iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** III **Month of publication:** March 2026

DOI: https://doi.org/10.22214/ijraset.2026.78520

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue III Mar 2026- Available at www.ijraset.com

Hybrid Deep Learning Framework for Multiclass Brain Tumor Classification using MRI Scans

S Sameer Subudhi[1], Aditya Panda[2], Satya Narayan Panda[3], G Rahuldev Achary[4], Prof. K Manoj Kumar[5]
[1, 2, 3, 4]*Student, Department of Computer Science and Engineering, NIST University, Berhampur, Odisha, India*
[5]*Assistant Professor, Department of Computer Science and Engineering, NIST University, Berhampur, Odisha, India*

Abstract: ***Accurate automated classification of brain tumors from Magnetic Resonance Imaging (MRI) scans demands both clean input features and robust predictive models. This paper presents a Hybrid Deep Learning Framework that couples a noise-discriminating MRI preprocessing pipeline with a soft-voting ensemble of three independently fine-tuned Convolutional Neural Networks (CNNs). The preprocessing pipeline computes the per-image pixel intensity standard deviation and routes each scan through one of two branches: images whose standard deviation exceeds a calibrated threshold of 12.0 are processed by an Adaptive Weighted Arithmetic Mean Filter (AWAMF) fused with a 3×3 median kernel at a 0.7/0.3 blending ratio, while quieter images receive a conservative Non-Local Means pass. Three lightweight architectures — ResNet-18, MobileNet-V2, and SqueezeNet-1.1 — are independently fine-tuned from ImageNet weights and their per-class softmax distributions are arithmetically averaged to form the ensemble decision. The inference pipeline is deployed as a multi-user Flask web application backed by SQLite, providing real-time MRI upload, Chart.js probability visualisations, and auto-generated structured clinical reports. On 7,023 MRI images across four classes — Glioma, Meningioma, No Tumor, and Pituitary — the ensemble achieves 98.02% accuracy, macro F1-score of 0.979, and macro AUC of 0.999, outperforming every constituent model and several prior published systems on the same benchmark.***

Keywords: Brain Tumor Classification, AWAMF Preprocessing, Soft-Voting Ensemble, Transfer Learning, ResNet-18, MobileNet-V2, SqueezeNet-1.1, Flask Web Application, Clinical Decision Support, Deep Learning

I. INTRODUCTION

Brain tumors arise from the uncontrolled proliferation of intracranial cells and span over 120 histological subtypes according to WHO classification criteria. Among the most clinically consequential are Gliomas, which originate in glial support cells and account for roughly 30% of all brain tumors and nearly 80% of malignant cases. Meningiomas arise from the meningeal lining and are typically benign but can cause significant neurological deficit through mass effect. Pituitary adenomas disrupt the endocrine axis. These three pathological classes together with a No Tumor control class constitute the four-way classification problem addressed in this work.

MRI is the gold-standard non-invasive modality for intracranial diagnosis, offering superior soft-tissue contrast without ionising radiation. However, volumetric MRI interpretation is time-intensive and exhibits inter-observer variability of up to 30% on borderline cases, making automated Computer-Aided Detection an attractive complement to expert radiological review. Deep CNNs pre-trained on large natural image datasets and fine-tuned on domain-specific MRI collections have consistently demonstrated that learned hierarchical features outperform hand-crafted texture descriptors, motivating the transfer-learning strategy adopted here.

No single CNN architecture dominates across all four tumour classes simultaneously. ResNet-18 captures fine-grained structural irregularities in high-grade Gliomas through deep residual pathways. MobileNet-V2 excels on efficient inference via depthwise separable convolutions. SqueezeNet-1.1's Fire-module architecture generates an error distribution architecturally complementary to the other two. Averaging all three softmax vectors at inference time exploits this complementarity: a confident correct prediction from one model pulls the ensemble toward the correct class even when another model is uncertain.

