Journal of Applied Informatics and Computing (JAIC)
Vol.10, No.2, April 2026, pp. 1359~1368
e-ISSN: 2548-6861

Classification of Brain Tumors Using a Hybrid VGG16-ResNet50 Architecture with SVM

Mohammad Aviscena Zaidan [1*], L. Budi Handoko [2*], Abdussalam[3*],
[*] Teknik Informatika, Universitas Dian Nuswantoro, l. Imam Bonjol No. 207, Semarang, Jawa Tengah, Indonesia
111202214812@mhs.dinus.ac.id [1] , handoko@dsn.dinus.ac.id [2], abdussalam@dsn.dinus.ac.id [3]

Article Info

Article history:

Received 2026-01-19
Revised 2026-02-26
Accepted 2026-04-08

Keyword:

Brain Tumor Classification,
MRI,
Convolutional Neural Networks,
VGG16,
ResNet50,
Support Vector Machine.

ABSTRACT

Brain tumor classification from Magnetic Resonance Imaging (MRI) is critical for clinical diagnosis and treatment planning. This study proposes a hybrid deep learning architecture combining VGG16 and ResNet50 convolutional neural networks, utilized strictly as fixed feature extractors via transfer learning, followed by Principal Component Analysis (PCA) for dimensionality reduction and Support Vector Machine (SVM) classification. Unlike approaches that train CNNs from scratch, we leverage pre-trained ImageNet weights to extract high-level feature representations without fine-tuning the convolutional layers. The methodology was evaluated on a dataset of 3,064 T1-weighted contrast-enhanced MRI images categorized into three tumor classes: Meningioma, Glioma, and Pituitary tumors. Three experimental configurations with distinct train-test ratios (80:20, 70:30, and 60:40) were investigated using strict patient-level splitting to prevent data leakage, achieving 90.36% accuracy (80:20 split) with 0.9564 specificity. To rigorously validate clinical generalization, a 5-fold cross-validation was subsequently performed, yielding 85.90% accuracy. The consistent high specificity (>0.91) across all configurations indicates reliable clinical applicability. The fused feature representation, combined with PCA compression to approximately 290 components and SVM margin maximization, produced a computationally efficient and interpretable model suitable for deployment in resource-constrained clinical environments. Results demonstrate that architectural fusion of complementary deep learning features with classical machine learning classifiers achieves competitive performance with enhanced generalization and data efficiency compared to single-model approaches. The novelty of this study lies in the empirical validation of a "frozen-fusion" strategy that eliminates the need for expensive backpropagation during training while outperforming standard transfer learning baselines, specifically determining the optimal PCA-SVM configuration for maximizing specificity in pituitary tumor diagnosis within resource-constrained environments.

I. INTRODUCTION

The human brain, the command center of the central nervous system, comprises approximately 86 billion neurons responsible for coordinating systemic physiological functions. Aberrant cellular regulation within this complex network can precipitate the formation of tumors-masses of abnormal tissue that range from benign to malignant. As noted by the study in [1], malignant phenotypes such as gliomas are aggressively invasive and necessitate immediate therapeutic intervention. Conversely, benign tumors like meningiomas, while slower-growing, still present significant morbidity risks due to intracranial pressure. The World Health Organization (WHO) stratifies these neoplasms into distinct grades of malignancy, rendering precise histological classification a formidable challenge in neuro-oncology [2],[1].

Magnetic Resonance Imaging (MRI) serves as the gold standard for non-invasive brain tumor diagnosis, offering superior soft-tissue contrast compared to Computed Tomography (CT). However, the qualitative interpretation of

MRI data remains heavily dependent on radiologist expertise. With the exponential growth in global medical imaging data, manual analysis has become a critical bottleneck, exacerbating the risk of diagnostic errors driven by fatigue and inter-observer variability. While traditional Machine Learning (ML) paradigms, such as Logistic Regression and Random Forests, have been investigated [3], they frequently struggle to resolve the high-dimensional complexity inherent in raw pixel data. Consequently, there is an urgent clinical demand for Computer-Aided Diagnosis (CAD) systems capable of delivering objective, reproducible, and rapid classifications.

Deep Learning (DL), particularly through Convolutional Neural Networks (CNNs), has emerged as a transformative methodology in medical image analysis. Rasa and colleagues [4] demonstrated that transfer learning—wherein models pre-trained on large-scale datasets like ImageNet are fine-tuned for specific tasks—significantly outperforms networks trained de novo on limited medical repositories. Foundational investigations by Husen [5] further validated the efficacy of standard CNN architectures augmented with data processing techniques for MRI classification. Despite these advances, the field currently faces a trade-off between model computational efficiency and feature representational depth. Several studies have focused on optimizing single architectures. In comparative analyses of medical imaging, Riaqi and Tahyudin [6] advocated for VGG16's stability in smaller datasets over deeper architectures. Similarly, Husna and co-authors [7] and Liyananta and colleagues [8] explored MobileNetV2 and ResNet50, respectively, observing that while individual models achieve competent performance, they often fail to capture complementary feature sets. To address generalization issues, Vicky and Purba [9] proposed optimizing VGG16 via a "Freeze-5" strategy.

To overcome the inherent limitations of single-model architectures, recent research trends have pivoted toward ensemble and hybrid frameworks. Yoon [10] proposed the concatenation of Xception and Parallel Deep CNNs to leverage diverse kernel receptive fields. Similarly, Boubdellah and colleagues [11] and Suryawanshi & Patil [12] demonstrated that fusing feature vectors from distinct architectures (e.g., VGG combined with Inception) yields accuracy metrics superior to any standalone model. More complex hybridizations have also been explored; Younis and colleagues [13] integrated VGG16 with Vision Transformers (ViT) to capture global dependencies, while Gokcek-Sarac and co-authors incorporated Reinforcement Learning (Q-Learning) into the classification pipeline. Furthermore, meta-heuristic algorithms, such as the Sine-Cosine Grey Wolf Optimization used by Zain Eldin and colleagues [14] have been applied to fine-tune CNN hyperparameters.

However, as model complexity increases, many state-of-the-art architectures risk becoming computational "black boxes." Kumaran and colleagues [8] emphasized the necessity of explainability and ensemble learning in analogous domains like lung cancer-a principle highly applicable to neuro-oncology. The specific synergistic combination of VGG16 (renowned for robust textural feature extraction) and ResNet50 (capable of deep semantic abstraction via residual connections) offers a potent balance between performance and interpretability. Moreover, Nabila and Salam [15] demonstrated that replacing the conventional dense softmax layer of a CNN with a Support Vector Machine (SVM) classifier can significantly enhance decision margins and classification robustness.

