Multi-Model Deep Feature Fusion for Robust Detection of Neuro-Oncological Abnormalities

Yasser Nizamli[1,2]*, Anton Filatov[1], Weaam Fadel[3], Yulia Shichkina[3], Kinda Mreish[4,5], Nahida Karaja[1], Tarek Alnajar[6]

[1]*Department of Software Engineering and Computer Applications, Saint Petersburg Electrotechnical University "LETI", Saint Petersburg, Russia; Email: yanizamli@stud.etu.ru*

[2]*Department of Computer Engineering and Automatic Control, Latakia University, Latakia, Syria*

[3]*Department of Computer Engineering, Saint Petersburg Electrotechnical University "LETI", Saint Petersburg, Russia*

[4]*Department of Computer Engineering, Aleppo University, Aleppo, Syria*

[5]*Department of Information Systems, Saint Petersburg Electrotechnical University "LETI", Saint Petersburg, Russia*

[6]*Department of Automatic Control Systems, Saint Petersburg Electrotechnical University "LETI", Saint Petersburg, Russia*

***Correspondence:** *Yasser Nizamli*, Email: *yanizamli@stud.etu.ru*

ABSTRACT- Accurate and early diagnosis of brain tumors can significantly reduce both the invasiveness and cost of therapeutic interventions while preserving neurological function. Current limitations in manual MRI analysis, including human error, variability in expertise, and interpretive inconsistencies, have created a pressing need for advanced diagnostic systems based on artificial intelligence and deep learning techniques. This study presents a hybrid transfer learning approach designed to enhance the detection of neuro-oncological abnormalities. Our methodology employs parallel processing of MRI images through pre-trained DenseNet121 and VGG16 architectures to extract discriminative numerical features. These feature sets are integrated and then processed through principal component analysis to improve computational efficiency. For the classification stage, we implement both support vector machine and k-nearest neighbors algorithms independently. A comprehensive evaluation, including an analysis of data processing sequences to ensure methodological rigor, demonstrates that the proposed feature fusion framework achieves robust performance and competitive accuracy, exceeding the performance of several contemporary deep learning models in this domain.

Keywords: Anomaly detection, Brain tumors, Medical Imaging, MRI Images, Deep learning, Feature fusion.

ARTICLE INFORMATION

Author(s): Yasser Nizamli, Anton Filatov, Weaam Fadel, Yulia Shichkina, Kinda Mreish, Nahida Karaja, and Tarek Alnajar;

Received: 03/10/2025; **Accepted:** 13/11/2025; **Published:** 15/12/2025;
E- ISSN: 2347-470X;
Paper Id: IJEER 0310A05;
Citation: 10.37391/ijeer.130420
Webpage-link:
https://ijeer.forexjournal.co.in/archive/volume-13/ijeer-130420.html

Publisher's Note: FOREX Publication stays neutral with regard to jurisdictional claims in Published maps and institutional affiliations.

1. INTRODUCTION

The brain is an organ located in the skull that controls and coordinates all processes in the human body. This specialized tissue can be affected by various diseases, among which tumors are particularly serious. A brain tumor is an abnormal growth of cells within the brain or central spinal canal, which may be either benign or malignant. These tumors develop from different cell types, including glial, meningeal, or pituitary cells [1, 2, 3].

Symptoms and treatment options vary depending on the tumor's size and type. Common symptoms include headaches, cognitive impairment, vision problems, and motor dysfunction. Treatment may involve surgery, radiation therapy, or chemotherapy [2, 4]. Magnetic resonance imaging (MRI) is the preferred diagnostic method for brain tumors due to its excellent soft tissue visualization. Clinicians analyze MRI images to assess pathological conditions and determine appropriate treatments. However, this process depends heavily on human expertise and faces limitations such as time constraints, potential errors, and high costs [5, 6, 7].

To address these challenges, researchers are developing automated systems using artificial intelligence and machine learning technologies to improve diagnostic accuracy and efficiency [8, 9]. Current approaches for brain tumor classification in MRI images predominantly employ deep convolutional neural networks (CNNs) or leverage established benchmark models through transfer learning.

In [10], a hybrid system combining deep feature integration with classical machine learning was introduced. MRI images underwent preprocessing to reduce noise and improve contrast before being processed by a specialized network with five convolutional layers, trained to extract deep numerical representations. An SVM algorithm then classified the extracted features into one of three central nervous system (CNS) tumor types: meningioma, glioma, or pituitary. The model achieved 96% accuracy, surpassing the performance of VGG16, GoogLeNet, and AlexNet models.

The authors of [11] designed a CNN model with three convolutional-pooling blocks followed by a four-layer fully connected neural network to detect brain anomalies. Their system reached a test accuracy of 95.87%, while training accuracy surpassed 99%.

In [12], a specialized architecture consisting of three convolutional blocks and a neural classifier was proposed. Using data augmentation, the model achieved a training accuracy of 99.31%, though validation accuracy remained at 93.1%. The researchers claimed their approach outperformed traditional feature extraction methods such as Local Binary Pattern (LBP) and Gray Level Co-occurrence Matrix (GLCM).

The work in [13] presented a complex system with five sequential feature extraction blocks, each containing three parallel convolutional layers. Their outputs were merged and processed by a fourth convolutional layer before down sampling through both max and average pooling. The combined features were then passed to the next block. After the final block, global average pooling was applied, followed by a fully connected three-layer neural network for classification. The model reached an accuracy of 96.4%.

To avoid the challenge of designing custom architectures, some studies employ standard pre-trained models. For instance, [14] proposed a brain tumor classification system using InceptionV3 and Xception for feature extraction, followed by an ensemble classifier (KNN, SVM, and RF). InceptionV3 slightly outperformed Xception, achieving 94.34% accuracy compared to 93.79%.

Following a methodology similar to [14], the authors of [15] developed a system where features were extracted using DenseNet169, and classification relied on majority voting among an ensemble (RF, SVM, and XGBoost). The system was benchmarked against multiple deep models, including ResNet, VGG, and EfficientNet variants, achieving 95.10% accuracy with data augmentation.

In [16], the researchers fine-tuned the You Only Look Once framework (YOLOv8) to detect brain tumors in MRI scans. The system attained an F1-score of 91.38% without augmentation and 92.47% with augmentation, calculated from the provided precision and recall values.

Another study [17] also utilized YOLO, specifically YOLOv5, for neuro-oncological lesion detection, achieving an inferred F1-score of 86.84%. The network was later enhanced with an improved spatial attention (ESA) layer, boosting performance to an F1-score of 89.85%.

The development of specialized deep models with numerous parameters trained on relatively small MRI datasets may lead to overfitting, where high training performance fails to generalize to validation data. While transfer learning eliminates the need for extensive parameterization, differences in task domains can limit the generation of effective features capable of achieving the required high performance. Additionally, producing an excessive number of features may negatively impact the final classification stage. This work aims to address these challenges through the following contributions:

- **Proposing a hybrid end-to-end system for enhanced recognition of central nervous system lesions in MRI images.**
- **Investigating the effectiveness of fusing high-level features obtained from pre-trained deep models.**
- **Validating the impact of applying class balancing in the feature space on the overall performance of the final classification model.**
- **Conducting an ablation study with multiple metrics to quantify each component's contribution in the proposed system pipeline.**

2. MATERIALS AND METHODS

Figure 1 outlines the workflow of the proposed hybrid system. The process consists of the following key steps:

1. Image acquisition and preprocessing: MRI images are loaded, processed, and prepared for feature extraction.
2. Deep feature extraction: Two parallel pre-trained models, DenseNet121 and VGG16, transform the processed images into high-level numerical feature sets.
3. Cumulative feature integration: The extracted features from both models are flattened and combined into a unified feature vector for each image.
4. Class balancing: To mitigate bias and improve model performance, minority classes near the decision boundary are augmented.
5.Dimensionality Optimization: Principal Component Analysis (PCA) is applied to optimize the feature space by reducing redundancy while preserving discriminative information.
6. Classification: The system evaluates the refined features separately using Support Vector Machine (SVM) and k-Nearest Neighbors (KNN) algorithms to determine the final neuro-oncological lesion classification.