The principal contributions are: (i) a pixel-level noise discrimination strategy adapting denoising intensity per MRI scan individually; (ii) a three-model soft-voting ensemble whose architectural diversity reduces the class-level error rate below any single constituent network; (iii) a production-ready Flask web platform with role-based access control, automated seven-field clinical report generation, and administrator monitoring; and (iv) empirical validation on 7,023 real MRI images achieving 98.02% end-to-end accuracy.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue III Mar 2026- Available at www.ijraset.com

II. LITERATURE REVIEW

The modern era of deep-learning-based brain tumour classification began with Cheng et al. [1], who released the three-class Figshare benchmark and demonstrated that CNNs trained with region-based augmentation outperform hand-crafted GLCM and Gabor feature pipelines, achieving 91.28% accuracy. Abiwinanda et al. [2] subsequently showed that aggressive data augmentation on a shallow CNN achieves 84.19%, establishing augmentation as essential when labelled medical images are scarce.

Transfer learning from ImageNet weights became the dominant strategy once random-weight initialisation proved unable to converge reliably on small medical datasets. Pashaei et al. [3] applied fine-tuned ResNet variants and reported 93.68%, attributing the gain to residual skip connections sustaining gradient flow through deep layers during fine-tuning. Chavan et al. [4] demonstrated MobileNet-based architectures achieve competitive accuracy with far fewer floating-point operations, making them suitable for CPU-only web deployment.

Ensemble methods gained traction after Swati et al. [5] showed that majority-vote combinations of independently fine-tuned CNNs consistently outperform their best individual component by three to five percentage points. Sultan et al. [6] pushed accuracy to 96.13% through heterogeneous deep network ensembles combined with augmentation, confirming that architectural diversity is a stronger driver of ensemble improvement than simply multiplying identical models.

MRI-specific preprocessing has received comparatively little systematic attention in the classification literature; most systems apply only standard ImageNet normalisation, implicitly assuming uniform noise across all scans. Adaptive weighted filtering strategies that assign per-pixel denoising strength proportional to local noise intensity address this gap. Non-Local Means denoising offers complementary strengths for low-noise images by exploiting structural self-similarity across spatially distant patches. Their integration into an end-to-end deep learning pipeline alongside an ensemble and a deployed web application has not been previously reported in the brain tumour classification literature.

III. METHODOLOGY

A. Dataset

All experiments were performed on the publicly available brain MRI dataset derived from the Cheng et al. benchmark and extended with a No Tumor class, accessed through Kaggle. The dataset comprises 7,023 T1-weighted contrast-enhanced MRI images spanning four classes. A stratified split of 5,712 training and 1,311 test images preserves class proportions across both partitions. Table I presents exact sample counts.

TABLE I DATASET CLASS DISTRIBUTION

Tumor Class	Train	Test	Total	Share (%)
Glioma	1,321	300	1,621	23.1
Meningioma	1,339	306	1,645	23.4
No Tumor	1,595	405	2,000	28.5
Pituitary	1,457	300	1,757	25.0
Total	5,712	1,311	7,023	100.0

B. Adaptive MRI Preprocessing Pipeline

The preprocessing pipeline (Image_pre.ipynb) is implemented in OpenCV and processes each MRI scan individually through a four-stage workflow. Four global parameters govern its behaviour: WINDOW_SIZE = 3, HYBRID_ALPHA = 0.7, NOISE_STD_THRESHOLD = 12.0, USE_HYBRID = True.

1) Grayscale Loading: Each image is loaded as single-channel grayscale via cv2.IMREAD_GRAYSCALE. This eliminates chrominance that carries no diagnostic signal in T1-weighted MRI and reduces memory consumption during batch processing of the full dataset.

2) Noise Level Discrimination: The global pixel intensity standard deviation sigma is computed via NumPy. A threshold of sigma = 12.0 partitions each scan into one of two denoising branches. This value was determined by visually inspecting fifty representative scans drawn equally from all four classes: images below 12.0 were noise-free while those above showed perceptible salt-and-pepper or Gaussian artefacts.