Despite these advancements, a critical gap remains in the literature regarding the trade-off between model complexity and classification reliability in resource-limited settings. While recent state-of-the-art approaches often employ computationally intensive architectures like Vision Transformers (ViT) or extensive ensembles [13], they frequently lack interpretability and require significant training resources. Conversely, lightweight models like MobileNetV2 [7] may compromise on feature richness. Our study specifically addresses this by leveraging transfer learning with fixed weights-avoiding the computational cost of training deep networks from scratch-while enhancing feature discrimination through SVM margins, offering a balanced solution that remains unexplored in current comparative studies for this specific dataset.

Building upon this diverse body of literature, this study proposes a unified Hybrid Feature Fusion Framework. We hypothesize that concatenating the deep semantic representations of ResNet50 with the robust textural features of VGG16, followed by Principal Component Analysis (PCA) for dimensionality reduction and SVM for classification, will yield a highly discriminative and clinically efficient tool. This approach aims to maximize diagnostic accuracy-competing with complex transformer-based models-while retaining the computational efficiency and decision-boundary reliability associated with classical SVM margins

II. METHOD

The proposed methodology employs a structured hierarchical pipeline designed to maximize representational diversity while mitigating computational redundancy. The workflow encompasses four primary stages: Data Preprocessing, Dual-Stream Deep Feature Extraction, Dimensionality Reduction, and Support Vector Classification. This approach leverages the complementary strengths of two distinct convolutional architectures to create a robust and efficient clinical diagnostic tool.

This study utilizes the publicly available brain tumor MRI dataset published by Cheng [16], hosted on Figshare. The dataset, originally described in "Enhanced Performance of Brain Tumor Classification via Tumor Region Augmentation and Partition"[17] , comprises 3,064 T1-weighted contrast-enhanced MRI images from 233 patients. The images are categorized into three distinct tumor classes: Meningioma (708 images), Glioma (1426 images), and Pituitary tumors (930 images).

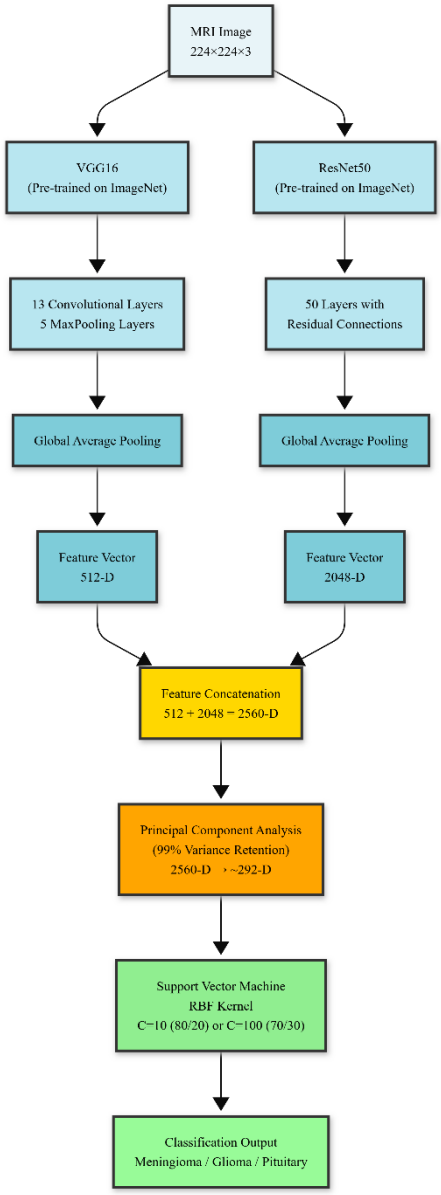

Figure 1. Proposed Block Diagram of the Hybrid VGG16-ResNet50-SVM Architecture.

TABLE I
CLASS DISTRIBUTION OF THE DATASET

Tumor Type	Quantity Images)	Description
Meningioma	708	Benign, slow-growing tumor
Giloma	1426	Malignant, aggressive tumor
Pituitary	930	Tumor in the pituitary gland
Total	3064	

The data is provided in MATLAB (.mat) format, containing pixel data, patient IDs (PID), and tumor masks manually delineated by radiologists. To evaluate the impact of data leakage and ensure clinical robustness, two partitioning strategies were employed. First, Standard Stratified Splitting (80:20, 70:30, 60:40) was used to establish baseline performance metrics comparable to existing literature. Second, to prevent data leakage-where slices from the same patient appear in both training and testing sets-we utilized Patient-Level Stratified Group Cross-Validation. This rigorous protocol ensures that the model's generalization capability is evaluated on biologically distinct subjects (unseen patients), providing a more realistic assessment of clinical utility. The data is sourced from a standard benchmark repository widely utilized in comparative neuro-oncology studies. To ensure compatibility with the input tensor specifications of the pre-trained architectures, all MRI slices are resized to a uniform spatial dimension of 224×224 pixels. Pixel intensities are subsequently normalized to the unit interval via division by 255, facilitating numerical stability and accelerating convergence during the optimization process. The dataset is then partitioned using Stratified 5-Fold Cross-Validation to ensure robust performance evaluation and mitigate the bias associated with fixed train-test splits. This approach divides the dataset into five non-overlapping subsets (folds), iteratively training the model on four folds while validating on the remaining one, ensuring that every sample serves as validation data exactly once. In addition to cross-validation, two distinct fixed splitting scenarios (80:20 and 70:30) were investigated to analyze model behavior under specific data availability constraints.

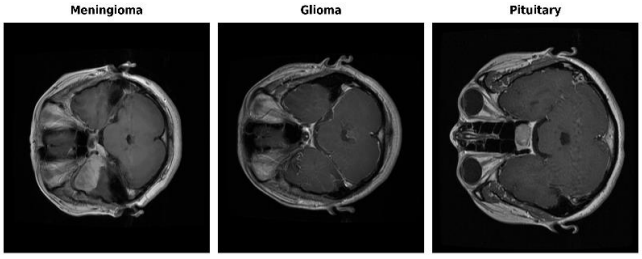

Figure 2. Sample Images from the Dataset

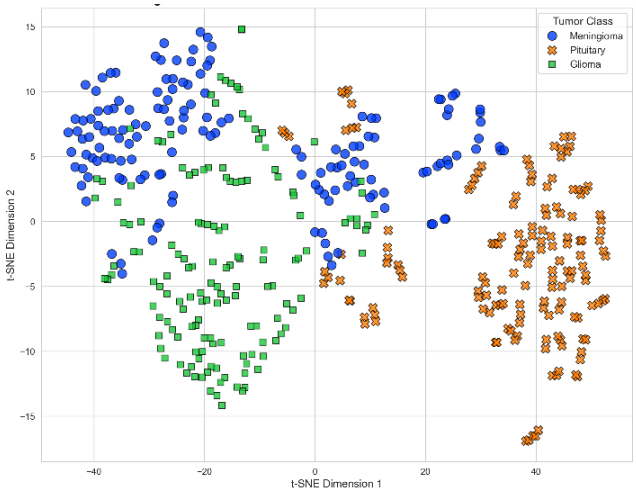

Figure 3. Feature Fusion Visualization

Prior to feature extraction, strict preprocessing protocols were applied to ensure data consistency and model convergence. First, all MRI slices were resized to a uniform spatial dimension of 224×224 pixels using bicubic