In line with common practice in recent literature [18, 19, 20, 21], the pipeline depicted in *figure 1* applies class balancing and dimensionality reduction to the overall data distribution. While this established approach is computationally efficient, the theoretical concern of potential data leakage is acknowledged. To thoroughly validate the findings, a comprehensive analysis was conducted. As detailed in *section 3.7*, performance was rigorously evaluated under both this paradigm and a strict alternative where these steps are applied only to training folds, confirming the substantial and genuine advantages of the proposed feature fusion approach.

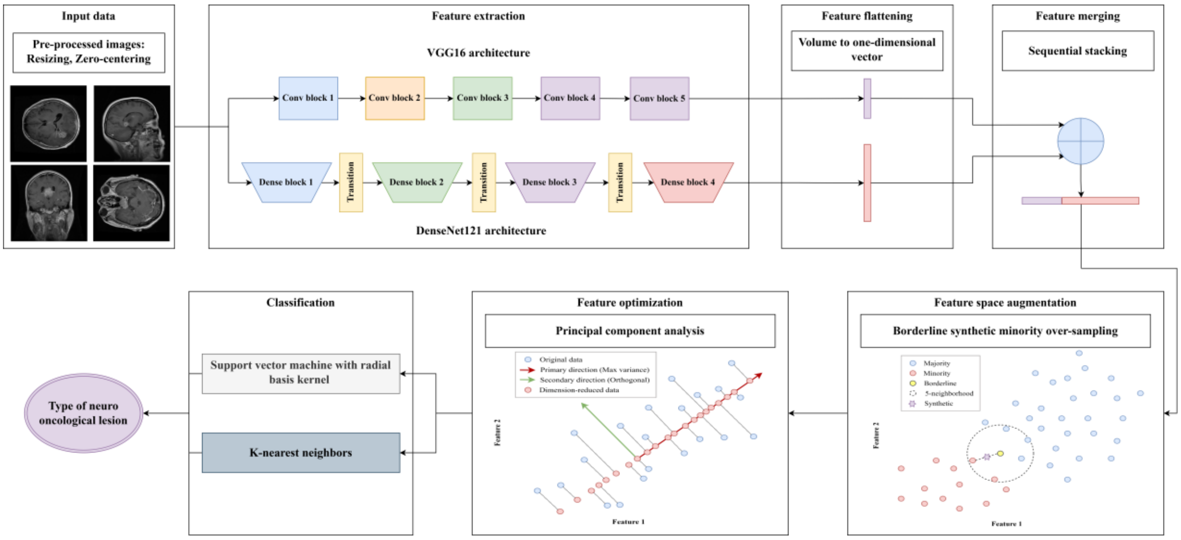


Figure 1. Proposed system architecture

2.1. Dataset and Preprocessing

The reliability of data-driven diagnostic systems depends critically on data quality. When models are trained on low-resolution images containing inaccurate labels or datasets of uncertain origin, they produce unreliable results with limited clinical utility. Consequently, careful selection of appropriate datasets becomes essential for developing valid and clinically relevant models.

This study employs the Figshare benchmark brain MRI dataset [22], consisting of 3064 images collected through collaborative efforts among medical experts and researchers across multiple Chinese hospitals. The dataset contains three clinically significant tumor categories with an imbalanced distribution: 1426 glioma cases, 706 meningioma cases, and 930 pituitary tumor cases. This composition reflects the prevalence of major brain tumor types in clinical practice, ensuring the developed models maintain practical relevance. Representative examples from the dataset are presented in *figure 2*.

All images underwent standard preprocessing procedures. First, they were uniformly resized to 224×224 pixels to optimize computational efficiency. The images were then normalized using mean subtraction to match ImageNet's distribution, zero-centering the pixel values for model compatibility.

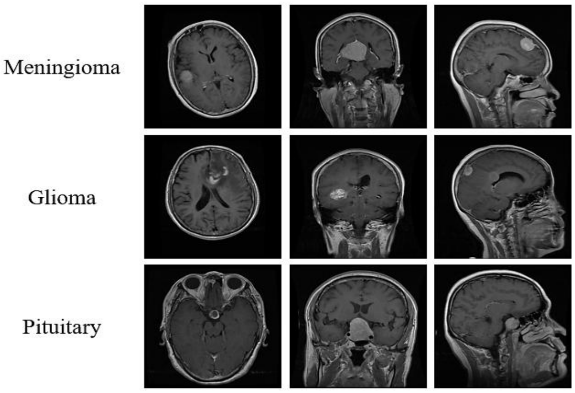


Figure 2. Samples from the dataset employed in the study

2.2. Feature Extraction

In this study, two deep learning models, VGG16 [23] and DenseNet121 [24], are employed to extract features from brain tumor images. Both models were initially trained on the ImageNet benchmark dataset for natural image classification. By freezing their optimized weights and removing the top classification layer, the networks can be repurposed for forward-pass feature extraction. The architectural details of each model are described below.

VGG16 is a convolutional neural network composed of 13 learnable layers arranged sequentially across five blocks, followed by a neural classifier. As illustrated in *table 1*, the first and second blocks each contain two convolutional layers, whereas the third, fourth, and fifth blocks consist of three layers each. All convolutional layers employ a uniform 3×3 filter size. Each block concludes with a pooling layer to reduce spatial dimensionality. Due to its simple and systematic design, this deep model generates discriminative numerical features that can be effectively utilized by subsequent classification algorithms.

DenseNet121, another convolutional neural network, distinguishes itself through skip connections that link each layer to every subsequent layer within a dense block. *Table 1* presents its architecture alongside VGG16 for comparison. The network comprises four dense blocks separated by transition layers. Each dense block contains multiple sets of convolutional layers and normalization components, enabling the extraction of high-level features efficiently. Transition layers incorporate a 1×1 convolution followed by pooling to maintain computational efficiency. This design promotes enhanced gradient flow during backpropagation, while feature reuse via skip connections serves as the primary advantage when employing the network as a feature extractor.

Models for image feature extraction vary widely in their size and complexity. For instance, the VGG16 network is relatively shallow and fast, yet on ImageNet, it generates higher-quality features than lightweight models like MobileNetV1. In contrast, DenseNet121 uses dense connections to reuse features more efficiently than ResNet architectures, and it is also less computationally demanding than its counterparts. As shown in *figure 3*, visualizations from different layers of VGG16 and DenseNet121 reveal their unique feature extraction capabilities. By using these two models in parallel and strategically merging their outputs, we can create a powerful set of features that may improve classifier performance for detecting brain anomalies.