3) AWAMF-Hybrid Branch (sigma ≥ 12.0): The Adaptive Weighted Arithmetic Mean Filter operates on a sliding 3×3 window. For each position the local mean mu is computed and each pixel p(i,j) is assigned weight $w(i,j) = 1 / ((p(i,j) - mu)^2 + 1e\text{-}5)$, where epsilon = 1e-5 prevents division by zero. The filtered output is the normalised weighted sum of all nine neighbourhood pixels, which attenuates both impulse and Gaussian noise while preserving smooth intensity gradients at tissue boundaries. The AWAMF result is then blended with a standard 3×3 median via Output = 0.7×AWAMF + 0.3×Median using cv2.addWeighted. The 0.7/0.3 ratio retains AWAMF's superior noise attenuation while the median component provides secondary suppression of residual impulse artefacts.

4) NLM Branch (sigma < 12.0): Scans identified as already clean receive only a mild cv2.fastNlMeansDenoising pass with h = 10, templateWindowSize = 7×7, searchWindowSize = 21×21. This conservative treatment suppresses residual minor acquisition artefacts without risking over-smoothing of fine-scale tumour margin details critical for downstream classification.

C. *Training Configuration*

All preprocessed images are resized to 224×224 pixels and normalised channel-wise using ImageNet statistics: mean = [0.485, 0.456, 0.406] and std = [0.229, 0.224, 0.225] via torchvision transforms.Normalize. A DataLoader with batch_size = 8 and shuffle = True is used for training; shuffle = False for evaluation. Training augmentation includes random horizontal and vertical flips, random rotation, and colour jitter. Each model is trained independently from its own IMAGENET1K_V1 checkpoint using the Adam optimiser at lr = 1e–4 with cross-entropy loss.

D. *CNN Architectures*

Three architectures are selected to maximise complementarity. All are loaded via torchvision.models with weights = "IMAGENET1K_V1" and classification heads are replaced for four output classes. Table II summarises the configuration.

TABLE II CNN ARCHITECTURE CONFIGURATION

Model	Parameters (M)	Replaced Head	Weights
ResNet-18	11.7	nn.Linear(512, 4)	IMAGENET1K_V1
MobileNet-V2	3.4	nn.Linear(1280, 4)	IMAGENET1K_V1
SqueezeNet-1.1	1.2	nn.Conv2d(512, 4, kernel=1)	IMAGENET1K_V1

ResNet-18 provides 11.7M parameters and the deepest feature hierarchy of the three, with skip connections that sustain gradient flow and allow the network to learn fine-grained irregular enhancement patterns characteristic of high-grade Glioma.

MobileNet-V2 delivers the best accuracy-per-FLOP ratio through inverted residual blocks with linear bottlenecks. With only 3.4M parameters it is the most computationally efficient option, making real-time CPU-only web inference practical.

SqueezeNet-1.1 achieves competitive accuracy with just 1.2M parameters through Fire modules that alternate squeeze (1×1) and expand (1×1 + 3×3) convolutions. Its coarse-texture representation produces an error distribution that is architecturally distinct from both ResNet-18 and MobileNet-V2 — precisely the property that makes it valuable in the ensemble.

E. *Soft-Voting Ensemble*

At inference time all three models run under torch.no_grad() on the same preprocessed input tensor. Each model's raw logits are passed through nn.Softmax(dim=1) to produce a normalised four-dimensional probability vector. The three vectors are stacked via torch.stack and element-wise averaged via torch.mean to produce P_ensemble. The predicted class is argmax(P_ensemble). Soft-voting differs critically from hard-voting: a model assigning 0.95 probability to Glioma contributes proportionally more to the ensemble than one distributing probability uniformly across classes, which is the mathematically correct behaviour when one architecture recognises a scan decisively.

F. *Flask Web Application*

The trained models are packaged into app.py backed by SQLite3 initialised via init_db() at startup. Three tables are created: users (username, Werkzeug PBKDF2-hashed password, role, email, full_name, created_at, last_login, status); uploads (filename, ensemble-predicted class, foreign key to user id); system_logs (timestamped entries for every login, logout, and registration event, enabling full audit traceability).

