizes better than the Baseline CNN on both datasets. As shown in Fig. 19 below, the 7-level configuration achieved higher validation accuracy and lower validation loss by the end of training. Baseline CNN's validation loss remained higher, while the WaveletFusion 8-level model also achieved better validation results than the baseline, but it was generally weaker than the 7-level model. Therefore, the validation curves confirm that the 7-level WaveletFusion model provides the best balance between learning capacity and generalization [46].

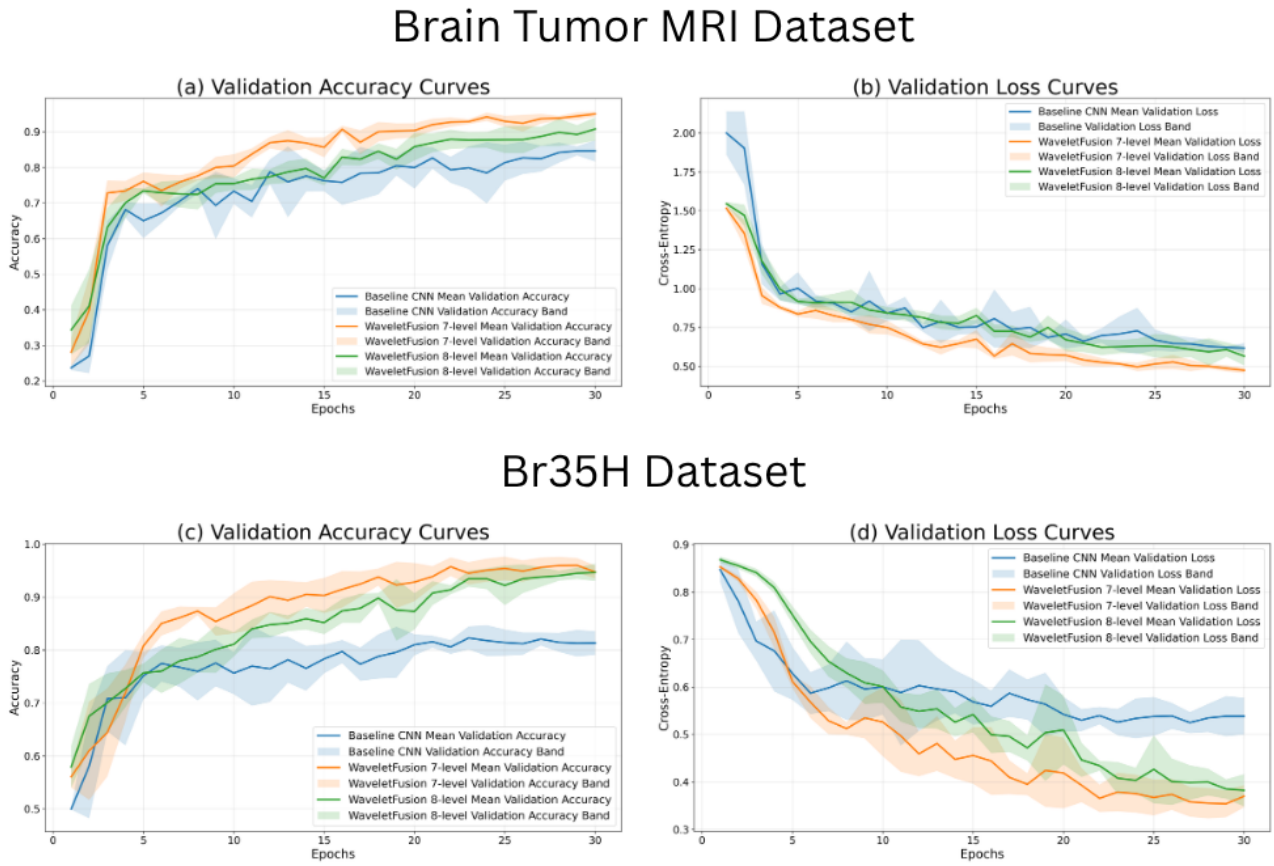

Fig. 19. Validation Accuracy and Validation Loss Curves for the Baseline CNN, WaveletFusion 7-level, and WaveletFusion 8-level Models on the Brain Tumor MRI and Br35H Datasets

Overall, the benchmark comparison results demonstrate that the WaveletFusion architecture improves classification performance compared with traditional convolutional neural network models. The strongest results were obtained by the WaveletFusion 7-level model, but it required more training time due to recursive decomposition and multi-scale fusion. In addition, the 7-level configuration provided a more compact parameter structure and better predictive performance. Therefore, the WaveletFusion 7-level model can be considered the most effective model among the evaluated architectures [46].

Conclusions

After completing the research project based on the interaction of wavelet transforms and artificial neural networks, the following conclusions are made:

1. Reviewing existing solutions illustrated that traditional image processing methods and machine learning approaches are helpful for image enhancement, segmentation, and feature extraction. Convolutional neural network models can automatically learn features, but they can still be sensitive to noise and information loss during downsampling. The existing methods also showed that wavelet transforms are useful, because they provide multi-resolution information and can preserve both structural and detail-related image features.
2. After identifying the limitations, a WaveletFusion architecture was designed that combines multi-level Haar wavelet decomposition with CNN-based feature extraction. The main limitations addressed in this research were the loss of high-frequency detail during downsampling, weak use of multi-scale information, and limited representation of texture and boundary-related features in standard CNN pipelines. The architecture uses both approximation and directional detail subbands, rather than traditional CNNs, allowing the model to learn information across different spatial and frequency scales. This design supports the use of low-frequency structural information together with high-frequency texture and boundary details.
3. The proposed architecture was implemented and evaluated on the Brain Tumor MRI and Br35H datasets. The experimental evaluation included decomposition depth analysis, directional detail subband analysis, level-group contribution analysis, benchmark comparison, and Grad-CAM explainability analysis. The results showed that increasing the wavelet decomposition depth improved classification performance but also indicated that excessive decomposition may lead to the loss of useful spatial information.
4. Comparing the obtained results with the Baseline CNN and previously mentioned architectures showed that the WaveletFusion 7-level model achieved the best overall performance among the proposed variants. On the Brain Tumor MRI dataset, the proposed solution improved test accuracy for the Baseline CNN model from 0.78 to 0.90 and test macro-F1 from 0.76 to 0.90. On the Br35H dataset, it improved both test accuracy and test macro-F1 from 0.82 to 0.96. The model also used fewer parameters than the Baseline CNN, reducing the parameter count to 267,864 on the Brain Tumor MRI dataset and 267,606 on the Br35H dataset. However, this improvement came with higher computational time, because recursive wavelet decomposition and multi-scale fusion increased training and inference time.

The findings indicate that future work could extend the proposed architecture by evaluating it on larger medical image datasets. In addition, other wavelet decomposition methods could be tested to improve the model by reducing inference time and optimizing the wavelet decomposition process for more efficient deployment. Finally, additional explainable artificial intelligence methods could be applied to further validate whether the model focuses on meaningful regions during the classification of medical images.

Acknowledgement

First and foremost, I would like to express my deepest gratitude to my supervisor, Dr. Dalia Čalnerytė, for her infinite support and encouragement in this master‘s thesis. During the various stages of this entire work, her guidance, patience, and recommendations helped me through the scientific conference and first article publication for which I am enormously thankful. Without her professional assistance and support, I do not believe this paperwork would have ever been accomplished. I must add that it was a great experience working with her in these last three years, and I recommend every student to collaborate at least once with her.

To Dr. Armantas Ostreika for giving suggestions and clarifications during the semester projects. His feedback and clarifications during the screenings helped me to make this master‘s thesis work unique and valuable in the end.

To linguist Dr. Virginija Tuomaitė for providing English consultations about grammatical errors and evaluating the quality of literacy of my master’s thesis. Her recommendations raised the quality of this work, ensuring its success.

Lastly, I would like to thank my family, especially my parents, Mr. Vasif Ismayilov and Mrs. Yegana Ismayilova, for their continued love and support. You are the only ones who saw all my successes and failures during my master’s journey, believed in me until the end, and always wished me all the best. Besides the fact that over the last five and a half years, I have been too far away from you, I am very thankful that you have supported my dreams from the first day of this journey. I hope that one day this distance between us will no longer exist, and we will all be reunited again.

As I close this chapter, I would also like to express my sincere gratitude to the Department of Applied Informatics once again for their motivation to grow professionally throughout my studies. Although many names may remain unmentioned, I am deeply grateful to everyone who became part of this journey and helped me reach its completion.

List of references

1. NEETU RANI. Image Processing Techniques: A Review. *Journal on Today's Ideas - Tomorrow's Technologies* [online]. 2017, **5**(1), 40–49 [viewed 11 April 2026]. ISSN 2321-7146. Available from: doi:10.15415/jotitt.2017.51003
2. PAL, Nikhil R., and Sankar K. PAL. A review on image segmentation techniques. *Pattern Recognition* [online]. 1993, **26**(9), 1277–1294 [viewed 11 April 2026]. ISSN 0031-3203. Available from: doi:10.1016/0031-3203(93)90135-j
3. CANNY, John. A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* [online]. 1986, **PAMI-8**(6), 679–698 [viewed 11 April 2026]. ISSN 0162-8828. Available from: doi:10.1109/tpami.1986.4767851
4. HARALICK, Robert M., K. SHANMUGAM, and Its'Hak DINSTEIN. Textural Features for Image Classification. *IEEE Transactions on Systems, Man, and Cybernetics* [online]. 1973, **SMC-3**(6), 610–621 [viewed 11 April 2026]. ISSN 2168-2909. Available from: doi:10.1109/tsmc.1973.4309314
5. LOWE, David G. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision* [online]. 2004, **60**(2), 91–110 [viewed 11 April 2026]. ISSN 0920-5691. Available from: doi:10.1023/b:visi.0000029664.99615.94
6. DALAL, N., and B. TRIGGS. Histograms of Oriented Gradients for Human Detection. In: *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)* [online]. IEEE, [viewed 11 April 2026] [viewed 11 April 2026]. ISBN 0769523722. Available from: doi:10.1109/cvpr.2005.177
7. COVER, T., and P. HART. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory* [online]. 1967, **13**(1), 21–27 [viewed 11 April 2026]. ISSN 1557-9654. Available from: doi:10.1109/tit.1967.1053964
8. CORTES, Corinna, and Vladimir VAPNIK. Support-vector networks. *Machine Learning* [online]. 1995, **20**(3), 273–297 [viewed 11 April 2026]. ISSN 1573-0565. Available from: doi:10.1007/bf00994018
9. BREIMAN, Leo. Random Forests. *Machine Learning* [online]. 2001, **45**, 5–32 [viewed 11 April 2026]. ISSN 0885-6125. Available from: doi:10.1023/A:1010933404324
10. LECUN, Yann, Yoshua BENGIO, and Geoffrey HINTON. Deep learning. *Nature* [online]. 2015, **521**(7553), 436–444 [viewed 11 April 2026]. ISSN 1476-4687. Available from: doi:10.1038/nature14539
11. KRIZHEVSKY, Alex, Ilya SUTSKEVER, and Geoffrey E. HINTON. ImageNet Classification with Deep Convolutional Neural Networks. In: *Advances in Neural Information Processing Systems 25 (NIPS 2012)* [online]. 2012, 1097–1105 [viewed 11 April 2026]. Available from: doi:10.5555/2999134.2999257
12. RONNEBERGER, Olaf, Philipp FISCHER, and Thomas BROX. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* [online]. 2015, 234–241 [viewed 11 April 2026]. ISBN 978-3-319-24574-4. Available from: doi:10.1007/978-3-319-24574-4_28
13. LUNDERVOLD, Alexander Selvikvåg, and Arvid LUNDERVOLD. An overview of deep learning in medical imaging focusing on MRI. *Zeitschrift für Medizinische Physik* [online].

2019, **29**(2), 102–127 [viewed 11 April 2026]. ISSN 0939-3889. Available from: doi:10.1016/j.zemedi.2018.11.002

14. RIZWAN, Muhammad, et al. Brain Tumor and Glioma Grade Classification Using Gaussian Convolutional Neural Network. *IEEE Access* [online]. 2022, **10**, 29731–29740 [viewed 11 April 2026]. ISSN 2169-3536. Available from: doi:10.1109/access.2022.3153108
15. DISCI, Rukiye, Fatih GURCAN, and Ahmet SOYLU. Advanced Brain Tumor Classification in MR Images Using Transfer Learning and Pre-Trained Deep CNN Models. *Cancers* [online]. 2025, **17**(1), 121 [viewed 11 April 2026]. ISSN 2072-6694. Available from: doi:10.3390/cancers17010121
16. YAGIS, Ekin, et al. Effect of data leakage in brain MRI classification using 2D convolutional neural networks. *Scientific Reports* [online]. 2021, **11**(1) [viewed 11 April 2026]. ISSN 2045-2322. Available from: doi:10.1038/s41598-021-01681-w
17. MONGAN, John, Linda MOY, and Charles E. KAHN. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): A Guide for Authors and Reviewers. *Radiology: Artificial Intelligence* [online]. 2020, **2**(2), e200029 [viewed 11 April 2026]. ISSN 2638-6100. Available from: doi:10.1148/ryai.2020200029
18. MALLAT, Stéphane. *Wavelet Tour of Signal Processing*. Elsevier Science & Technology Books, 1999. ISBN 9780080520834.
19. FAYAZ, Muhammad, et al. An Efficient Methodology for Brain MRI Classification Based on DWT and Convolutional Neural Network. *Sensors* [online]. 2021, **21**(22), 7480 [viewed 11 April 2026]. ISSN 1424-8220. Available from: doi:10.3390/s21227480
20. GUO, Wenhui, et al. Hyperspectral image classification using CNN-enhanced multi-level Haar wavelet features fusion network. *IEEE Geoscience and Remote Sensing Letters* [online]. 2022, 1 [viewed 11 April 2026]. ISSN 1558-0571. Available from: doi:10.1109/lgrs.2022.3167535
21. MASQUELIN, Axel H., et al. Wavelet decomposition facilitates training on small datasets for medical image classification by deep learning. *Histochemistry and Cell Biology* [online]. 2021, **155**(2), 309–317 [viewed 11 April 2026]. ISSN 1432-119X. Available from: doi:10.1007/s00418-020-01961-y
22. LI, Qiufu, et al. WaveCNet: Wavelet Integrated CNNs to Suppress Aliasing Effect for Noise-Robust Image Classification. *IEEE Transactions on Image Processing* [online]. 2021, **30**, 7074–7089 [viewed 11 April 2026]. ISSN 1941-0042. Available from: doi:10.1109/tip.2021.3101395
23. ABD EL KADER, Isselmou, et al. Brain Tumor Detection and Classification on MR Images by a Deep Wavelet Auto-Encoder Model. *Diagnostics* [online]. 2021, **11**(9), 1589 [viewed 11 April 2026]. ISSN 2075-4418. Available from: doi:10.3390/diagnostics11091589
24. LIU, Pengju, et al. Multi-Level Wavelet Convolutional Neural Networks. *IEEE Access* [online]. 2019, **7**, 74973–74985 [viewed 11 April 2026]. ISSN 2169-3536. Available from: doi:10.1109/access.2019.2921451
25. SAVAREH, Behrouz Alizadeh, et al. Wavelet-enhanced convolutional neural network: a new idea in a deep learning paradigm. *Biomedical Engineering / Biomedizinische Technik* [online]. 2019, **64**(2), 195–205 [viewed 11 April 2026]. ISSN 1862-278X. Available from: doi:10.1515/bmt-2017-0178
26. ALZUBAIDI, Laith, et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data* [online]. 2021, **8**(1) [viewed 11 April 2026]. ISSN 2196-1115. Available from: doi:10.1186/s40537-021-00444-8

27. GEEKSFORGEEKS. Introduction to Convolution Neural Network - GeeksforGeeks. *GeeksforGeeks* [online]. 21 August 2017 [viewed 12 May 2026]. Available from: https://www.geeksforgeeks.org/machine-learning/introduction-convolution-neural-network/
28. AZULAY, Aharon and WEISS, Yair. *Why do deep convolutional networks generalize so poorly to small image transformations?* Journal of Machine Learning Research [online]. 2019, 20, 1–25 [viewed 11 April 2026]. Available from: https://jmlr.org/papers/v20/19-519.html
29. LUO, Wenjie, LI, Yujia, URTASUN, Raquel and ZEMEL, Richard. Understanding the Effective Receptive Field in Deep Convolutional Neural Networks [online]. In: *Proceedings of the 30th Conference on Neural Information Processing Systems (NIPS 2016)*. Barcelona, Spain, 2016 [viewed 11 April 2026]. Available from: https://papers.nips.cc/paper_files/paper/2016/hash/c8067ad1937f728f51288b3eb986afaa-Abstract.html
30. SHARMA, Shagun, and Kalpna GULERIA. Deep Learning Models for Image Classification: Comparison and Applications. In: *2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)* [online]. IEEE, 2022 [viewed 13 May 2026]. Available from: doi:10.1109/icacite53722.2022.9823516
31. NIZAMLI, Yasser, et al. A Lightweight CNN Architecture for Efficient Brain Tumor Detection in MRI Scans. *International Journal of Electrical and Electronics Research* [online]. 2025, **13**(2), 296–305 [viewed 13 May 2026]. ISSN 2347-470X. Available from: doi:10.37391/ijeer.130213
32. AHMAD, Fahad, et al. NeuroAssist: A Hybrid SEMoNet Framework for Brain Tumour Classification From MRI With Patient-Centred Conversational Support. *Applied Computational Intelligence and Soft Computing* [online]. 2026, **2026**(1) [viewed 13 May 2026]. ISSN 1687-9732. Available from: doi:10.1155/acis/7216102
33. HEKMAT, Arash, et al. Brain tumor diagnosis redefined: Leveraging image fusion for MRI enhancement classification. *Biomedical Signal Processing and Control* [online]. 2025, **109**, 108040 [viewed 13 May 2026]. ISSN 1746-8094. Available from: doi:10.1016/j.bspc.2025.108040
34. DHINGRA, Mrigank, and Vikas DHINGRA. Assessing machine learning model efficacy for brain tumor MRI classification: a multi-model approach. *Journal of Emerging Investigators* [online]. 2026 [viewed 13 May 2026]. Available from: doi:10.59720/25-127
35. ELEYAN, Gülden, et al. Brain Tumor Detection Via Ensemble CNN-Based Deep Learning Models. In: *2025 6th International Conference on Bio-engineering for Smart Technologies (BioSMART)* [online]. IEEE, 2025, pp. 1–4 [viewed 13 May 2026]. Available from: doi:10.1109/biosmart66413.2025.11046093
36. SHAP: A Comprehensive Guide to SHapley exPlanations [online]. GeeksforGeeks. 14 July 2025 [viewed 11 April 2026]. Available from: https://www.geeksforgeeks.org/machine-learning/shap-a-comprehensive-guide-to-shapley-additive-explanations/
37. DEV, Vikram. A Complete Guide to SHAP Values [online]. Medium, 21 July 2024 [viewed 13 May 2026]. Available from: https://easyml.medium.com/a-complete-guide-to-shap-values-ae2de0d38fbd
38. ERDEM, Kemal. XAI Methods — Integrated Gradients [online]. Medium. 21 April 2022 [viewed 11 April 2026]. Available from: https://medium.com/@kemalpiro/xai-methods-integrated-gradients-6ee1fe4120d8

39. SUN, Yi, and Mukund SUNDARARAJAN. Axiomatic attribution for multilinear functions. In: *the 12th ACM conference* [online]. New York, New York, USA: ACM Press, 2011 [viewed 13 May 2026]. ISBN 9781450302616. Available from: doi:10.1145/1993574.1993601
40. NICKPARVAR, Masoud. Brain Tumor MRI Dataset [dataset online]. Kaggle [viewed 11 April 2026]. Available from: https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset
41. Br35H :: Brain Tumor Detection 2020. Kaggle: The World's AI Proving Ground [online]. [no date] [viewed 7 May 2026]. Available from: https://www.kaggle.com/datasets/ahmedhamada0/brain-tumor-detection
42. TALUKDER, Kamrul Hasan and HARADA, Koichi. Haar Wavelet Based Approach for Image Compression and Quality Assessment of Compressed Image [online]. arXiv, 2010 [viewed 11 April 2026]. Available from: https://arxiv.org/abs/1010.4084
43. PARIDA, Priyadarsan, and Nilamani BHOI. Wavelet based transition region extraction for image segmentation. Future Computing and Informatics Journal [online]. 2017, 2(2), 65–78 [viewed 11 April 2026]. ISSN 2314-7288. Available from: doi:10.1016/j.fcij.2017.10.005
44. H.L., Sneha. 2D Convolution in Image Processing [online]. All About Circuits, 30 November 2018 [viewed 12 April 2026]. Available from: https://www.allaboutcircuits.com/technical-articles/two-dimensional-convolution-in-image-processing/
45. KIM, Dong-Keon. Grad-CAM: A Gradient-based Approach to Explainability in Deep Learning [online]. Medium, 17 February 2025 [viewed 11 April 2026]. Available from: https://medium.com/@kdk199604/grad-cam-a-gradient-based-approach-to-explainability-in-deep-learning-871b3ab8a6ce
46. ISMAYILOV, Mahammad, and Dalia ČALNERYTĖ. Integration of Multi-Level Wavelet Decomposition and CNN for Brain Tumor MRI Classification. Applied Sciences [online]. 2026, 16(9), 4482 [viewed 19 May 2026]. ISSN 2076-3417. Available from: doi:10.3390/app16094482
47. ULLAH, Atta, et al. An Explainable Deep Learning Framework for Brain Tumor Detection Using MRI Images. *VFAST Transactions on Software Engineering* [online]. 2026, **14**(1), 278–300 [viewed 13 May 2026]. ISSN 2309-3978. Available from: doi:10.21015/vtse.v14i1.2335

Appendices

Appendix 1. Lithuanian MSc Research in Informatics and ICT 2025 Publication

Ismayilov M., Čalnerytė D. (2025). *Integration of Image Decomposition Methods and CNN for Image Classification. doi: 10.15388/LMITT.2025.6.*

Appendix 2. Engineering, Multidisciplinary Journal Publication – Applied Sciences (MDPI)

Ismayilov M., Čalnerytė D. (2026). *Integration of Multi-Level Wavelet Decomposition and CNN for Brain Tumor MRI Classification. doi: 10.3390/app16094482.*