Table 1. Architecture comparison of feature extraction layers in VGG16 and DenseNet121

Component	VGG16	DenseNet121
Input	224×224×3	224×224×3
Initial Layers	(None – starts directly with conv layers)	7×7 Conv (64), stride=2 → ReLU 3×3 MaxPool, stride=2
Block 1	2 × [3×3 Conv (64) → ReLU] 2×2 MaxPool, stride=2	Dense block: 6 × [1×1 Conv (128) → ReLU → 3×3 Conv (32)] Transition layer: 1×1 Conv (128) → 2×2 AvgPool
Block 2	2 × [3×3 Conv (128) → ReLU] 2×2 MaxPool, stride=2	Dense block: 12 × [1×1 Conv (128) → ReLU → 3×3 Conv (32)] Transition layer: 1×1 Conv (256) → 2×2 AvgPool
Block 3	3 × [3×3 Conv (256) → ReLU] 2×2 MaxPool, stride=2	Dense block: 24 × [1×1 Conv (128) → ReLU → 3×3 Conv (32)] Transition layer: 1×1 Conv (512) → 2×2 AvgPool
Block 4	3 × [3×3 Conv (512) → ReLU] 2×2 MaxPool, stride=2	Dense block: 16 × [1×1 Conv (128) → ReLU → 3×3 Conv (32)]
Block 5	3 × [3×3 Conv (512) → ReLU] 2×2 MaxPool, stride=2	—
Output Features	7×7×512	7×7×1024

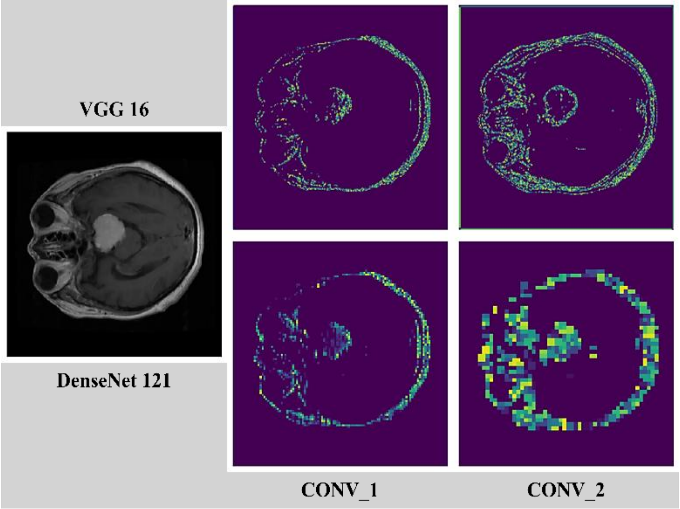


Figure 3. Reconstruction of MRI image features extracted from different layers of VGG16 and DenseNet121

2.3. Feature Fusion

Features obtained from different models often capture different information. Combining these features in a complementary manner can create a robust differential representation that mitigates the weaknesses and biases of individual models [25]. Several feature fusion methods are available, including element-wise operations such as addition and multiplication, or weighted fusion, where learned weights are assigned to features.

In this study, the simplest method for fusion, sequential stacking (also known as concatenation) [26], is used. Given feature vector $F_1 \in \mathbb{R}^n$ from the first model and $F_2 \in \mathbb{R}^m$ from the second model, the fused representation is:

$$Z = F_1 \oplus F_2 = [f_{11}, ..., f_{1n}, f_{21}, ..., f_{2m}] \in \mathbb{R}^{n+m}$$

where $\oplus$ denotes the concatenation operation. This method uniquely preserves all original features without requiring complex parameter optimization. The resulting higher-dimensional feature space retains maximum discriminative information, though subsequent processing may need to address the increased dimensionality.

2.4. Data Balancing in Feature Space

Despite the quality and reliability of the dataset used, it exhibits class imbalance, characterized by majority and minority classes. This skewed distribution may lead the model to disproportionately learn patterns from the majority class, potentially biasing its classification performance toward specific tumor type. While conventional input space augmentation through image transformations could address this imbalance, such approaches risk exacerbating existing domain mismatches when utilizing models trained on different tasks.

In contrast, this study implements feature space augmentation using the Borderline Synthetic Minority Oversampling Technique (Borderline-SMOTE) [27], an enhanced variant of the original SMOTE algorithm. Unlike standard SMOTE that generates synthetic samples through interpolation between arbitrary minority class pairs, Borderline-SMOTE specifically focuses on minority samples near classification boundaries, which represent the most challenging cases for accurate classification.

The technique operates through the following precise steps:

- Finding borderline samples. For each minority sample x:
    1. Find m nearest neighbors from all classes.
    2. Calculate minority_ratio = count(minority_neighbors)/m.
    3. Label x as:
        - Safe when minority_ratio > 0.5 (no action).
        - Borderline when 0 < minority_ratio ≤ 0.5 (keep for synthesis).
        - Noise when minority_ratio = 0 (discard).
- Generating synthetic samples. For each borderline x:
    1. Find k nearest neighbors from minority class only.
    2. Randomly select one neighbor s.
    3. Create a data point by interpolating between x and s according to the following formula:

$$x\_new = x + \text{random}(0, 1) \times (s - x)$$

The algorithm's two critical parameters, *m* (defining the neighborhood size for borderline detection) and *k* (determining neighbors for synthetic sample generation), were configured to optimize the synthesis of informative minority class samples. This approach effectively augments the minority classes (meningiomas and pituitary tumors) while preserving the original distribution of the majority class (gliomas), with *figure 4* demonstrating the balanced outcome.

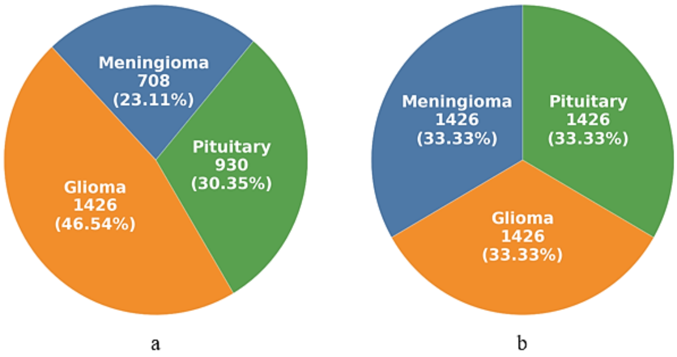


Figure 4. Data distribution: (a) Before augmentation, (b) After augmentation

2.5. Classification

The numerical representations extracted from MRI images enable classification of brain lesion types. However, deep models such as DenseNet121 and VGG16 produce an extensive set of features relative to the limited sample size. Direct feature fusion followed by classification may compromise generalization performance while unnecessarily increasing computational complexity and processing time. Principal component analysis (PCA) serves as an effective solution for dimensionality reduction while preserving essential information [28]. In this implementation, the number of principal components was selected to ensure adequate variance capture. Figure 5 demonstrates the progressive dimensionality changes across all processing stages. The resulting reduced features feed into two machine learning classifiers: a support vector machine employing an RBF kernel and a *k*-nearest neighbors classifier utilizing Euclidean distance metrics, with their respective hyperparameters tuned for optimal accuracy.