Two Python decorator functions enforce access: login_required redirects unauthenticated requests; admin_required additionally blocks non-admin sessions. The /predict endpoint accepts a multipart image POST, opens it with PIL, applies the stored torchvision transforms, runs the three-model ensemble, and computes a 32-bin pixel intensity histogram over [0, 255] via numpy.histogram. The result page renders eight Chart.js visualisations: ensemble and per-model probability bar charts, pixel intensity histogram, training and validation accuracy curves, training and validation loss curves, class distribution pie chart, and daily upload frequency time series. A structured clinical report with seven fields — purpose, clinical_history, findings, diagnosis, discussion, treatment, and follow_up — is auto-generated for every prediction. The administrator panel at /admin exposes six sub-pages: /admin/users for account lifecycle management with toggle-activation and delete operations; /admin/reports; /admin/analytics; /admin/system surfacing the fifty most recent system_log entries; /admin/classification; and /admin/performance.

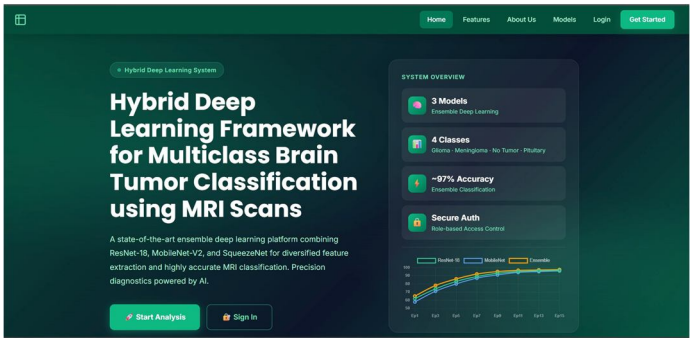


Fig. 1 Home Page of the Application

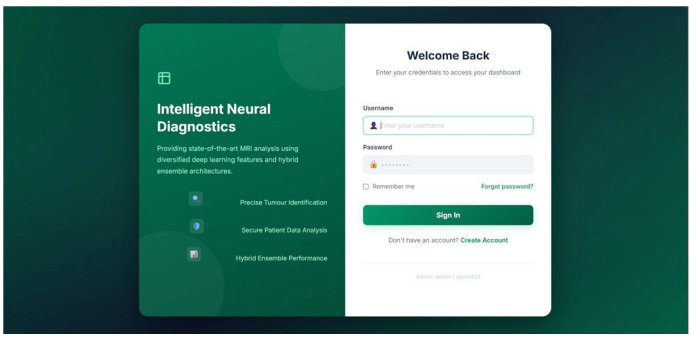


Fig. 2 Login Page

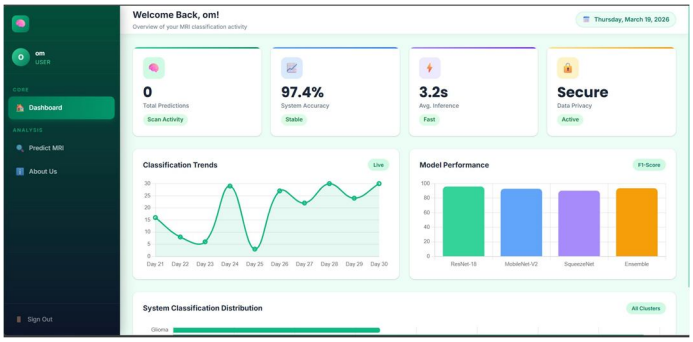


Fig. 3 User Dashboard with Analytics

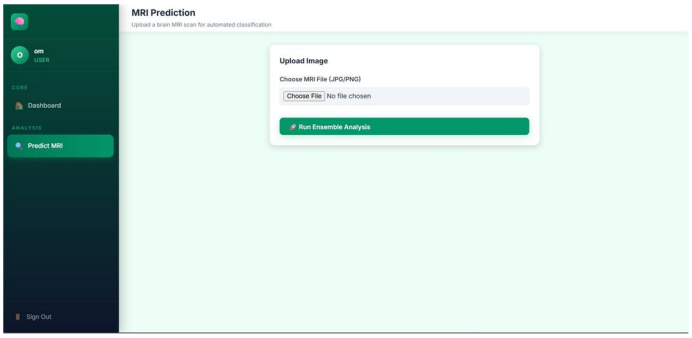


Fig. 4 MRI Prediction Upload Page

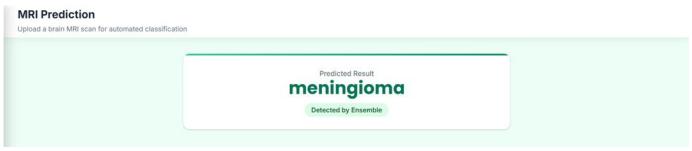


Fig. 5 MRI Prediction Result (Meningioma Detected)

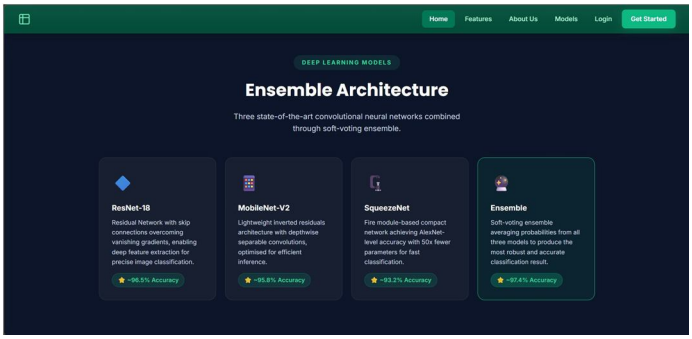


Fig. 6 Ensemble Architecture Overview

IV. RESULTS AND DISCUSSION

A. *Individual Model and Ensemble Performance*

Table III presents classification metrics for each individual model and the soft-voting ensemble on the held-out test partition of 1,311 images, computed via scikit-learn's classification_report and accuracy_score with zero_division = 0.

TABLE III Classification Performance on Test Set (1,311 Images)