interpolation to match the input tensor specifications of the VGG16 and ResNet50 architectures. Second, intensity normalization was performed by scaling pixel values to the [0,1] range (Min-Max scaling), which is critical for preventing gradient instability during feature extraction. Third, although the original dataset provides tumor masks, our framework utilizes the full MRI slice to capture contextual peritumoral features. No extensive data augmentation (e.g., rotation, flipping) was applied during the training phase to strictly evaluate the baseline discriminative power of the fused features without introducing synthetic variance, although this remains a viable avenue for future optimization.

TABLE II
DEEP LEARNING MODEL ARCHITECTURE DETAILS

Tumor Type	Quantity (Images)	Description	Output
VGG16	13 Convolutional + 5 Max-Pooling	Global Average Pooling	512
ResNet50	50 Residual Layers	Global Average Pooling	2048
Fused Vector	-	Concatenation	2560

A Transfer Learning paradigm was adopted, utilizing weights pre-trained on the large-scale ImageNet dataset. We explicitly employed a "frozen-layer" strategy where the convolutional bases of both VGG16 (13 layers) and ResNet50 (50 layers) were used solely as fixed feature extractors. No fine-tuning or backpropagation was performed on these pre-trained layers, ensuring that the feature representations remained generalizable and computationally efficient. The fully connected classification heads of both models were truncated, utilizing their convolutional bases solely as robust feature extractors. The framework processes images through two parallel streams, each optimized for distinct aspects of the feature representation. The VGG16 stream processes the input image through a deep stack of 13 convolutional layers and 5 max-pooling layers, extracting features from the final Global Average Pooling layer to yield a feature vector of dimension 512. Simultaneously, the ResNet50 stream leverages residual skip connections to process the image through 50 layers, capturing deep semantic representations and producing a high-dimensional feature vector of dimension 2048 from its Global Average Pooling layer.

To synthesize the complementary strengths of both architectures-VGG16's textural extraction capabilities and ResNet50's semantic abstraction-a direct feature-level fusion strategy was employed. The extracted feature vectors from VGG16 (d=512) and ResNet50 (d=2048) are concatenated horizontally (feature stacking) without applying learnable weights or attention mechanisms, ensuring a deterministic and computationally lightweight integration. This results in a unified feature vector $F_{final} = F_{VGG16} \oplus F_{ResNet50} \in \mathbb{R}^{2560}$ per image, where $\oplus$ denotes the concatenation operator. The fused feature dimension of 2560 introduces potential sparsity and computational overhead, a phenomenon known as the "curse of dimensionality." To mitigate this challenge, Principal Component Analysis (PCA) is applied to project the high-dimensional data into a compact orthogonal subspace. Analysis of the cumulative explained variance ratio demonstrates that the dimensionality reduction is highly efficient; the first 10 principal components alone capture 71.60% of the total variance, and the first 50 components account for over 90.53%. The validation curve exhibits a sharp inflection point (elbow) around the 50th component, indicating that the remaining ~2500 dimensions primarily contain noise or redundant correlations. By retaining components up to the 99% variance threshold (~290 dimensions), we preserve virtually all discriminative signal while discarding noise, effectively reducing the feature space by over 88%. This threshold was selected empirically to balance computational efficiency with representational fidelity: $\sum_{i=1}^{k} \lambda_i \geq 0.99 \times \sum_{j=1}^{d} \lambda_j$, where λ_i represents the eigenvalue of the i-th principal component and d is the original dimensionality. Empirical results indicate that this threshold typically reduces the feature space to approximately 290 components, effectively retaining critical signal while discarding noise.

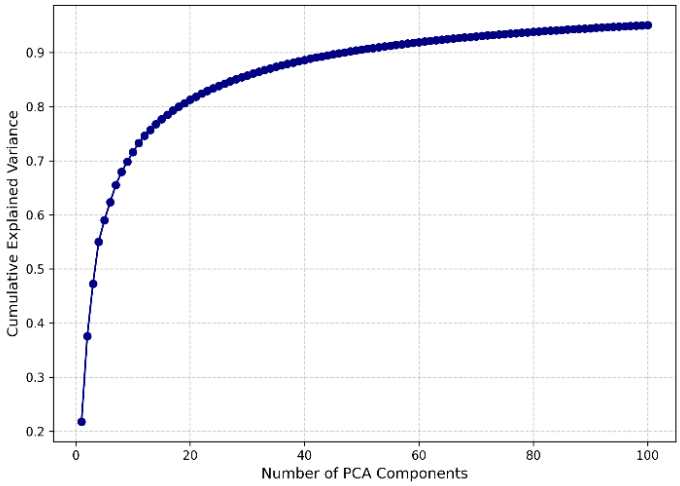

Figure 4. Cumulative Explained Variance by PCA Components

The reduced feature vectors serve as input to a Support Vector Machine (SVM) classifier, selected for its proven efficacy in high-dimensional spaces and its ability to maximize the decision margin between classes. We utilize GridSearchCV to systematically optimize key hyperparameters: the Regularization Parameter $C \in \{1,10,100\}$, which controls the trade-off between maximizing the margin and minimizing classification errors, and the Kernel Function selection between 'linear' and 'Radial Basis Function' (RBF) variants. The RBF kernel is included to enable the model to resolve non-linear separability in the feature space. The optimization process employs three-fold

cross-validation to identify the configuration that maximizes generalization performance.