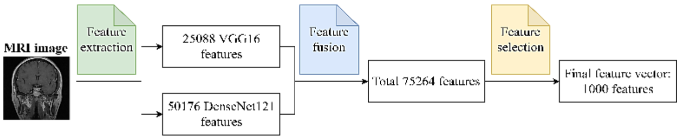


Figure 5. Dimensional shifts after each processing step

3. EXPERIMENTS, RESULTS AND DISCUSSION

All experiments were conducted in a Google Colab environment using Python, leveraging cloud GPU resources for computational efficiency. The models were implemented primarily with the Keras deep learning API, supported by the scikit-learn library for general machine learning tasks. To assess model performance, we used an 80% training and 20% validation split (hold-out) and applied a rigorous 5-fold cross-validation. The individual contributions of model components were also evaluated through a comprehensive ablation study. It should be noted that all results presented and discussed in this work correspond exclusively to the test dataset. This approach is critical because a model's true effectiveness is determined by its performance on data that was not used during training. The ability to generalize to unseen samples represents the most meaningful evaluation of a model's practical utility, as opposed to assessing performance on training data, which could yield unreliable performance measures.

3.1. Hyperparameter Configuration

The hyperparameters for all components of the pipeline were selected to balance model performance, generalization, and computational efficiency. A fixed random state of 42 was used across all stochastic processes to ensure the reproducibility of results. The specific choices are justified as follows:

Borderline-SMOTE: The parameters were set to $k = 1$ and $m = 5$. A small *k* value ensures that synthetic samples are generated very close to the borderline minority instances, preserving the local structure and avoiding the creation of noisy samples. The *m* value of 5 provides a sufficiently large neighborhood to reliably identify samples on the class decision boundary without being overly influenced by outliers.

Principal Component Analysis (PCA): The number of components was set to 1000. This value was selected because it guarantees the capture of at least 90% of the explained variance while maintaining a consistent feature dimensionality across all model configurations for a fair comparative analysis.

Support Vector Machine (SVM): An RBF kernel was chosen for its ability to model complex, non-linear decision boundaries. The regularization parameter *C* was set to 10. This value was found to provide a good trade-off between maximizing the margin and minimizing classification error, preventing the model from being either too rigid or too prone to overfitting.

K-Nearest Neighbors (KNN): The number of neighbors was set to $k = 1$, using the Minkowski metric with power parameter $p = 2$ (Euclidean distance). This choice is principled and aligns with the feature transformation: since the features were processed by PCA, which itself operates by maximizing variance in the Euclidean space, using the Euclidean distance for KNN ensures geometric consistency. The strong performance of this $k = 1$ configuration suggests that the feature space created by our fusion and reduction pipeline is highly discriminative, making the closest neighbor a reliable predictor.

3.2. VGG16 Feature-Based Model

In this experiment, deep features are extracted by processing input images through a pretrained VGG16 convolutional neural network. The high-dimensional feature vectors then undergo principal component analysis (PCA) for dimensionality reduction while retaining the most discriminative information. As shown in *figure 6*, selecting the top 1000 principal components preserves 90.13% of the original data variance. These reduced-dimension features serve as input to the classifiers. Model performance is evaluated through the classification report in *table 2* and the confusion matrices in *figure 7*. The support vector classifier achieves 95.76% overall accuracy with an average F1 score of 95.31%, misclassifying 26 test samples. Comparatively, the k-nearest neighbors classifier demonstrates slightly better performance with 95.92% accuracy, an average F1 score of 95.51%, and 25 misclassified instances. Both classifiers exhibit strong agreement in their predictive performance.

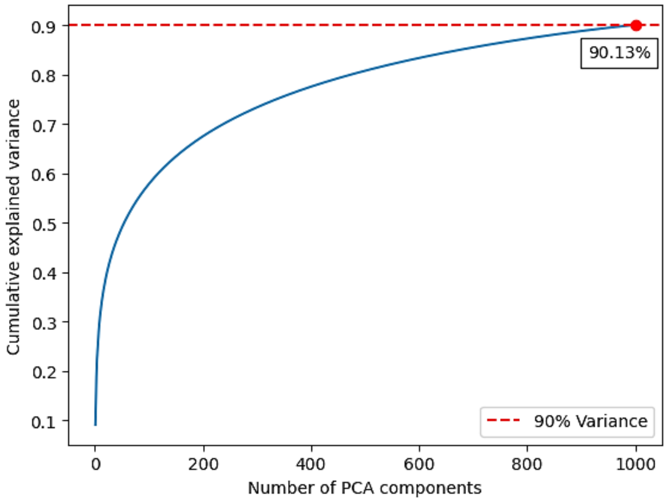
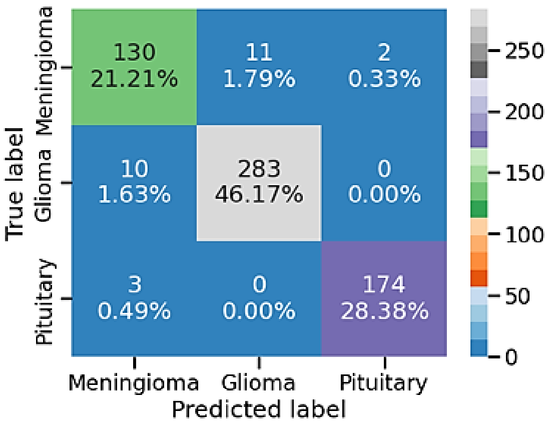


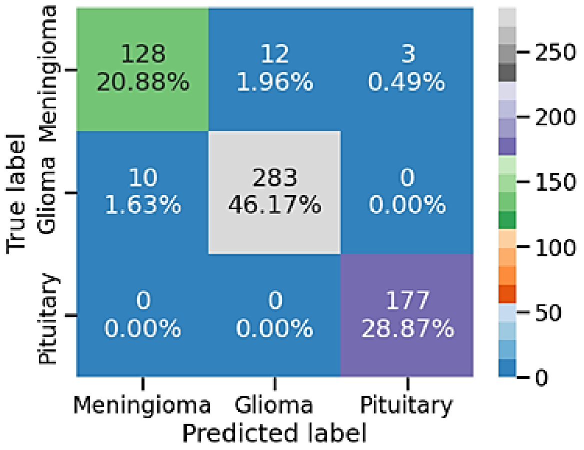
Figure 6. PCA cumulative explained variance of the VGG16 features

Table 2. Classification report of the VGG16 feature-based model.

End classifier	Category	Precision	Sensitivity	F1-score	Mean-F1	ACC
SVC	Meningioma	90.91	90.91	90.91	95.31	95.76
	Glioma	96.26	96.59	96.42		
	Pituitary	98.86	98.31	98.58		
KNN	Meningioma	92.75	89.51	91.10	95.51	95.92
	Glioma	95.93	96.59	96.26		
	Pituitary	98.33	100	99.16		



a



b

Figure 7. Confusion matrices of the VGG16 feature-based model: (a) SVM classifier, (b) KNN classifier

3.3. DenseNet121 Feature-Based Model

The system pipeline undergoes modification through replacement of the VGG16 feature extractor with a pretrained DenseNet121 architecture, while preserving all other system components. Variance analysis in *figure 8* indicates that a minimal set of principal components explains over 90% of total variance, meeting standard analytical thresholds. However, experimental consistency is maintained through use of 1000 features, capturing 99.53% of variance. Classification metrics in *table 3* demonstrate model performance, with the SVM classifier achieving 96.25% accuracy and an average F1 score of 95.81%, while the KNN classifier achieves superior results of 97.06% accuracy and 96.84% average F1 score. These outcomes represent accuracy improvements of 0.49% and 1.14% respectively compared to the VGG16 implementation. Corresponding confusion matrices in *figure 9* show reduced error rates, with 23 misclassified samples for SVM and 18 for KNN.