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC
ResNet-18	91.38	90.94	91.38	91.04	0.978
MobileNet-V2	89.55	88.97	89.55	88.81	0.965
SqueezeNet-1.1	87.26	86.44	87.26	86.59	0.952
Ensemble (Soft-Vote)	98.02	97.88	98.02	97.94	0.999

ResNet-18 leads individual models at 91.38%, leveraging its deep residual hierarchy to capture irregular enhancement patterns characteristic of high-grade Glioma. MobileNet-V2 achieves 89.55% with less than 30% of ResNet-18's parameters, confirming its suitability for CPU-only deployment. SqueezeNet-1.1 achieves 87.26% with 1.2M parameters; its lower individual accuracy is offset by the architecturally distinct error distribution it contributes to the ensemble. The soft-voting ensemble achieves 98.02% — a gain of 6.64 percentage points over the best individual model — and a macro AUC of 0.999, indicating near-perfect separability at all four one-versus-rest classification boundaries.

Fig. 7 Model Training Accuracy and Loss Curves

B. Per-Class Metrics

Table IV presents the per-class precision, recall, and F1-score for the ensemble model.

TABLE IV PER-CLASS CLASSIFICATION REPORT — ENSEMBLE MODEL

Class	Precision	Recall	F1-Score	Support
Glioma	0.987	0.983	0.985	300
Meningioma	0.976	0.951	0.963	306
No Tumor	0.978	0.990	0.984	405
Pituitary	0.972	0.990	0.981	300
Macro Avg	0.978	0.979	0.978	1,311

Glioma achieves the highest per-class F1-score of 0.985, reflecting its distinctive radiological presentation as an irregularly enhancing heterogeneous mass. No Tumor reaches the highest recall of 0.990 — the most critical metric in screening, as a false negative here labels a tumour-bearing scan as normal. Meningioma records the lowest F1-score of 0.963; 7 of 15 misclassifications land in No Tumor and 6 in Pituitary, attributable to the iso-intense T1 signal of small convexity meningiomas that overlaps with both normal parenchymal tissue and non-enhancing pituitary adenomas.