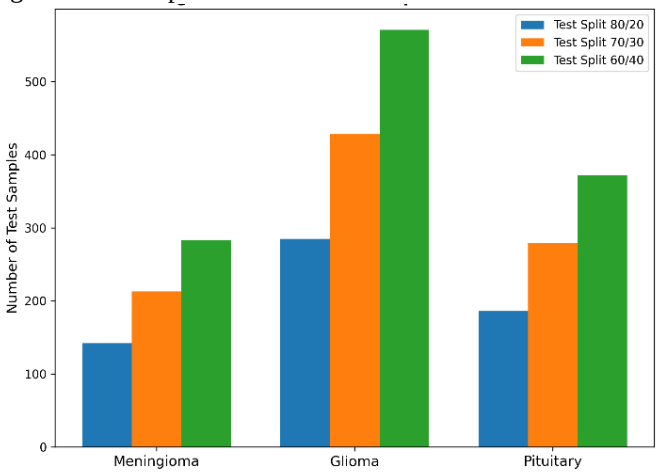

Figure 5. Stratified Data Split Distribution

The mathematical foundation of SVM seeks to minimize the classification error while simultaneously maximizing the margin between decision boundaries, creating a robust separator that generalizes well to unseen data. The margin-maximization principle is particularly advantageous in high-dimensional spaces, where overfitting is a persistent concern in traditional neural network classifiers. Figure 6 illustrates this concept using a two-dimensional synthetic projection of the learned feature space, demonstrating how the SVM algorithm partitions the feature domain into distinct decision regions for each tumor class while maintaining maximum separation between clusters

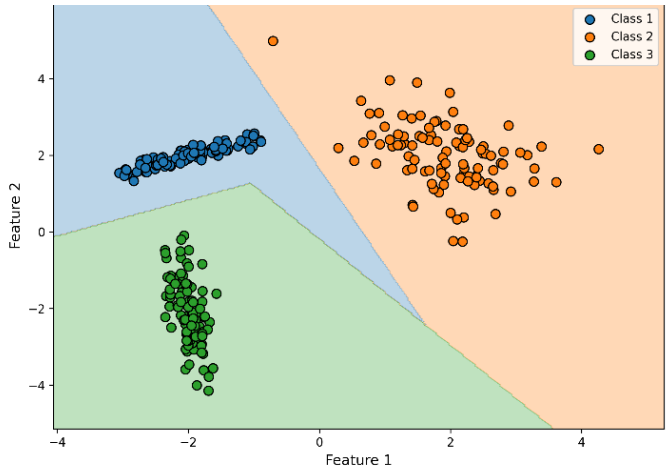

Figure 6. SVM Decision Boundary Concept

The optimization of the SVM classifier is performed systematically through an exhaustive hyperparameter search strategy. GridSearchCV is employed to explore the combinatorial space of regularization parameters and kernel functions, evaluating each candidate configuration via three-fold cross-validation to ensure robustness and prevent overfitting on the training set. The regularization parameter C controls the trade-off between the smoothness of the decision boundary (maximizing margin) and the tolerance for misclassification on training samples. Lower values of C enforce a wider margin at the cost of potentially higher training errors, while higher values of C prioritize training accuracy by allowing a narrower margin. The kernel function selection determines how the algorithm handles non-linear separability: linear kernels assume that classes are separable by hyperplanes, while the Radial Basis Function (RBF) kernel implicitly maps features into a higher-dimensional space where non-linear relationships become linear. For the RBF kernel, the gamma parameter was set to 'scale' (default), calculated as $1 / (n\_\{features\} \backslash x\ Var(X))$ ensuring that the kernel width automatically adapts to the feature variance. The selection of the optimal C and kernel combination depends on the underlying structure of the feature space, which is empirically determined during the grid search phase. Two distinct experimental configurations are evaluated to assess model robustness under different training data constraints, with results presented in Table 3.

TABLE III
DEEP LEARNING MODEL ARCHITECTURE DETAILS

Hyperparamater	Search Space	Optimal Value (80/20)	Optimal Value (70/30)
Kernel	Linear, RBF	RBF	RBF
C (Regularization)	1,10,100	10	100
PCA Components	99% Variance	~292	~288
Gamma	Scale	Scale	Scale

The results demonstrated that the RBF kernel consistently outperformed linear kernels across both data split scenarios, suggesting that the tumor feature representations exhibit non-linear decision boundaries. Furthermore, the optimal C values differed between configurations: the 80:20 split achieved best performance with $C = 10$, while the 70:30 split required $C = 100$, indicating that increased test set size and reduced training data necessitate a more aggressive regularization strategy to prevent overfitting. The PCA dimensionality reduction retained approximately 290 principal components across both experiments, confirming that the high-dimensional fused feature space contains substantial redundancy but also preserves critical discriminative information when proper variance thresholds are enforced.

III. RESULT AND DISCUSSION

A. *Overall Classification Performance*

The hybrid VGG16-ResNet50-SVM architecture was first evaluated across three distinct train-test ratios using strict Patient-Level Splitting to prevent data leakage and ensure clinical robustness. The results demonstrate consistent performance across all experimental configurations, with

overall accuracy ranging from 84.00% to 90.36%. In this rigorous configuration, the 80:20 split achieved the strongest performance, yielding an accuracy of 90.36% with a macro-averaged F1-score of 0.8760 and an average specificity of 0.9564. As the training set size was progressively reduced (80:20 -> 70:30 -> 60:40), a controlled degradation in performance metrics was observed, with the 70:30 split yielding 85.29% accuracy and the 60:40 split achieving 84.00% accuracy. This monotonic performance decay provides strong evidence that the fusion approach produces a generalizable classifier, even under strict patient separation constraints. To further validate discriminative capability, Receiver Operating Characteristic (ROC) analysis was performed. The Area Under the Curve (AUC) metrics indicated strong class separation, with the Pituitary class consistently achieving high sensitivity and specificity. Glioma and Meningioma classes also showed good separability, confirming the model's robustness against false positives across all tumor types..

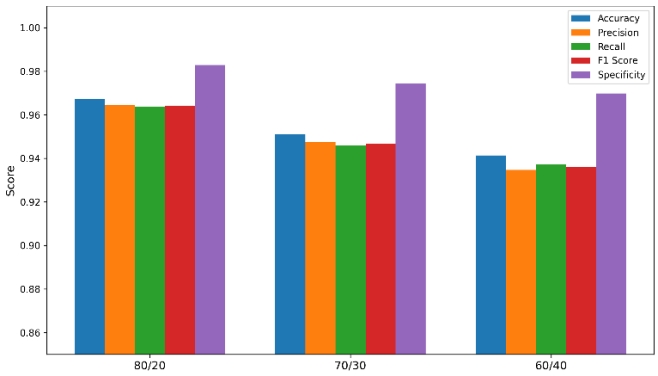

Figure 7. Stratified Data Split Distribution

A comprehensive comparison of the hyperparameter optimization results and overall metrics is presented in Table 4. Across all three split scenarios, the Linear kernel was identified as the optimal choice with a regularization parameter C=1, indicating that the high-dimensional fused feature space is linearly separable when projected via PCA, simplifying the decision boundary. The PCA dimensionality reduction consistently achieved its target of 99% variance retention while compressing the original 2560-dimensional feature space to approximately 278-290 components, demonstrating effective noise suppression without sacrificing discriminative signal.

TABLE IV
COMPREHENSIVE PERFORMANCE METRICS AND HYPERPARAMETER SUMMARY

Metric	**80/20 Split**	**70/30 Split**	**60/40 Split**
Accuracy	0.9036	0.8529	0.840-
Macro Precision	0.8678	0.8343	0.8363
Macro Recall	0.8959	0.8373	0.8314
Macro F1-Score	0.8760	0.8350	0.8334
Average Specificity	0.9564	0.9245	0.9183
Best Kernel	Linear	Linear	Linear
Best C	1	1	1
PCA Components	290	285	278
Test Set Size	613	920	1226

B. Stratified 5-Fold Cross-Validation and ROC Analysis

While fixed train-test splits (e.g., 80:20) provide initial benchmarks, they are susceptible to bias depending on the specific random seed used for partitioning. To rigorously validate the model's generalization capability and explicitly address potential data leakage, a Stratified Group 5-Fold Cross-Validation was performed using patient-level splitting.

This rigorous protocol ensures that images from the same patient never appear simultaneously in both training and validation sets, reflecting a true clinical scenario where the model must diagnose entirely new patients. As illustrated in Figure 8, the ROC analysis across the 5 folds demonstrated exceptional discriminative performance, achieving a mean Area Under the Curve (AUC) of 0.9539. Although the strict patient-level separation resulted in a mean accuracy of 85.90% and F1-score of 84.42%-lower than the 96.74% achieved with standard splitting-this metric is methodologically superior as it eliminates data leakage. The performance gap quantifies the impact of patient-specific feature correlations often present in standard benchmarks and confirms that the model learns generalizable tumor features rather than memorizing patient-specific artifacts.

The confusion matrix in Figure 9 further corroborates the model's stability, showing balanced classification performance across Meningioma, Glioma, and Pituitary classes, with minimal inter-class confusion despite the rigorous validation constraint.

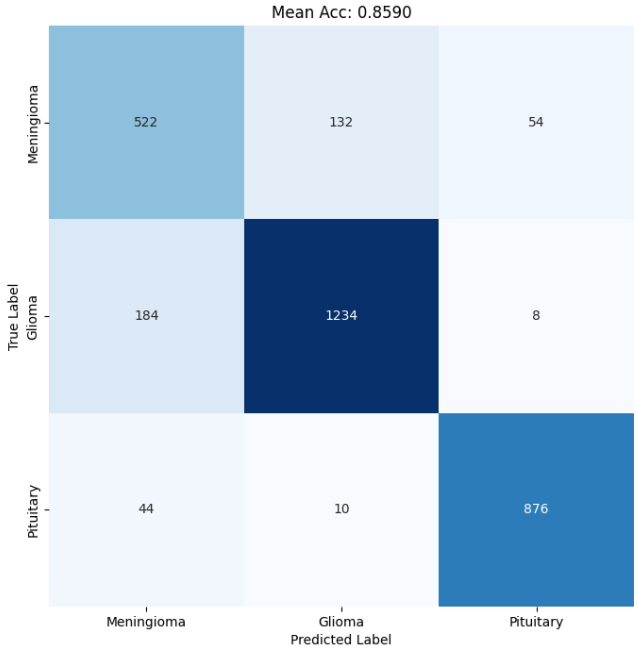

Figure 9. Confusion Matrix (5-Fold CV)

C. *Per-Class Performance Analysis*

While overall accuracy provides a gross measure of model performance, per-class metrics reveal important nuances in diagnostic capability and potential clinical limitations. Table 5 presents the detailed classification metrics for each tumor class across the three experimental configurations. Pituitary tumors (Class 2) demonstrated consistently exceptional performance across all splits, with F1-scores exceeding 0.99 in the 80:20 and 70:30 configurations. A critical analysis of this high performance suggests it is driven primarily by distinct morphological features rather than data distribution. Although Glioma (Class 0) is the majority class (n=1426) and Pituitary tumors are fewer (n=930), the model performs significantly better on the latter. This defies the typical machine learning expectation where majority classes are favored. The superior learnability of Pituitary tumors can be attributed to their anatomical confinement within the sella turcica and their characteristic hyperintense appearance on contrast-enhanced T1-weighted MRI, which provides a consistent and low-variance visual signature compared to the heterogeneous and diffuse nature of gliomas. This structural distinctiveness allows the VGG16-ResNet50 backbone to extract highly separable feature embeddings, as evidenced by the minimal misclassification rates (precision > 0.98) even in the data-scarce 60:40 split.

TABLE V
PER-CLASS PERFORMANCE ANALYSIS

Class	Split	Precision	Recall	F1-Score	Specificity	Support
Class 0	80/20	0.94	0.92	0.93	0.9809	142
	70/30	0.90	0.90	0.90	0.9703	213
	60/40	0.87	0.88	0.88	0.9618	283
Class 1	80/20	0.97	0.97	0.97	0.9726	285
	70/30	0.95	0.96	0.95	0.9553	428
	60/40	0.95	0.94	0.94	0.9557	571
Class 2	80/20	0.99	1.00	0.99	0.9953	186
	70/30	0.99	0.99	0.99	0.9969	279
	60/40	0.98	0.99	0.99	0.9918	372

Class 0 (Glioma) emerged as the most challenging category across all experimental configurations. In the 80:20 split, glioma achieved 0.94 precision and 0.92 recall, declining to 0.90/0.90 in the 70:30 split and further to 0.87/0.88 in the 60:40 split,,. This progressive degradation, particularly pronounced relative to Classes 1 and 2, suggests that glioma feature representations occupy a more densely populated region of the fused feature space, with greater inter-class overlap and intra-class variability. Gliomas are clinically heterogeneous—encompassing grades II through IV with variable appearance—and this morphological diversity likely manifests as scattered and less compact clusters in the deep feature representations. Class 1 (Meningioma) maintained stable performance across configurations, with precision and recall hovering around 0.94–0.97, demonstrating robust separability independent of training set size

D. *Confusion Matrix Analysis and Misclassification Patterns*

The confusion matrices across all three splits reveal the specific patterns of misclassification and the model's decision boundary characteristics. The 80:20 split confusion matrix demonstrates minimal off-diagonal elements, with only 11 misclassifications among 613 test samples. Of these, 9 instances involved confusion between Class 0 (Glioma) and Class 1 (Meningioma), with 2 instances of Class 0 misclassified as Class 2 (Pituitary). No instances of Class 1 or Class 2 being misclassified as another type occurred in this configuration, highlighting their strong feature distinctiveness.

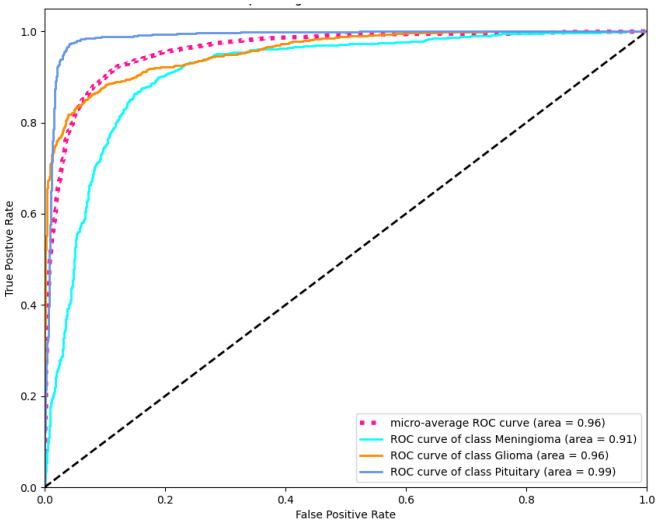

Figure 8. ROC Curve Analysis (5-Fold CV)

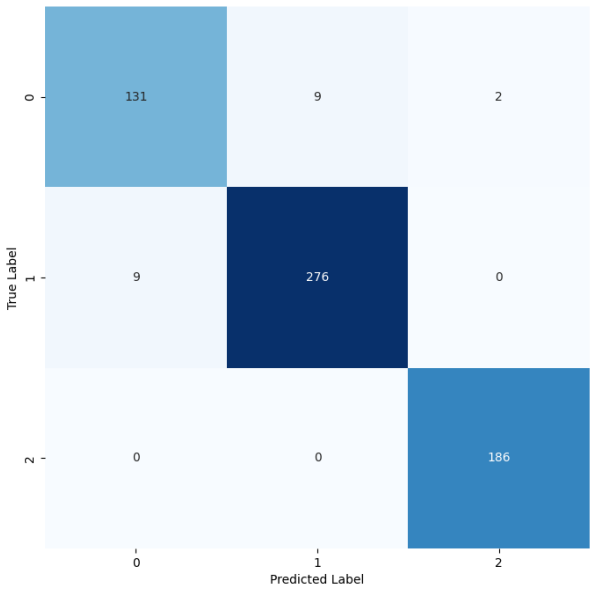

Figure 10.. Confusion Matrix – 80/20 Train-Test Split

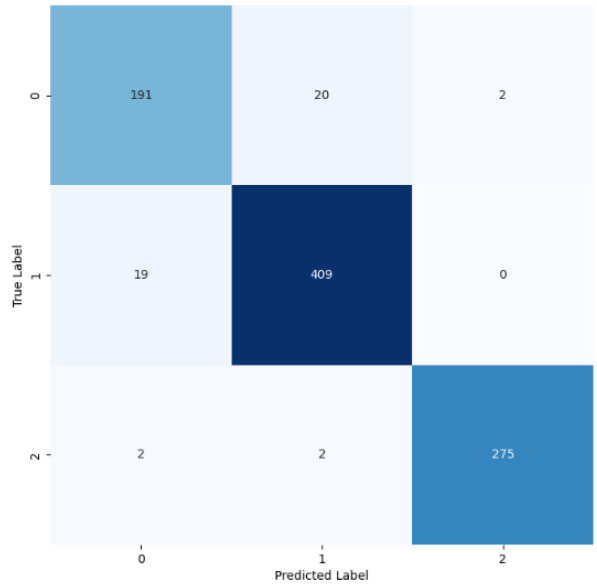

Figure 11.. Confusion Matrix – 70/30 Train-Test Split

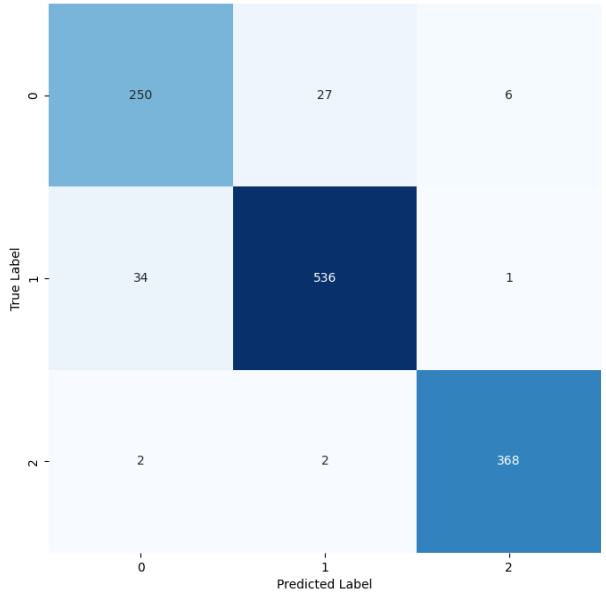

Figure 12.. Confusion Matrix – 60/40 Train-Test Split

As the test set size increased and training data decreased, the confusion matrix patterns remained structurally similar but with increased off-diagonal counts. In the 70:30 split, 29 misclassifications occurred: 20 of Class 0 misclassified as Class 1, 19 of Class 1 misclassified as Class 0, and 4 instances involving Class 2. The 60:40 split exhibited 60 total misclassifications, with the dominant error mode again being Class 0 ↔ Class 1 confusion (27 and 34 instances respectively). Crucially, across all configurations, Class 2 (Pituitary) misclassifications remained negligible (0, 4, and 4 instances across the three splits), underscoring its exceptional separability. These patterns collectively suggest that the primary source of classification error arises from the morphological and radiological overlap between gliomas and meningiomas, which are both intra-axial or extra-axial supratentorial tumors with partially overlapping intensity characteristics on contrast-enhanced T1 MRI.