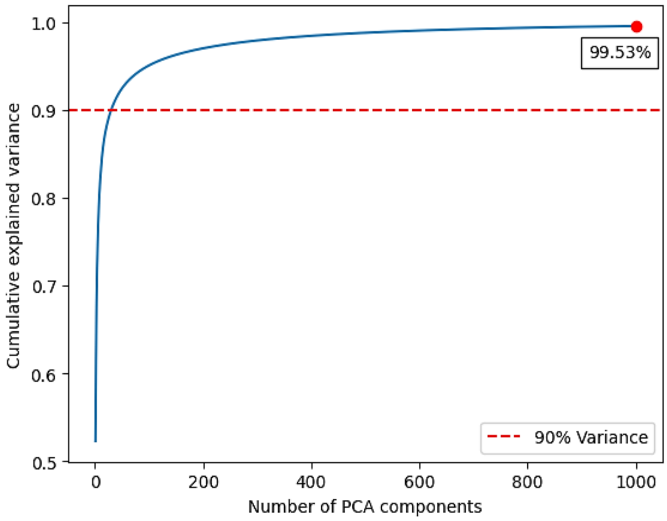


Figure 8. PCA cumulative explained variance of the DenseNet121 features

Table 3. Classification report of the DenseNet121 feature-based model.

End classifier	Category	Precision	Sensitivity	F1-score	Mean-F1	ACC
SVC	Meningioma	91.10	93.01	92.04	95.81	96.25
	Glioma	97.59	96.59	97.08		
	Pituitary	98.31	98.31	98.31		
KNN	Meningioma	93.71	93.71	93.71	96.84	97.06
	Glioma	96.94	97.27	97.10		
	Pituitary	100	99.44	99.72		

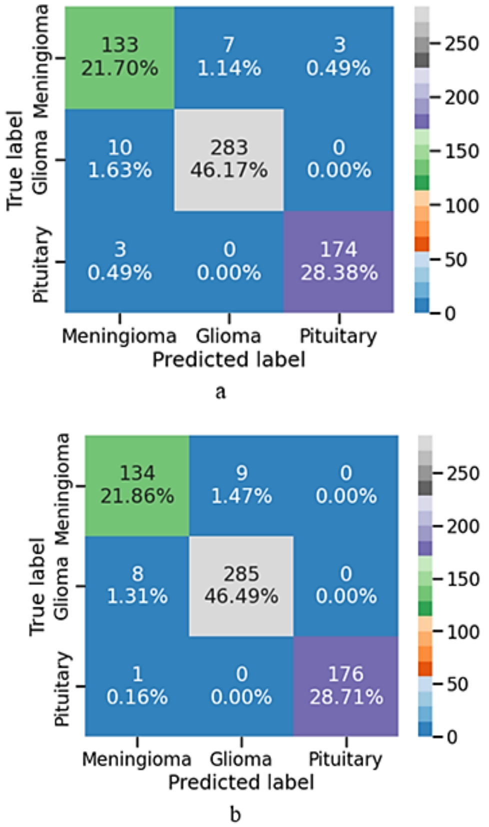


Figure 9. Confusion matrices of the DenseNet121 feature-based model: (a) SVM classifier, (b) KNN classifier

3.4. Deep Feature Fusion Model

Fusing features extracted from VGG16 and DenseNet121 networks provides more diverse information, improving representational power. However, this process can produce overlapping features and increase dimensionality. *Figure 10* shows that using PCA to reduce the dimensions of the deep fused features to a fixed number of 1000 components captures less variance (94.18%) than DenseNet121 alone but still exceeds the variance obtained with VGG16. The classification report in *table 4* highlights the superiority of the fused feature model over both individual networks. The SVM classifier achieved an accuracy of 97.39% and an average F1-score of 97.06%, while the KNN classifier reached an accuracy of 98.21% and an average F1-score of 98.10%. The confusion matrices in *figure 11* indicate a low number of misclassified samples, with no more than 16 for SVM and only 11 for KNN.

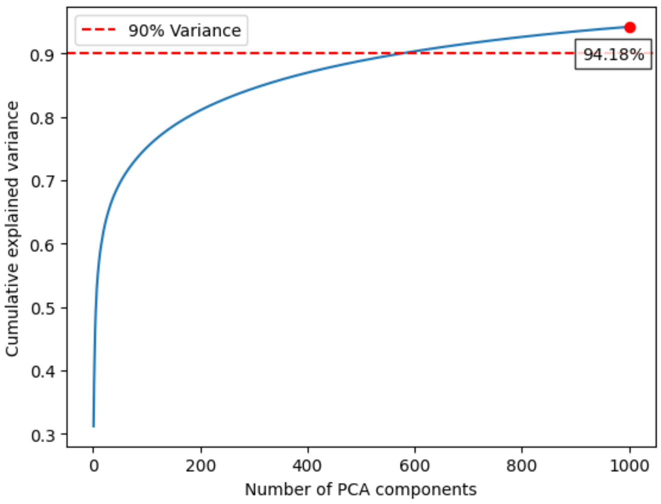


Figure 10. PCA cumulative explained variance of the fused features

Table 4. Classification report of the deep feature fusion model

End classifier	Category	Precision	Sensitivity	F1-score	Mean-F1	ACC
SVC	Meningioma	93.20	95.80	94.48	97.06	97.39
	Glioma	98.62	97.61	98.11		
	Pituitary	98.86	98.31	98.58		
KNN	Meningioma	95.83	96.50	96.17	98.10	98.21
	Glioma	98.29	97.95	98.12		
	Pituitary	100	100	100		

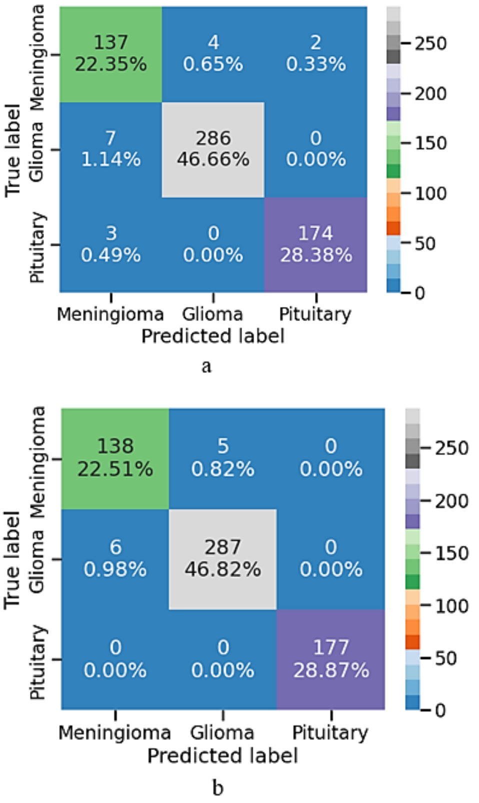


Figure 11. Confusion matrices of the deep feature fusion model: (a) SVM classifier, (b) KNN classifier

3.5. Augmented Feature Fusion Model

This experiment integrates a Borderline SMOTE component into the deep feature fusion model to balance the dataset and increase its size. Following the same approach as previous cases, the high-dimensional feature space is processed using PCA. *Figure 12* displays the explained variance relative to the number of principal components, with the first 1000 components capturing over 95% of the variance. The classification results in *table 5* demonstrate outstanding performance, with SVM achieving near-perfect accuracy of 99.53% and average F1 score of 99.51%, while KNN reached 98.71% accuracy and 98.67% average F1 score. The confusion matrices in *figure 13* reveal minimal misclassifications: only 4 samples for SVM and 11 for KNN. The proposed system shows clear improvements over individually employed pretrained networks, with SVM accuracy increasing by more than 3% compared to both VGG16 and DenseNet121 based models, while KNN shows improvements of 2.79% over VGG16 and 1.65% over DenseNet121.