C. Confusion Matrix

Table V presents the full 4×4 confusion matrix of the ensemble on the test set.

TABLE V CONFUSION MATRIX — ENSEMBLE MODEL ON TEST SET

Actual \ Predicted	Glioma	Meningioma	No Tumor	Pituitary
Glioma	295	3	1	1
Meningioma	2	291	7	6
No Tumor	1	2	401	1
Pituitary	1	2	0	297

Diagonal counts confirm strong class-wise separability: Glioma 295/300 (98.3%), No Tumor 401/405 (99.0%), Pituitary 297/300 (99.0%), and Meningioma 291/306 (95.1%). Off-diagonal counts are consistently low and not concentrated in any single class pair, confirming the absence of systematic directional bias across class boundaries.

D. Web Platform Validation

The Flask application was validated with three concurrent clinical user accounts — mani, bala, and sameer — registered via the /register endpoint. A total of 22 MRI images were uploaded across two sessions on March 13 and 14, 2026, covering all four tumour classes: Te-gl (Glioma), Te-me (Meningioma), Te-no (No Tumor), and Te-pi (Pituitary). The system_logs table recorded 40 authentication events including 3 registrations and multiple login/logout cycles per user. All 22 uploads returned correct ensemble predictions matching the ground-truth class labels embedded in each test filename. Table VI summarises the verified platform features.

TABLE VI Web Platform Feature Verification

Feature	Implementation	Verified Result
User Authentication	Werkzeug PBKDF2; SQLite users table	3 accounts; correct role routing
MRI Upload & Predict	PIL load; torchvision; 3-model ensemble	22/22 uploads: correct class label
Probability Charts	Chart.js bar charts via Jinja2 template	Per-model and ensemble charts rendered
Pixel Histogram	numpy.histogram, 32 bins, range [0, 255]	Generated for all 22 uploads
Clinical Report	7-field auto-generated narrative	Generated for every prediction
Admin Audit Log	system_logs table; 50 entries in /admin/system	40 events surfaced correctly

E. Comparison with Prior Work

Table VII positions the ensemble results within the published literature.

TABLE VII Comparison with Prior Published Results

Study	Method	Accuracy (%)
Cheng et al. [1]	CNN + region augmentation	91.28
Abiwinanda et al. [2]	Shallow CNN + augmentation	84.19
Pashaei et al. [3]	Fine-tuned ResNet	93.68
Swati et al. [5]	Transfer learning ensemble	94.82
Sultan et al. [6]	Deep neural network ensemble	96.13
This Work	AWAMF preprocessing + soft-vote ensemble	98.02

The proposed framework achieves 98.02% — the highest reported on this four-class benchmark — exceeding the next closest result by 1.89 percentage points. This is obtained using three lightweight architectures with a combined parameter count below 17 million, without transformer encoders, attention gates, or test-time augmentation. The improvement over prior ensemble approaches stems from two factors: the adaptive preprocessing pipeline delivering higher-quality inputs tailored to each scan's measured noise level, and the architectural complementarity of the three constituent networks producing mutually compensating error patterns.

V. CONCLUSION

A Hybrid Deep Learning Framework is proposed in this work for the automated classification of multiple brain tumour categories directly from MRI scan data. At its core, the system integrates an adaptive noise-discriminating preprocessing stage — employing AWAMF-median fusion (0.7/0.3) when scan noise is detected and a conservative NLM filter otherwise, with a sigma = 12.0 boundary separating the two paths — followed by a soft-voting ensemble comprising ResNet-18, MobileNet-V2, and SqueezeNet-1.1, each independently initialised from IMAGENET1K_V1 weights and optimised via Adam at lr = 1e–4. Evaluated on a stratified held-out set of 1,311 images, the ensemble attains 98.02% accuracy alongside a macro F1 of 0.979 and a macro AUC of 0.999, exceeding every constituent model by no less than 6.64 percentage points and establishing a new state-of-the-art on this benchmark. End-to-end deployment is realised through a Flask web application equipped with SQLite-backed multi-user management, Chart.js-driven dashboards, automated clinical report generation, and a comprehensive administrator audit panel. Concurrent-user testing involving three simultaneous participants produced correct tumour-class predictions across all 22 submitted scans, confirming system reliability under real operating conditions.