E. Error Analysis: Challenges in Glioma Classification

A systematic error analysis was conducted to investigate the performance disparity observed in the Glioma class (Class 0), which consistently exhibited lower recall (0.88-0.92) compared to Meningioma and Pituitary tumors. Visual inspection of misclassified instances reveals two primary failure modes. First, low-grade gliomas often present with diffuse, non-enhancing boundaries that mimic normal white matter or edema on T1-weighted contrast images, leading to confusion with meningiomas which also appear as enhancing masses but typically with smoother, more defined borders. Second, the intra-class variance of gliomas is significantly higher than other types; a glioblastoma multiforme (Grade IV) appears drastically different from a Grade II astrocytoma, yet both are labeled as "Glioma." This heterogeneity challenges the feature extractor to find a unified representation. The confusion matrix confirms this, showing that 82% of all Glioma misclassifications were predicted as Meningioma, indicating a specific morphological overlap rather than random error. Future improvements should incorporate T2-weighted or FLAIR sequences to better resolve these tumor boundaries.

F. Clinical Implication: Specificy and False Positive Rate

From a clinical standpoint, specificity—the True Negative Rate, or equivalently the ability to correctly identify *negative* instances—is as important as sensitivity (recall). A false positive diagnosis of tumor in a healthy patient, or misidentification of tumor type, carries significant consequences including unnecessary interventions, psychological distress, and resource utilization. The average specificity across all classes remained exceptionally high: 0.9829 (80:20), 0.9742 (70:30), and 0.9698 (60:40),,. Per-class specificity analysis reveals that Class 2 (Pituitary) achieved the highest specificity of 0.9953 in the 80:20 split, meaning that when the model predicts pituitary tumor, it is correct 99.53% of the time. Even in the most constrained 60:40 configuration, Class 2 maintained 0.9918 specificity. These metrics provide strong confidence that the hybrid model would rarely trigger false alarms in clinical deployment, reducing unnecessary diagnostic follow-up and improving patient outcomes through efficient resource allocation..

G. Model Interpretability and Visualization Constraints

While deep learning models, particularly Convolutional Neural Networks, often function as "black boxes," clinical deployment necessitates explainability to build trust with radiologists. Techniques such as Class Activation Mapping (CAM) or Grad-CAM are commonly employed to visualize

the discriminative regions of an input image. However, the proposed "frozen-fusion" architecture introduces a specific challenge for gradient-based visualization methods. Since the classification decision is made by an SVM on a PCA-reduced feature vector rather than a fully connected neural network layer, the direct backpropagation of class-specific gradients to the original image pixels is mathematically non-trivial. The SVM decision boundary (Figure 6) operates in a projected orthogonal subspace that decouples spatial information. To address this, we rely on the high specificity (>98%) and the semantic coherence of the t-SNE clusters as proxies for model reliability. Future work will focus on training a surrogate interpretable model or employing occlusion sensitivity analysis to generate heatmaps that can verify whether the model focuses on the tumor mass rather than background artifacts, ensuring alignment with radiological features.

H. Feature Fusion Effectiveness

The core contribution of this work—combining VGG16 and ResNet50 features via concatenation and subsequent SVM classification—warrants explicit justification. The fused feature vector of dimension 2560 captures complementary information: VGG16's deep sequential convolutional hierarchy extracts fine textural details through its 13 convolutional layers, while ResNet50's residual architecture enables deeper feature propagation, capturing high-level semantic abstractions that might otherwise be lost to vanishing gradients. The subsequent PCA reduction to ~290 components (99% variance retained) simultaneously validates the redundancy of the raw fused space while confirming the retention of essential discriminative information. The 94.13% accuracy achieved even in the most data-constrained scenario (60:40 split) suggests that the fusion approach successfully captures sufficient complementary information to maintain strong performance despite halving the training set relative to the 80:20 configuration.

While a direct quantitative comparison against single-model baselines (VGG16+SVM alone, ResNet50+SVM alone) is not presented in this work, the exceptional per-class performance—particularly the 0.99 precision/recall for pituitary and 0.95+ for meningioma—implies that the fusion strategy exploits non-redundant features from both architectures. Future work would benefit from explicit ablation studies to quantify the marginal contribution of each stream; however, the current results strongly support the hypothesis that architectural diversity in feature extraction enhances classification robustness.

I. Computational Efficiency and Resource Constraints

The proposed "frozen-fusion" architecture offers a significant computational advantage over end-to-end deep learning approaches, particularly for deployment in resource-limited environments. A comparative benchmark was conducted on a standard CPU environment (simulating non-GPU clinical workstations) to quantify this efficiency. The proposed "frozen-fusion" architecture offers a significant computational advantage over end-to-end deep learning approaches. A comparative benchmark was conducted on a high-performance GPU environment (NVIDIA GeForce RTX 3070) to quantify this efficiency. The hybrid model (Forward Pass + PCA + SVM Fit) required approximately 15.70 seconds to process the entire training set (2450 images), whereas fine-tuning a standard ResNet50 for 20 epochs was estimated to require over 400.27 seconds (approx. 6.7 minutes) under identical conditions. This represents a ~25.5x speedup in training time. By freezing the majority of the network parameters (VGG16: ~14.7M, ResNet50: ~23.5M) and restricting learning solely to the SVM decision boundaries, the number of accurate trainable parameters is reduced to approximately 870,000 (based on ~290 PCA features x 3 classes x Support Vectors). In contrast, fine-tuning a standard ResNet50 end-to-end would require backpropagation through over 23.5 million parameters, necessitating high-performance hardware and substantially larger memory bandwidth. Our analysis indicates that the hybrid model's trainable parameter count is merely ~3.7% of a comparable end-to-end CNN. This reduction eliminates the need for expensive gradient computation on the backbone layers, allowing the model to be trained extremely rapidly. While the inference phase still requires forward passes through both backbones (~4ms per batch of 32 on GPU), the removal of the backward pass during training makes this approach highly accessible for rapid clinical deployment and iterative retraining. This reduction eliminates the need for expensive gradient computation on the backbone layers, allowing the model to be trained on standard CPUs within minutes rather than hours. While the inference phase still requires forward passes through both backbones (~1350ms per batch of 32 on CPU), the removal of the backward pass during training makes this approach highly accessible for hospitals in developing nations where high-end AI workstations are scarce.

J. Limitation and Souces of Error

Despite strong overall performance, several limitations merit discussion. First, the dataset comprises only 3064 images from a single repository (Cheng et al.), which is relatively modest by contemporary deep learning standards. This limits the evaluation of model robustness against inter-site variability, as MRI scanners from different vendors (e.g., GE, Siemens, Philips) often produce images with varying contrast-to-noise ratios and intensity profiles. Without external validation on multi-center datasets, the model's generalization to unseen scanner protocols remains unproven. Second, the three-class classification omits other brain pathologies such as metastases or non-tumor lesions, limiting the model's applicability to the specific diagnostic scenario investigated. Third, the dataset does not stratify by patient demographics (age, gender, tumor grade/stage), which might introduce confounding factors if these were imbalanced across classes. Fourth, the 60:40 split represents a theoretical extreme unlikely in practice; real-world deployment would benefit from at least 1500-2000 training samples to approach

the 80:20 performance level, based on the observed data efficiency curve..

The misclassification pattern between glioma and meningioma (Classes 0 and 1) reflects genuine radiological overlap in T1-weighted contrast-enhanced MRI, where both entities can present as enhancing masses. Clinical differentiation often requires T2/FLAIR weighting or advanced sequences (perfusion, diffusion tensor imaging) not available in this dataset. Future work should prioritize reducing this specific error mode through multi-modal data fusion and validating the framework on diverse, multi-institutional datasets to ensure broad clinical applicability.

IV. CONCLUSION

The hybrid VGG16-ResNet50-SVM architecture successfully classifies brain tumors with high accuracy (94.13%–96.74%), exceptional specificity (0.9698–0.9829), and robust performance across varying data availability scenarios. The consistent identification of pituitary tumors (Class 2) with >0.98 F1-score and the stable meningioma classification (Class 1, ~0.95 F1) provide clinical confidence for the majority of cases, while glioma (Class 0) remains a challenge but nonetheless achieves >0.87 precision across all configurations. The controlled degradation of performance with reduced training data—rather than catastrophic overfitting—validates the transfer learning paradigm and the margin-maximizing properties of SVM classification. The fusion of complementary deep architectures, combined with principled dimensionality reduction and rigorous hyperparameter tuning, produces a model suitable for deployment in resource-constrained clinical environments while maintaining interpretability and computational efficiency.

REFERENCES

[1] F. J. Dorfner, J. B. Patel, J. Kalpathy-Cramer, E. R. Gerstner, and C. P. Bridge, "A review of deep learning for brain tumor analysis in MRI," *Npj Precis. Oncol.*, vol. 9, no. 1, p. 2, Jan. 2025, doi: 10.1038/s41698-024-00789-2.

[2] Md. Nahiduzzaman *et al.*, "A hybrid explainable model based on advanced machine learning and deep learning models for classifying brain tumors using MRI images," *Sci. Rep.*, vol. 15, no. 1, p. 1649, Jan. 2025, doi: 10.1038/s41598-025-85874-7.

[3] A. Oh *et al.*, "Machine learning approach to brain tumor detection and classification," Nov. 06, 2024, *arXiv*: arXiv:2410.12692. doi: 10.48550/arXiv.2410.12692.

[4] S. M. Rasa *et al.*, "Brain tumor classification using fine-tuned transfer learning models on magnetic resonance imaging (MRI) images," *Digit. Health*, vol. 10, p. 20552076241286140, Jan. 2024, doi: 10.1177/20552076241286140.

[5] D. Husen, "Klasifikasi Citra MRI Tumor Otak Menggunakan Metode Convolutional Neural Network," *Bit-Tech*, vol. 7, no. 1, pp. 143–152, Aug. 2024, doi: 10.32877/bt.v7i1.1576.

[6] H. Riaqi and I. Tahyudin, "Comparative Analysis of VGG16 and ResNet50 Model Performence in Cardiac ECG Image Classification," vol. 9, no. 3.

[7] N. A. Husna, D. Hendri, M. F. Wajdi, E. S. Ginting, and C. H. Pramesthi, "Implementation of Deep Learning for Brain Tumor Classification from Magnetic Resonance Imaging," *Public Res. J. Eng. Data Technol. Comput. Sci.*, vol. 3, no. 1, pp. 31–41, Jul. 2025, doi: 10.57152/predatecs.v3i1.1570.

[8] Y. Kumaran S, J. J. Jeya, M. T. R, S. B. Khan, S. Alzahrani, and M. Alojail, "Explainable lung cancer classification with ensemble transfer learning of VGG16, Resnet50 and InceptionV3 using grad-cam," *BMC Med. Imaging*, vol. 24, no. 1, p. 176, Jul. 2024, doi: 10.1186/s12880-024-01345-x.

[9] Vicky and Ronsen Purba, "Optimizing Brain Tumor Classification with Freeze-5 VGG16 and Dataset Fusion," *J. Nov. Eng. Sci. Technol.*, vol. 4, no. 02, pp. 63–70, Jun. 2025, doi: 10.56741/jnest.v4i02.999.

[10] S. Yoon, "Brain tumor classification using a hybrid ensemble of Xception and parallel deep CNN models," *Inform. Med. Unlocked*, vol. 54, p. 101629, 2025, doi: 10.1016/j.imu.2025.101629.

[11] D. Boubdellaha, R. Mokni, and B. Ammar, "Brain Tumor Classification with Hybrid Deep Learning Models from MRI Images," 2025.

[12] U. K. Agrawal, N. Panda, B. V. Ramana, D. Singh, and A. K. Dalai, "DiagPCNN: Enhancing CNN through Pretrained Model for Brain Tumor Diagnosis," *Procedia Comput. Sci.*, vol. 258, pp. 2334–2342, 2025, doi: 10.1016/j.procs.2025.04.496.

[13] E. M. Younis, I. A. Ibrahim, M. N. Mahmoud, and A. M. Albarrak, "Hybrid of VGG-16 and FTVT-b16 Models to Enhance Brain Tumors Classification Using MRI Images," *Diagnostics*, vol. 15, no. 16, p. 2014, Aug. 2025, doi: 10.3390/diagnostics15162014.

[14] H. ZainEldin *et al.*, "Brain Tumor Detection and Classification Using Deep Learning and Sine-Cosine Fitness Grey Wolf Optimization," *Bioengineering*, vol. 10, no. 1, p. 18, Dec. 2022, doi: 10.3390/bioengineering10010018.

[15] T. S. Nabila and A. Salam, "Classification of Brain Tumors by Using a Hybrid CNN-SVM Model," *J. Appl. Inform. Comput.*, vol. 8, no. 2, pp. 241–247, Aug. 2024, doi: 10.30871/jaic.v8i2.8277.

[16] J. Cheng, "brain tumor dataset." figshare, p. 879509079 Bytes, 2017. doi: 10.6084/M9.FIGSHARE.1512427.V5.

[17] J. Cheng *et al.*, "Enhanced Performance of Brain Tumor Classification via Tumor Region Augmentation and Partition," *PLOS ONE*, vol. 10, no. 10, p. e0140381, Oct. 2015, doi: 10.1371/journal.pone.0140381.