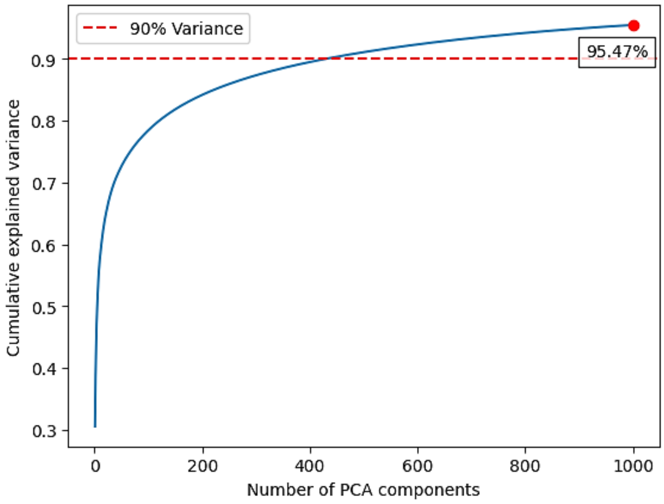


Figure 12. PCA cumulative explained variance of the augmented fused features

Table 5. Classification report of the augmented feature fusion model

End classifier	Category	Precision	Sensitivity	F1-score	Mean-F1	ACC
SVC	Meningioma	98.39	100	99.19	99.51	99.53
	Glioma	100	99.05	99.52		
	Pituitary	100	99.66	99.83		
KNN	Meningioma	97.17	98.36	97.76	98.67	98.71
	Glioma	98.73	97.79	98.26		
	Pituitary	100	100	100		

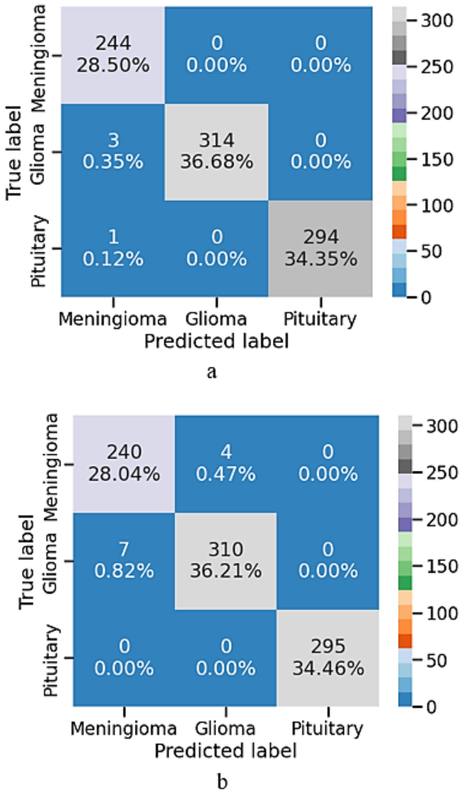


Figure 13. Confusion matrices of the augmented feature fusion model: (a) SVM classifier, (b) KNN classifier.

3.6. Statistical Significance of Performance Improvements

To assess whether significant differences exist between model configurations, researchers typically employ the paired t-test. This statistical method compares means from two related groups to verify that observed differences reflect true effects rather than random variation. Since the hold-out method yields only a single test value, it cannot support paired t-test analysis. Consequently, a cross-validation approach becomes essential to generate multiple evaluation metrics for each model.

Table 6 presents performance statistics, while *figure 14* compares the accuracy values for both the 5-fold cross validation and hold-out approaches. The close alignment of accuracy values between these evaluation approaches for each model suggests minimal bias toward specific data subsets and indicates strong generalizability. We utilize the cross-validation fold accuracy values to perform paired t-tests comparing instances of the developed model. *Table 7* displays the resulting p-values and Cohen's d effect sizes for each model pair comparison. The *p*-values estimate the probability that observed performance differences occurred by random chance (with $p < 0.05$ indicating statistical significance), while Cohen's *d* quantifies the magnitude of these differences, where values ≥0.2, ≥0.5, and ≥0.8 typically represent small, medium, and large effects respectively.

The analysis reveals that both classifiers achieve comparable performance with either VGG16 or DenseNet121 as the feature extractor. However, fusing features from both networks produces statistically significant performance gains over either individual model. The augmented feature fusion model shows additional gains, though these require further methodological consideration as discussed in *section 3. 7.*

Table 6. Statistical performance evaluation using 5-fold cross-validation

Model configuration	Classifier	Mean accuracy	Standard deviation	95% Confidence interval
VGG16	SVC	96.38	0.79	(95.39-97.36)
	KNN	96.15	0.63	(95.37-96.93)
DenseNet121	SVC	96.70	0.71	(95.82-97.58)
	KNN	96.31	0.66	(95.49-97.13)
Feature fusion	SVC	97.49	0.62	(96.72-98.25)
	KNN	97.68	0.57	(96.97-98.39)
Augmented feature fusion	SVC	98.83	0.65	(98.03-99.63)
	KNN	98.60	0.39	(98.12-99.08)

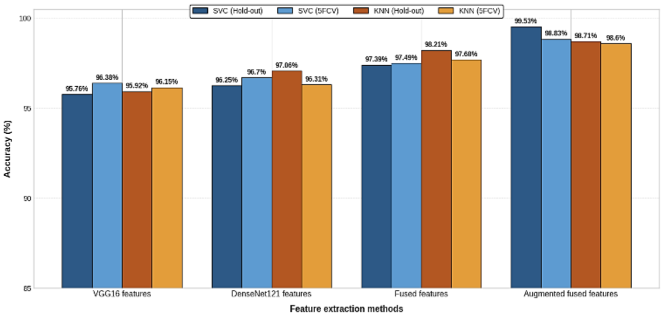


Figure 14. Accuracy performance across models: 5-Fold cross-validation and hold-out results

Table 7. Pairwise statistical significance across tested models

Comparison	SCV		KNN	
	p-value	Cohen's *d*	*p*-value	Cohen's *d*
VGG16 vs. DenseNet121	0.3253	0.50	0.5778	0.27
VGG16 vs. Fused features	**0.0022**	**3.14**	**0.0072**	**2.26**
DenseNet121 vs. Fused features	**0.0218**	**1.63**	**0.0014**	**3.52**
Fused features vs. Augmented fused features	**0.0121**	**1.95**	**0.0033**	**2.81**

3.7. Robustness to Data Processing Sequence

Principal Component Analysis (PCA) and Borderline-Synthetic Minority Oversampling Technique (Borderline-SMOTE) are commonly applied as preprocessing techniques prior to dataset partitioning in machine learning workflows [18, 19, 20, 21]. Although efficient, this approach introduces a potential risk of data leakage. To ensure the validity of the findings, a systematic investigation into the impact of processing sequence on model performance was conducted using 5-fold cross-validation.

The analysis evaluated the proposed feature fusion model against baseline individual models (VGG16 and DenseNet121) across key processing sequences. The results, detailed in *table 8*, yield two critical observations. Firstly, the performance of the feature fusion model exhibits remarkable stability concerning the timing of PCA application. The minimal performance degradation observed when PCA is applied post-split confirms that the superior discriminative power of the fused features is an inherent property and not an artifact of data leakage.

Secondly, the analysis of Borderline-SMOTE provides two key insights. When applied in a post-split manner, it offers no performance benefit for the SVC and only a minor decrement for KNN compared to the baseline feature fusion. This indicates that the feature representation learned by the multi-model fusion is inherently robust to class imbalance, effectively mitigating the need for synthetic oversampling. However, the notable performance difference between the pre-split and post-split application of Borderline-SMOTE suggests that the high results reported by the former can be attributed to data leakage rather than a genuine improvement in the model's generalization capability.

Table 8. Performance comparison across model configurations (5-fold cross-validation mean accuracy %).

Model configuration	PCA before split	PCA after split	Borderline-SMOTE before split	Borderline-SMOTE after split
VGG16 + SVC	96.38	95.73	—	—
VGG16 + KNN	96.15	94.91	—	—
DenseNet121 + SVC	96.70	96.31	—	—
DenseNet121 + KNN	96.31	95.99	—	—
Feature fusion + SVC	**97.49**	**97.03**	98.83	97.03
Feature fusion + KNN	**97.68**	**97.00**	98.60	96.77

This comprehensive evaluation confirms that the performance advantages of the multi-model feature fusion approach remain substantial and are not dependent on specific data processing implementation details.

4. LIMITATIONS AND FUTURE WORK

This work is subject to several important limitations that help define a clear pathway for future research.

4.1. Dataset and Benchmarking Limitations

The study utilizes the Figshare brain tumor dataset, which is characterized by its high-quality, professional curation under institutional oversight, ensuring label integrity and ethical compliance. This standard of quality differentiates it from other publicly available brain tumor datasets that are often scraped from the internet without rigorous validation and are consequently less reliable for robust benchmarking [29].

However, a key limitation is that the data was collected exclusively from Chinese institutions, meaning the trained models may exhibit institutional and demographic biases, potentially limiting their immediate global applicability. Furthermore, a critical consideration is the dataset's splitting methodology. The dataset does not contain duplicated images from the same patient, which reduces the risk of data leakage and makes the established image-level splitting protocol a sound and standard choice. To our knowledge, no prior work on this dataset has implemented a patient-level split; therefore, deviating from this universal precedent would invalidate all direct comparisons with the state-of-the-art, which is a fundamental aspect of our performance evaluation. Future work should pursue patient-level splits on larger, multi-institutional datasets.

4.2. Clinical Translation and Explainability

The clinical applicability of the proposed system requires significant further development. Our current work serves as a proof-of-concept, demonstrating the performance gains from multi-model feature fusion. For real-world clinical adoption, essential next steps include rigorous validation in live diagnostic settings and on data from diverse medical centers.

While this study provided visualizations of discriminative features from the individual VGG16 and DenseNet121 models, generating a unified, human-interpretable explanation for the final prediction based on the fused feature space remains a complex challenge to be addressed in future work. Finally, while the model successfully classifies the three most prevalent tumor types (comprising approximately 75% of clinical cases), expanding its diagnostic scope to include a wider spectrum of rarer neuro-oncological pathologies is crucial for enhancing its overall clinical utility and impact.

7. CONCLUSION

This work presents an effective deep learning approach for detecting brain lesion types in MRI images, specifically targeting meningioma, glioma, and pituitary tumors. The proposed system employs a dual-network architecture combining VGG16 and DenseNet121 for parallel feature extraction from processed MRI scans. These complementary networks produce deep features that are fused through sequential stacking, yielding a robust numerical representation of the input images.

To enhance the system's efficiency, principal component analysis optimizes the feature space by preserving maximum variance during dimensionality reduction. A comprehensive investigation into class balancing revealed that the proposed feature fusion inherently mitigates class imbalance, demonstrating robust performance without relying on synthetic augmentation. For final classification, the system leverages both support vector machines and k-nearest neighbors, confirming the feature representation's quality across different classifier types. The model's validity is examined through hold-out testing and 5-fold cross-validation, supported by ablation studies and paired t-test analysis that verify each component's contribution. A methodological robustness analysis further confirms that the performance advantages are genuine and not an artifact of data processing sequences. These results demonstrate the system's superior performance compared to existing methods and suggest the proposed approach could serve as a valuable decision-support tool for clinical specialists, potentially improving both the speed and accuracy of diagnostic processes for brain lesions.

Author Contributions: Yasser Nizamli: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review and editing. Anton Filatov: Validation, Supervision. Weaam Fadel: Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review and editing. Yulia Shichkina: Validation, Supervision. Kinda Mreish: Data curation, Resources, Validation, Visualization. Nahida Karaja: Data curation, Resources, Validation, Visualization, Editing. Tarek Alnajar: Resources, Validation, Visualization, Editing.

Data Availability Statement: The brain tumor dataset analyzed during this study is publicly available from the Figshare repository at https://figshare.com/articles/dataset/brain_tumor_dataset/1512427/5.

The preprocessed version (JPG format with normalization) used to achieve the reported results is available on Zenodo: https://doi.org/10.5281/zenodo.17530045.

Code Availability: The source code and implementations necessary to replicate the findings of this study are available in the GitHub repository at https://github.com/Yasser-A-Nizamli/neuro-oncology-feature-fusion.

Ethical Considerations: This research utilized a de-identified, publicly available benchmark dataset. No new patient data was collected, and thus the study did not require institutional review board (IRB) approval. All experiments were conducted in accordance with relevant guidelines and regulations.

Funding Source: This research has been supported by the Ministry of Science and Higher Education of the Russian Federation (assignment No. FSEE-2025-0015).

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

Acknowledgments: The authors sincerely thank the anonymous reviewers for their insightful comments and constructive suggestions, which have significantly strengthened the methodological rigor and clarity of this manuscript.

REFERENCES

[1] Huang, T.; Yin, X.; Jiang, E. Decision-making in clinical diagnostic for brain tumor detection based on advanced machine learning algorithm. International Journal for Simulation and Multidisciplinary Design Optimization (IJSMDO) 2025, Volume 16, Issue 1. DOI: 10.1051/smdo/2024021.

[2] Perkins, A.; Liu, G. Primary brain tumors in adults: diagnosis and treatment. Am Fam Physician 2016, Volume 93, Issue 3, pp. 211–217. PMID: 26926614.

[3] Barkade, G.; Bhosale, P.; Shirsath, S. Overview of brain cancer, its symptoms, diagnosis and treatment. IP International Journal of Comprehensive and Advanced Pharmacology 2023, Volume 8, Issue 3, pp. 159–164. DOI: 10.18231/j.ijcaap.2023.027.

[4] Alther, B.; Mylius, V.; Weller, M.; Gantenbein, A. From first symptoms to diagnosis: Initial clinical presentation of primary brain tumors. Clinical and Translational Neuroscience 2020, Volume 4, Issue 2. DOI: 10.1177/2514183X20968368.

[5] Kang, J.; Ullah, Z.; Gwak, J. MRI-Based brain tumor classification using ensemble of deep features and machine learning classifiers. Sensors 2021, Volume 21, Issue 6. DOI: 10.3390/s21062222.

[6] Abda, O.; Naimi, H. Enhanced brain tumor MRI classification using stationary wavelet transform, ResNet50V2, and LSTM networks. ITEGAM-JETIA 2025, Volume 11, Issue 51, pp. 127–133. DOI: 10.5935/jetia.v11i51.1457.

[7] Nizamli, Y.; Filatov, A.; Fadel, W.; Shichkina, Yu. Accurate anomaly detection in medical images using transfer learning and data optimization: MRI and CT as case studies. 2024 V International Conference on Neural Networks and Neurotechnologies (NeuroNT), Saint Petersburg, Russia, 2024, pp. 170–173. DOI: 10.1109/NeuroNT62606.2024.10585603.

[8] Alahmed, H.; Al-Suhail, G. Exploring transfer learning techniques for brain tumor diagnosis in MRI data. 2024 1st International Conference on Emerging Technologies for Dependable Internet of Things (ICETI), Sana'a, Yemen, 2024, pp. 1–8. DOI: 10.1109/ICETI63946.2024.10777184.

[9] Bouguerra, O.; Attallah, B.; Brik, Y. MRI-based brain tumor ensemble classification using two stage score level fusion and CNN models. Egyptian Informatics Journal 2024, Volume 28. DOI: 10.1016/j.eij.2024.100565.

[10] Biswas, A.; Islam, M. S. A hybrid deep CNN-SVM approach for brain tumor classification. Journal of Information Systems Engineering and Business Intelligence 2023, Volume 9, pp. 1–15. DOI: 10.20473/jisebi.9.1.1-15.

[11] Jaspin, K.; Selvan, S. Multiclass convolutional neural network-based classification for the diagnosis of brain MRI images. Biomedical Signal Processing and Control 2023, Volume 83. DOI: 10.1016/j.bspc.2022.104542.

[12] Sowrirajan, S. R.; Balasubramanian, S. Brain tumor classification using machine learning and deep learning algorithms. International Journal of Electrical and Electronics Research 2022, Volume 10, pp. 999–1004. DOI: 10.37391/IJEER.100441.

[13] Agrawal, T.; Choudhary, P.; Shankar, A.; Singh, P.; Diwakar, M. MultiFeNet: Multi-scale feature scaling in deep neural network for the brain tumour classification in MRI images. Int J Imaging Syst Technol 2024. Volume 34, Issue 1. DOI: 10.1002/ima.22956.

[14] Noreen, N.; Palaniappan, S.; Qayyum, A.; Ahmad, I.; Alassafi, M. O. Brain Tumor Classification Based on Fine-Tuned Models and the Ensemble Method. Computers, Materials & Continua 2021, Volume 67, Issue 3, pp. 3967–3982. DOI: 10.32604/cmc.2021.014158.

[15] Khan, S. U. R.; Zhao, M.; Asif, S.; Chen, X. Hybrid-NET: A fusion of DenseNet169 and advanced machine learning classifiers for enhanced brain tumor diagnosis. International Journal of Imaging Systems and Technology 2024, Volume 34, Issue 1. DOI: 10.1002/ima.22975.

[16] Passa, R. S.; Nurmaini, S.; Rini, D. P. YOLOv8 based on data augmentation for MRI brain tumor detection. Scientific Journal of Informatics 2023. Volume 10, Issue 3. DOI: 10.15294/sji.v10i3.45361.

[17] Muksimova, S.; Umirzakova, S.; Mardieva, S.; Iskhakova, N.; Sultanov, M.; Cho, Y. I. A lightweight attention-driven YOLOv5m model for improved brain tumor detection. Computers in Biology and Medicine 2025, Volume 188. DOI: 10.1016/j.compbiomed.2025.109893.

[18] Salakapuri, R.; Terlapu, P. V.; Kalidindi, K. R.; Balaka, R. N.; Jayaram, D.; Ravikumar , T. Intelligent brain tumor detection using hybrid finetuned deep transfer features and ensemble machine learning algorithms. Scientific Reports 2025, Volume 15. DOI: 10.1038/s41598-025-08689-6.

[19] Nagaraju, G.; Gujjeti, S.; Varaprasad Rao, M.; Patil, A.; Yamsani, N. Deep learning-driven behavioral analysis for real-time threat detection and classification in network traffic. International Journal of Electrical and Electronics Research 2025, Volume 13, Issue 1, pp. 80–88. DOI: 10.37391/IJEER.130112.

[20] Malik, H.; Anees, T.; Khalil, W.; Alharthi, S. Z.; Al-Shamaylehs, A. S.; khunzada, A. Deep learning-based classification of chest diseases using X-rays, CT scans, and cough sound images 2023, Volume 13, Issue 17. DOI: 10.3390/diagnostics13172772.

[21] Merlin, R. T.; Ravi, R. Empowering smart city IoT network intrusion detection with advanced ensemble learning-based feature selection 2024, Volume 12, Issue 2, pp. 367–374. 10.37391/IJEER.120206.

[22] Figshare brain tumor dataset. Available online: https://figshare.com/articles/dataset/brain_tumor_dataset/1512427/5 (accessed 08.01.2025).

[23] Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. arXiv 2015. DOI: 10.48550/arXiv.1409.1556.

[24] Huang, G.; Liu, Z.; van der Maaten, L.; Weinberger, K. Q. Densely Connected Convolutional Networks. . arXiv 2017. DOI: 10.48550/arXiv.1608.06993.

[25] Belal, M.; Hassan, T.; Hassan, A.; Alsheikh, N.; Elhendawi, N.; Hussain, I. Integrating features for recognizing human activities through optimized parameters in graph convolutional networks and transformer architectures. 2024 International Conference on Digital Image Computing: Techniques and Applications (DICTA), Perth, Australia, 2024, pp. 449-453. DOI: 10.1109/DICTA63115.2024.00072.

[26] Zhang, T.; Fan, S.; Hu, J.; Guo, X.; Li, Q.; Zhang, Y.; Wulamu, A. A feature fusion method with guided training for classification tasks, Computational Intelligence and Neuroscience 2021, Volume 2021, Issue 1. DOI: 10.1155/2021/6647220.

[27] Han, H.; Wang, WY.; Mao, BH. Borderline-SMOTE: A new over-sampling method in imbalanced data sets learning. 2005 International Conference on Intelligent Computing (ICIC), Hefei, China, 2005, pp. 878-887. DOI: 10.1007/11538059_91.

[28] Alyabrodi, A.; Al-Daja, S.; Injadat, M.; Moubayed, A.; Elrashidi, A. Evaluation of chest CT-scan anomaly detection models using principal component analysis. 2024 25th International Arab Conference on Information Technology (ACIT), Zarqa, Jordan, 2024, pp. 1-6. DOI: 10.1109/ACIT62805.2024.10877172.

[29] Nizamli, Y.; Filatov, A. On the validity of public MRI datasets for neuro-oncological abnormality detection and classification: a forensic audit. TechRxiv 2025. DOI: 10.36227/techrxiv.176287946.66170971/v1.