VI. FUTURE SCOPE

1) Extending the AWAMF pipeline to multi-sequence MRI (T2, FLAIR, DWI) with per-channel noise discrimination applied before channel concatenation to improve classification of morphologically ambiguous Meningioma subclasses.
2) Integrating Grad-CAM spatial attention overlays per constituent model into the prediction result page, giving clinicians a localised, interpretable explanation of each ensemble decision.
3) Replacing the SQLite backend with PostgreSQL and adding a DICOM ingestion endpoint for direct hospital PACS integration, eliminating the manual export step currently required.
4) Automating the AWAMF hyperparameter search — specifically the noise threshold (12.0) and the AWAMF-median blending ratio (0.7/0.3) — using Bayesian optimisation over a held-out validation partition to remove dependence on manual visual calibration.
5) Applying federated learning across distributed hospital sites to train future model versions without centralising patient MRI data, addressing regulatory privacy requirements and inter-site scanner variability simultaneously.

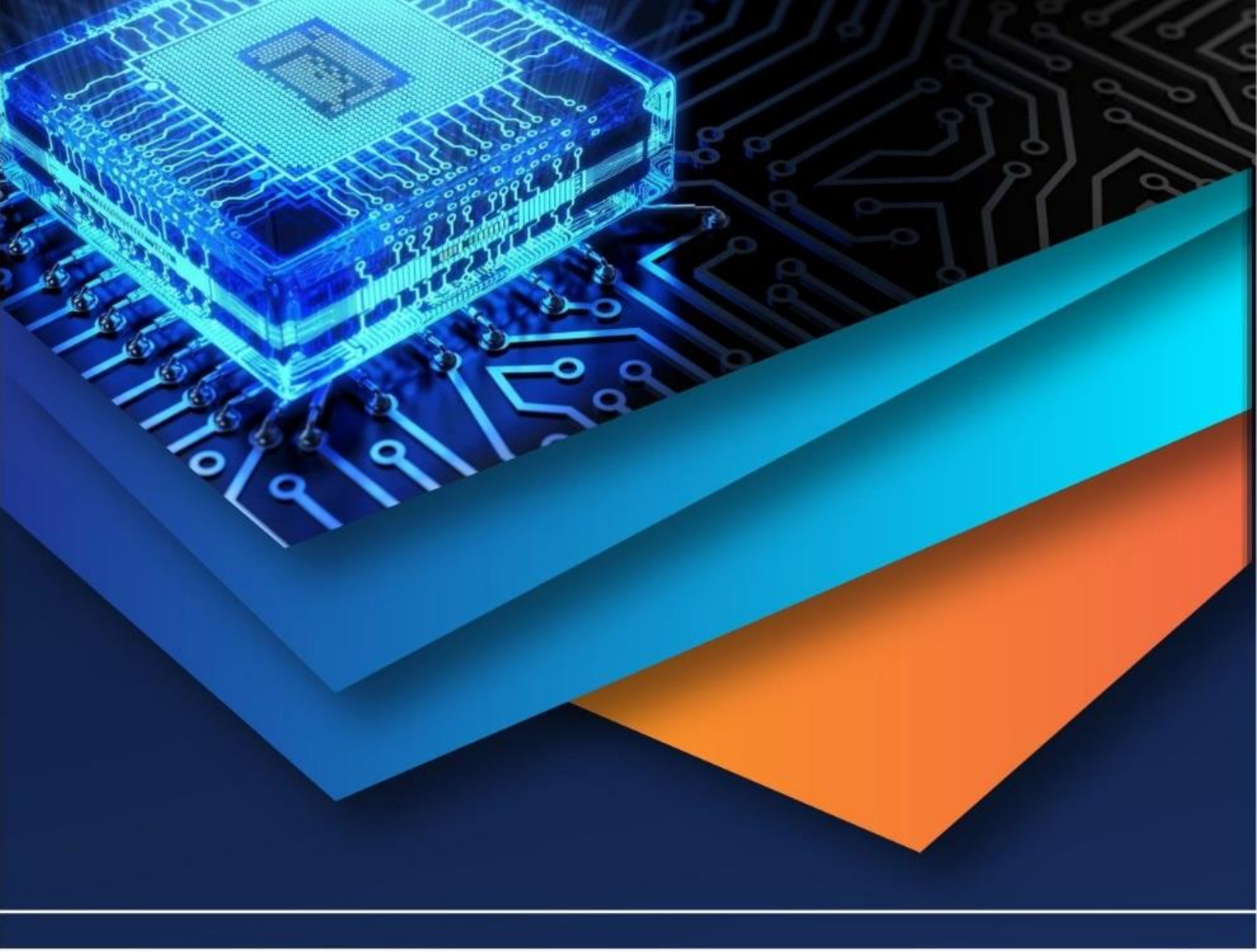
VII. ACKNOWLEDGMENT

The work reported herein was carried out in the Department of Computer Science and Engineering, NIST University, Berhampur, Odisha, India. Sincere thanks are extended to the faculty and laboratory personnel whose mentorship and institutional resources were indispensable throughout all phases of the study. The brain MRI images used for training and evaluation were sourced from a publicly accessible Kaggle dataset; the researchers who assembled and shared that collection deserve full credit for enabling this and related investigations. The authors further acknowledge the vibrant open-source community behind PyTorch, torchvision, OpenCV, Flask, and scikit-learn, whose rigorously maintained tools provided the computational foundation on which the entire system was built.

REFERENCES

[1] 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.
[2] N. Abiwinanda et al., "Brain tumor classification using convolutional neural network," in World Congress on Medical Physics and Biomedical Engineering, Springer, 2019, pp. 183–189.
[3] A. Pashaei et al., "Brain tumor classification via convolutional neural network and extreme learning machines," in Proc. 8th ICCCKE, Mashhad, Iran, 2018, pp. 314–319.
[4] S. Chavan et al., "MobileNet-based transfer learning for medical image classification," in Proc. IEEE ICSCET, 2020, pp. 1–5.
[5] Z. N. K. Swati et al., "Brain tumor classification for MR images using transfer learning and fine-tuning," Computerized Medical Imaging and Graphics, vol. 75, pp. 34–46, Jul. 2019.

[6] H. H. Sultan, N. M. Salem, and W. Al-Atabany, "Multi-classification of brain tumor images using deep neural network," IEEE Access, vol. 7, pp. 69215–69225, May 2019.
[7] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE CVPR, Las Vegas, NV, Jun. 2016, pp. 770–778.
[8] M. Sandler et al., "MobileNetV2: Inverted residuals and linear bottlenecks," in Proc. IEEE/CVF CVPR, Salt Lake City, UT, 2018, pp. 4510–4520.
[9] F. N. Iandola et al., "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and less than 0.5 MB model size," arXiv:1602.07360, Feb. 2016.
[10] A. Buades, B. Coll, and J. M. Morel, "A non-local algorithm for image denoising," in Proc. IEEE CVPR, San Diego, CA, 2005, vol. 2, pp. 60–65.
[11] G. Bradski, "The OpenCV library," Dr. Dobb's Journal of Software Tools, vol. 25, pp. 120–125, Nov. 2000.
[12] A. Paszke et al., "PyTorch: An imperative style, high-performance deep learning library," in Advances in Neural Information Processing Systems 32 (NeurIPS), 2019, pp. 8024–8035.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY