Journal of Medical and Health Studies (JMHS)
ISSN: 2710-1452
DOI: 10.32996/jmhs
Journal Homepage: www.al-kindipublisher.com/index.php/jmhs

JMHS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

CatBoost-Stacked Heterogeneous Deep Ensembles for Explainable Multi-Class Brain Tumor MRI Classification

Partha Chakraborty
School of Business, International American University, 3440 Wilshire Blvd STE 1000, Los Angeles, CA 90010, USA
Corresponding Author: Partha Chakraborty, **E-mail**: parthachk64@gmail.com

| ABSTRACT

Accurate brain tumor screening from MRI requires reliable four-class recognition (glioma, meningioma, pituitary, and no-tumor) under substantial multi-source variability, class imbalance, and the need for clinically interpretable decisions beyond aggregate accuracy. We propose a stacking ensemble that couples heterogeneous feature extractors (EfficientNetB0, MobileNetV2, GoogLeNet, a multi-level CapsuleNet, and a CNN) with a CatBoost meta-learner trained on concatenated class-probability vectors, enabling non-linear fusion of complementary error patterns across backbones. Experiments are conducted on M2, a merged dataset of 24,618 MR images assembled from four public sources, using an 80/5/15 split and stratified 10-fold cross-validation; to prevent evaluation leakage, imbalance mitigation (Borderline-SMOTE and label-preserving augmentation) is applied exclusively to the training folds. The proposed model attains 98.99±0.41% accuracy, 98.32±0.49% micro-F1, 99.33±0.27% PR-AUC, and 94.14±0.13% MCC on M2, consistently outperforming individual backbones, and Grad-CAM audits typically highlight tumor-relevant regions while residual failures concentrate in morphologically similar subtypes. This work contributes a compact, explainable stacking framework with a deployment-oriented inference workflow and visual auditing, validated at scale on heterogeneous MRI sources to support trustworthy brain tumor decision support in resource-constrained settings and real-world clinical practice.

| KEYWORDS

Brain tumor, MRI, stacking ensemble, CatBoost, class imbalance, Grad-CAM, explainable AI.

| ARTICLE INFORMATION

ACCEPTED: 21 February 2026 **PUBLISHED:** 05 March 2026 **DOI:** 10.32996/jmhs.2026.7.5.4

1. Introduction

Brain tumors remain a major cause of neurological morbidity and mortality, where delayed or inaccurate diagnosis can directly affect treatment planning, surgical strategy, and long-term prognosis [1]. Magnetic resonance imaging (MRI) is central to neuro-oncology workflows due to its noninvasive, high-contrast depiction of soft tissue; however, routine clinical screening still depends heavily on expert interpretation, which is time-intensive and susceptible to inter-reader variability, especially in resource-constrained settings. In this work, we address automated multi-class brain tumor screening from MRI as a four-class classification task comprising glioma, meningioma, pituitary tumor, and no tumor [2]. This formulation is clinically relevant because these categories correspond to distinct disease characteristics and treatment pathways, while also reflecting a realistic screening scenario in which normal cases must be reliably separated from tumor subtypes.

Recent advances in artificial intelligence have shown that data-driven learning frameworks can support complex recognition, prediction, and decision-making tasks across a wide spectrum of domains, including plant disease analysis, cybersecurity, citation-network modeling, clinical web systems, and biomedical decision support [10]–[13]. In visual and health-related applications, deep learning has been extended to oral cancer segmentation, prostate MRI analysis, cotton leaf diagnosis, cervical cancer diagnosis, precision wound care, glioblastoma-oriented therapeutic modeling, and cancer immunology, reflecting the broader momentum toward intelligent systems for clinically relevant problems [14]–[22]. Additional progress in leukemia screening, lung cancer detection, pneumonia diagnosis, breast cancer diagnosis, lung cancer classification, retinal disease recognition, microorganism classification, multimodal object recognition, and explainable brain tumor diagnosis further indicates that modern diagnostic pipelines are increasingly expected to deliver not only high predictive accuracy, but also interpretability, efficiency, and real-world usability [24], [27], [29]–[41]. Within this context, automated brain MRI classification using CNN-based models remains an important research direction because it aligns with the wider movement toward robust, explainable, and deployment-aware AI for high-stakes medical imaging.

Despite rapid progress in deep learning for brain tumor classification, several limitations persist in current practice. First, many studies rely on single-backbone architectures that can overfit dataset-specific artifacts and exhibit brittle performance under multi-source heterogeneity, including variations in scanners, acquisition protocols, intensity distributions, and patient populations [3]. Such domain shift is particularly consequential when models are trained on curated datasets yet deployed in heterogeneous clinical environments. Second, class imbalance is common in public MRI repositories and merged datasets; reporting accuracy alone can therefore be misleading, as dominant classes may drive high accuracy while minority classes suffer degraded sensitivity, precisely the failure mode that is unacceptable in screening contexts. Third, while explainability techniques such as Grad-CAM are frequently included, they are often presented as qualitative overlays without systematic error analysis or clinically grounded interpretation, limiting their utility for trust calibration and failure awareness.

To address these challenges, we adopt a stacking ensemble formulation that is principled for heterogeneous medical imaging data. Rather than depending on a single representation, stacking explicitly exploits complementary inductive biases from diverse base learners, e.g., efficient convolutional feature extraction, multi-scale processing, and capsule-based part–whole modeling, so that errors made by one backbone can be corrected by others. We further employ CatBoost as a meta-learner over stacked class-probability vectors. This design is computationally attractive because the meta-level input is low-dimensional and tabular, and it is expressive because CatBoost can model non-linear interactions among base predictions, yielding a more reliable decision rule than fixed-weight averaging when class boundaries are subtle and domain variability is high. The main contributions of this paper are:

1. A CatBoost-stacked heterogeneous deep ensemble for four-class brain tumor MRI classification, using concatenated base-model probability vectors for meta-level fusion.
2. Large-scale validation on a multi-source merged dataset (M2), with training-only imbalance handling (Borderline-SMOTE plus augmentation) and stratified cross-validation to reduce evaluation leakage.
3. Explainability auditing with Grad-CAM, complemented by concise failure-pattern commentary to support clinically meaningful trust and model debugging.

The remainder of the paper is organized as follows: Section II reviews related work; Section III details the proposed methodology; Section IV reports experimental settings and results; Section V presents explainability and error analysis; and Section VI concludes the paper with limitations and future directions.

2. Related Works

Deep convolutional neural networks (CNNs), typically trained via transfer learning, dominate MRI-based brain tumor classification pipelines, leveraging hierarchical feature extraction and large-scale pretraining to achieve strong in-dataset performance. Recent studies and reviews report high accuracies for four-category tumor screening on curated public datasets, often by fine-tuning established backbones and standardizing preprocessing; however, these results can overstate real-world reliability when evaluation is confined to homogeneous sources or single repositories [4]. Capsule networks (CapsNets) have been explored as an alternative that may better preserve part–whole relationships and exhibit improved robustness to certain geometric transformations, which is attractive for tumor morphology that varies across slices and acquisition conditions. Nonetheless, CapsNet variants are typically more sensitive to training design choices and are less commonly validated at scale than CNN baselines [5].

Ensembles improve predictive stability by aggregating diverse learners [6]; classical paradigms such as bagging and boosting primarily induce diversity through resampling or iterative reweighting, respectively. In contrast, stacking (stacked generalization) learns a second-level model over base learners' outputs, enabling non-linear correction of systematic error modes and more expressive fusion than fixed averaging, particularly relevant when different architectures capture complementary cues [7]. Within biomedical applications, stacking has shown consistent gains across heterogeneous feature spaces (including deep representations and radiomics), and CatBoost-based meta-learning is increasingly used due to its strong performance on tabular inputs and its ability to model complex interactions with robust training behavior [8].

Saliency approaches such as Grad-CAM and its variants are widely used to provide post-hoc visual rationales in medical imaging, supporting qualitative sanity checks and clinician-facing communication [9]. However, multiple studies caution that saliency maps may be sensitive to model choice, preprocessing, and perturbations; importantly, highlighted regions do not necessarily imply causal evidence, motivating explanation protocols that include failure inspection and localization-fidelity scrutiny rather than relying on standalone overlays.

The recent literature suggests that brain MRI classification can benefit substantially from moving beyond single-backbone CNNs toward hybrid and ensemble-oriented designs. Transformer-based or hybrid vision models have been explored in prostate MRI, oral cancer, cervical cancer, lung cancer, breast cancer, and retinal disease analysis, while explainable stacking and ensemble formulations have also been investigated for brain tumor diagnosis and related medical classification tasks [14], [16], [18], [22],

[24], [27], [29]–[31], [35], [37], [40]. At the same time, several studies on transfer learning for leukemia, lung cancer, and pneumonia detection show that pretrained deep architectures remain highly relevant in medical imaging when dataset size is limited or class structure is challenging [24], [35], [37]. Beyond medicine, advances in weed recognition, mango and soybean disease analysis, depression and emotion detection, medicinal plant recognition, and multimodal object understanding reinforce the value of heterogeneous feature extraction, efficient transformer design, and post-hoc explainability, all of which are directly relevant to strengthening CNN-based brain MRI classification pipelines [15], [17], [27], [28], [32]–[34], [36], [38], [41].

Few works simultaneously integrate multi-source merged validation, stacking-based meta-learning, and explicit interpretability auditing under a strict training-only imbalance handling protocol suitable for trustworthy multi-class brain tumor MRI screening.

3. Materials and Methods

Brain MRI CNN classification framework can be methodologically grounded in three complementary principles: transfer learning for stable representation initialization, ensemble or stacking strategies for improved robustness, and explainability modules for transparent clinical interpretation. Prior work supports the use of pretrained deep networks in medical imaging [24], [35], [37], the integration of stacking or ensemble learners for difficult disease classification tasks [18], [22], [29], [30], [36], [40], and the incorporation of transformer-enhanced or hybrid feature extractors to capture both local texture and broader contextual structure [14], [16], [17], [27], [28], [31], [33], [34]. In addition, studies on intelligent web-based diagnosis, cybersecurity response systems, citation-aware graph learning, sentiment-focused predictive analytics, classroom activity recognition, and multimodal fusion collectively suggest that practical AI systems should be scalable, computationally efficient, and adaptable to heterogeneous data settings [11]–[13], [23], [25], [26], [32], [38], [39]. Therefore, for brain MRI classification, a CNN-centered methodology augmented with transfer learning, ensemble fusion, and explainable inference is well supported by contemporary trends across both medical and broader AI research. Figure 1 summarizes the proposed end-to-end workflow. Each MRI slice is first standardized through lightweight preprocessing to reduce non-brain background and contrast variability. A set of heterogeneous base learners then produce class-probability vectors for the four target categories. These probability vectors are concatenated into a compact meta-level feature representation and passed to CatBoost, which learns a non-linear stacking rule to generate the final prediction. Finally, Grad-CAM is computed to visually audit whether discriminative evidence is concentrated within tumor-relevant regions and to characterize typical failure patterns.

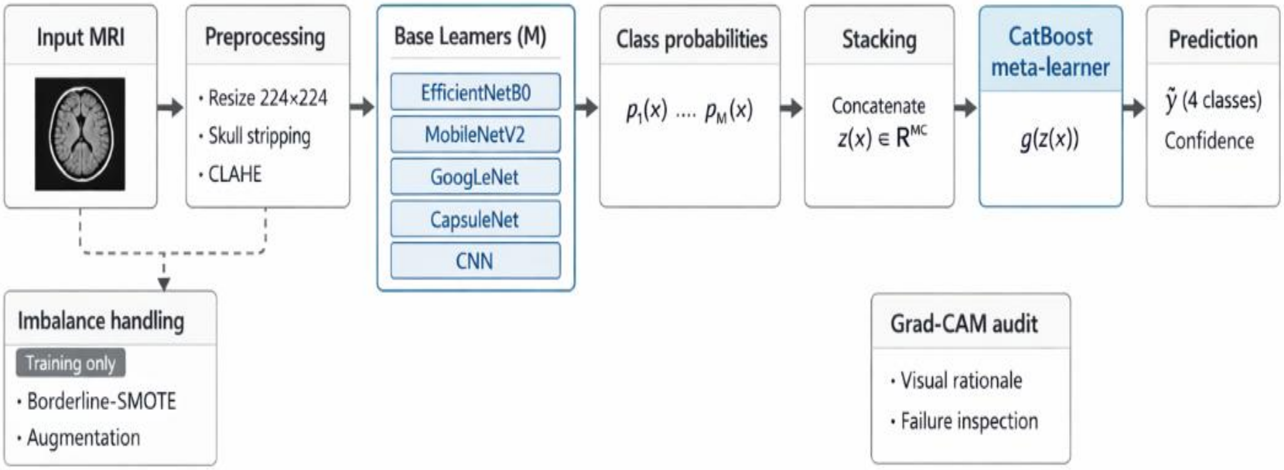

Fig. 1. Overall methodology

3.1 Image Preprocessing

All images are resized to 224×224 to enable consistent batch processing across backbones. Skull stripping is applied to suppress non-brain regions and reduce background-driven shortcuts. Contrast Limited Adaptive Histogram Equalization (CLAHE) is used to improve local contrast consistency across sources. The same preprocessing pipeline is applied to training, validation, and test splits. All preprocessing parameters are fixed a priori (no split-specific tuning) to avoid implicit leakage.

3.2 Training-only imbalance handling

Class imbalance is addressed only within training folds. Borderline-SMOTE is applied after the train/validation/test partitioning to generate synthetic samples for under-represented classes using decision-boundary neighborhood information, while ensuring that neither validation nor test sets are altered. In addition, label-preserving augmentation is used on training data (rotation, horizontal/vertical flip, scaling, center cropping, and brightness adjustment) to improve invariance to common acquisition and pose variations without contaminating the evaluation splits.

3.3 Base learners

We employ a heterogeneous pool of backbones to promote complementary error patterns. EfficientNetB0 contributes balanced compound scaling that often yields strong accuracy with moderate capacity. MobileNetV2 provides an efficient inverted-residual design that improves generalization under constrained computes and helps reduce overfitting. GoogLeNet (Inception) supplies multi-scale receptive fields via inception modules, which is useful for tumors with diverse sizes and texture granularity. A multi-level CapsuleNet introduces part–whole routing behavior that can better preserve spatial relationships compared with pure convolutional pooling. A standard CNN baseline is included as a controlled-capacity learner to diversify representations and to serve as a stable reference within the ensemble.

3.4 Proposed Model

Let $C = 4$ denote the number of classes and $M$the number of base learners. For an input image $\mathbf{x}$, each base model outputs a class-probability vector $\mathbf{p}_m(\mathbf{x})$. We form a meta-level feature vector by concatenating these probabilities and train CatBoost to map the stacked vector to the final decision:

$$p_m(x) \in R^C, \qquad \sum_{c=1}^{C} p_{m,c}(x) = 1, \quad m \in \{1, \dots, M\}$$

$$z(x) = [p_1(x); \dots; p_M(x)] \in R^{MC}$$

$$\hat{y} = \arg\max_c g\big(\boldsymbol{z}(\boldsymbol{x})\big)_c$$

To reduce meta-level leakage, CatBoost is trained on out-of-fold (OOF) probability vectors: base learners are fit on training folds, OOF predictions are collected for the corresponding held-out folds, and the meta-learner is trained on these OOF features. At inference, base learners trained on the full training set produce $\mathrm{z}(\mathrm{x})$for CatBoost prediction.

3.5 Implementation details

All backbones are trained for 30 epochs with batch size 32 using early stopping (patience = 8) to prevent overfitting; Nadam is used as the default optimizer with a step-based learning-rate decay schedule. The stacking stage uses CatBoost trained on OOF probability features with multi-class loss. Hardware details may be reported briefly if space permits (e.g., a single GPU workstation) but are not required for reproducibility in a 6-page format unless runtime comparisons are emphasized.

4. Experimental Setup and Results

4.1 Dataset Construction and Splits

We evaluate the proposed approach on M2, a merged four-class brain MRI dataset created by integrating four public sources (BraTS, Msoud, Br35H, and SARTAJ). M2 contains 24,618 images distributed across Glioma (6,085), Meningioma (6,400), Pituitary (6,452), and No Tumor/Normal (5,681). Data are partitioned using a training/validation/testing = 80%/5%/15% protocol with stratification to preserve class proportions. For robust performance estimation, we additionally report outcomes under 10-fold stratified cross-validation, summarized as mean ± standard deviation across folds.

Table 1. M2 dataset class-wise distribution (80/5/15 split).

Class	Total	Train	Val	Test
Glioma	6085	4868	304	913
Meningioma	6400	5120	320	960
Pituitary	6452	5162	323	967
No Tumor	5681	4544	285	852
All	24618	19694	1232	3692

4.2 Evaluation Metrics

We report Accuracy for comparability, but emphasize imbalance-robust measures: micro-F1, PR-AUC, and Matthews Correlation Coefficient (MCC). Micro-F1 aggregates errors across classes, PR-AUC better reflects precision–recall behavior under skewed distributions, and MCC summarizes all confusion-matrix terms into a single correlation-like score. Results are presented as mean ± std over stratified folds.

4.3 Baselines

We compare the proposed stacked model against strong single-backbone baselines trained under the same preprocessing and split protocol. EfficientNetB0 is used as the primary high-performing single model baseline, while CapsuleNet (CapNet) represents an alternative inductive bias that preserves part–whole spatial relationships. Additional backbones (e.g., MobileNetV2, GoogLeNet, and a standard CNN) are evaluated in the complete study; however, to preserve space in a 6-page format, the results table below highlights the most informative baselines.

4.4 Main Quantitative Results

Table 2 summarizes the principal results on M2. The CatBoost-stacked ensemble achieves the strongest performance, reaching 98.99±0.41% Accuracy, 98.32±0.49% micro-F1, 99.33±0.27% PR-AUC, and 94.14±0.13% MCC. Relative to the best single model (EfficientNetB0), stacking improves Accuracy by +1.26 percentage points, micro-F1 by +1.91, PR-AUC by +2.06, and MCC by +3.04, indicating more reliable discrimination than accuracy gains alone. These improvements are consistent with the role of meta-learning on probability vectors, which can correct systematic confusions from individual backbones and improve stability under multi-source heterogeneity.

Table 2 Performance on M2 (mean ± std)

Model	Accuracy (%)	micro-F1 (%)	PR-AUC (%)	MCC (%)
EfficientNetB0	97.73 ± 0.51	96.41 ± 0.38	97.27 ± 0.52	91.10 ± 0.46
CapNet	96.28 ± 0.43	95.75 ± 0.71	97.85 ± 0.49	91.94 ± 0.60
Proposed (Stack + CatBoost)	98.99 ± 0.41	98.32 ± 0.49	99.33 ± 0.27	94.14 ± 0.13

4.5 Ablation (Single-Factor Study)

To quantify the contribution of training enhancements, we analyze the ensemble under combinations of feature selection (FS) and data augmentation (DA) on M2 in Table 3. Using FS+DA yields 98.94±0.35% Accuracy and 93.34±0.28% MCC, whereas removing both (No FS + No DA) reduces performance to 98.03±0.55% Accuracy and 92.92±0.44% MCC, with increased variability across folds. Using DA alone (No FS + DA) further underperforms FS+DA (97.91±0.48% Accuracy, 92.78±0.61% MCC), suggesting that augmentation without complementary feature prioritization is insufficient for peak robustness.

Table 3 Ablation summary (M2)

Setting	Accuracy (%)	MCC (%)
No FS + No DA	98.03 ± 0.55	92.92 ± 0.44
No FS + DA	97.91 ± 0.48	92.78 ± 0.61
FS + DA	98.94 ± 0.35	93.34 ± 0.28

5. Explainability and Error Analysis

5.1 Grad-CAM audit protocol

To assess whether predictions rely on clinically plausible evidence, we perform a post-hoc explainability audit using Grad-CAM on representative test-set MR images from each class. For each selected case, we compute the class-discriminative activation map at the final convolutional stage of the backbone and overlay the normalized heatmap on the corresponding MRI slice. A “good” explanation is characterized by compact, high-intensity activation aligned with the visible tumor mass (or anatomically relevant region), while a “bad” explanation exhibits diffuse attention, edge-focused artifacts, or activation shifted to non-brain/background areas. We also include at least one misclassified example to inspect whether failure cases arise from ambiguous morphology, low contrast, or evidence concentration outside the lesion region.

5.2 Qualitative results

Fig. 2 presents a multi-panel visualization with one representative case per class, glioma, meningioma, pituitary, and no tumor, and an additional failure example (optional) to highlight a typical misclassification mode. Each panel includes the original MRI slice and its Grad-CAM heatmap overlay using identical scaling and colormap normalization to enable direct visual comparison across classes. Qualitatively, correct predictions typically show localized activation around lesion-bearing regions for tumor classes, whereas normal cases exhibit low-intensity or anatomically diffuse responses without spurious focal hotspots. The failure panel demonstrates a common scenario where Grad-CAM attention is either partially shifted away from the lesion boundary or broadly distributed across structurally salient regions, consistent with confusion between morphologically similar tumor subtypes.

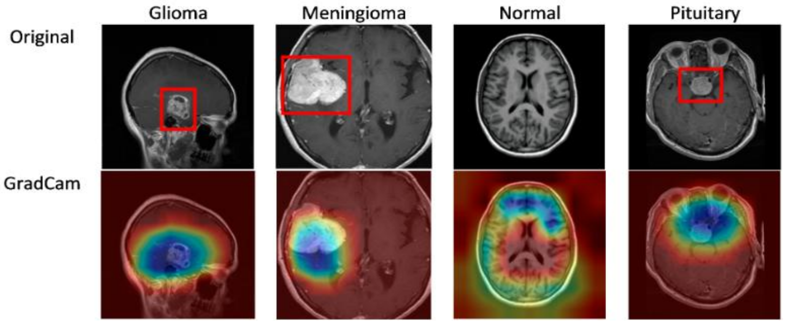

Fig. 2. Grad-CAM visual audit on M2 test images.

5.3 Error patterns

Beyond qualitative evidence, we analyze misclassification tendencies to identify clinically meaningful failure modes. Errors are most frequently observed between benign-appearing tumors (e.g., meningioma or pituitary) and glioma, which can share overlapping intensity profiles or produce similar mass effects depending on slice location and tumor extent. Additional confusions occur when lesion boundaries faint or when tumors occupy small regions relative to surrounding anatomy, leading to partial attention drift and reduced confidence separation at the meta-learner stage. To keep the conference format compact, any confusion matrix is best included as a small insect within Fig. 2, thereby supporting the narrative that the primary remaining errors occur among morphologically similar subclasses rather than between tumor and normal cases.

6. Conclusion

This paper addressed four-class brain tumor screening from MRI (glioma, meningioma, pituitary, and no tumor) using a CatBoost-stacked heterogeneous deep ensemble that fuses complementary base learners through concatenated class-probability vectors. On the large multi-source M2 dataset (24,618 images), the proposed approach achieved 98.99±0.41% accuracy, 98.32±0.49% micro-F1, 99.33±0.27% PR-AUC, and 94.14±0.13% MCC, consistently outperforming strong single-backbone baselines. Grad-CAM audits further indicated that, for most correctly classified cases, discriminative activation concentrates on tumor-relevant regions, while misclassifications often coincide with diffuse or shifted attention under ambiguous morphology. Despite these gains, ensemble inference increases computational cost and multi-source merging cannot fully eliminate residual bias stemming from scanner and protocol variability. Future work will focus on uncertainty-aware decision support, external-site validation on unseen clinical cohorts, and improved calibration to enable reliable confidence estimation for deployment.

Funding: This research received no external funding.

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

References

[1] Chattopadhyay A, Maitra M. MRI-based brain tumour image detection using CNN based deep learning method. Neuroscience Informatics. 2022 Dec 1;2(4):100060. doi:10.1016/j.neuri.2022.100060

[2] Özkaraca O, Bağrıaçık Oİ, Gürüler H, Khan F, Hussain J, Khan J, et al. Multiple Brain Tumor Classification with Dense CNN Architecture Using Brain MRI Images. Life 2023, Vol 13, Page 349. 2023 Jan 28;13(2):349. doi:10.3390/life13020349

[3] Naseer A, Yasir T, Azhar A, Shakeel T, Zafar K. Computer-Aided Brain Tumor Diagnosis: Performance Evaluation of Deep Learner CNN Using Augmented Brain MRI. Int J Biomed Imaging. 2021 Jan 1;2021(1):5513500. doi:10.1155/2021/5513500

[4] Dolz J, Desrosiers C, Wang L, Yuan J, Shen D, Ben Ayed I. Deep CNN ensembles and suggestive annotations for infant brain MRI segmentation. Computerized Medical Imaging and Graphics. 2020 Jan 1;79:101660. doi:10.1016/j.compmedimag.2019.101660 PubMed PMID: 31785402.

[5] Yamanakkanavar N, Lee B. A novel M-SegNet with global attention CNN architecture for automatic segmentation of brain MRI. Comput Biol Med. 2021 Sep 1;136(1):104761. doi:10.1016/j.compbiomed.2021.104761 PubMed PMID: 34426168.

[6] Khagi B, Kwon GR. 3D CNN Design for the Classification of Alzheimer's Disease Using Brain MRI and PET. IEEE Access. 2020;8:217830–47. doi:10.1109/ACCESS.2020.3040486

[7] Tripathy BK, Parikh S, Ajay P, Magapu C. Brain MRI segmentation techniques based on CNN and its variants. Brain Tumor MRI Image Segmentation Using Deep Learning Techniques. 2022 Jan 1;161–83. doi:10.1016/B978-0-323-91171-9.00001-6

[8] Deepak S, Ameer PM. Automated Categorization of Brain Tumor from MRI Using CNN features and SVM. Journal of Ambient Intelligence and Humanized Computing 2020 12:8. 2020 Oct 1;12(8):8357–69. doi:10.1007/s12652-020-02568-w

[9] Khairandish MO, Sharma M, Jain V, Chatterjee JM, Jhanjhi NZ. A Hybrid CNN-SVM Threshold Segmentation Approach for Tumor Detection and Classification of MRI Brain Images. IRBM. 2022 Aug 1;43(4):290–9. doi:10.1016/j.irbm.2021.06.003

[10] Haque R, Khan M, Pranto MN, et al. Data-Centric Approach for Leather Quality Control Using Advanced Vision Transformer Models. Proceedings - International Conference on Next Generation Communication and Information Processing, INCIP 2025. Published online 2025:200-205. doi:10.1109/INCIP64058.2025.11019741

[11] Sultana S, Rahman MM, Hossain MS, Gony MdN, Rafy A. AI-powered threat detection in modern cybersecurity systems: Enhancing real-time response in enterprise environments. World Journal of Advanced Engineering Technology and Sciences. 2022 Aug 30;6(2):136–46. doi:10.30574/wjaets.2022.6.2.0079

[12] Abid SM, Xiaoping Q, Islam MM, Islam MA, Rahman MM, Alam ARM. Edge-Conditioned GAT for Journal Ranking in Citation Networks. 2025 IEEE 6th International Seminar on Artificial Intelligence, Networking and Information Technology, AINIT 2025. 2025;1612–9. doi:10.1109/AINIT65432.2025.11035687

[13] Al Masum A, Limon ZH, Islam MA, Rahman MS, Khan M, Afridi SS, et al. Web Application-Based Enhanced Esophageal Disease Diagnosis in Low-Resource Settings. 2024 IEEE International Conference on Biomedical Engineering, Computer and Information Technology for Health, BECITHCON 2024. 2024;153–8. doi:10.1109/BECITHCON64160.2024.10962580

[14] Hossain A, Sakib A, Pranta ASUK, Debnath J, Tarafder MTR, Islam S, et al. Transformer-Based Ensemble Model for Binary and Multiclass Oral Cancer Segmentation. 2025 International Conference on Electrical, Computer and Communication Engineering, ECCE 2025. 2025. doi:10.1109/ECCE64574.2025.11012921

[15] Rahman H, Khan MA, Khan S, Limon ZH, Siddiqui MIH, Chakroborty SK, et al. Automated Weed Species Classification in Rice Cultivation Using Deep Learning. 2025 International Conference on Electrical, Computer and Communication Engineering, ECCE 2025. 2025. doi:10.1109/ECCE64574.2025.11014047

[16] Debnath J, Bin Mohiuddin A, Pranta ASUK, Sakib A, Hossain A, Shanto MM, et al. Hybrid Vision Transformer Model for Accurate Prostate Cancer Classification in MRI Images. 2025 International Conference on Electrical, Computer and Communication Engineering, ECCE 2025. 2025. doi:10.1109/ECCE64574.2025.11013952

[17] Rahman MM, Hossain MS, Dhakal K, Poudel R, Islam MM, Ahmed MR, et al. A Novel Transformer Model for Accelerated and Efficient Cotton Leaf Disease Identification. 2025 IEEE International Conference on Quantum Photonics, Artificial Intelligence, and Networking, QPAIN 2025. 2025. doi:10.1109/QPAIN66474.2025.11172151

[18] Bin Mohiuddin A, Rahman MM, Gony MN, Shuvra SMK, Rafy A, Ahmed MR, et al. Accelerated and Accurate Cervical Cancer Diagnosis Using a Novel Stacking Ensemble Method with Explainable AI. 2025 IEEE International Conference on Quantum Photonics, Artificial Intelligence, and Networking, QPAIN 2025. 2025. doi:10.1109/QPAIN66474.2025.11171850

[19] Malik AH, Rahman S. Toward precision wound healing: Integrating regenerative therapies and smart technologies. International Journal of Science and Research Archive. 2025 Sep 30;16(3):244–57. doi:10.30574/ijsra.2025.16.3.2492

[20] Malik AH, Rahman S. Hybrid Temozolomide Nanoconjugates: A polymer–drug strategy for enhanced stability and glioblastoma therapy. International Journal of Science and Research Archive. 2025 Sep 30;16(3):258–68. doi:10.30574/ijsra.2025.16.3.2493

[21] Malik AH, Rahman S. Molecular erasers: Reprogramming cancer immunity through protein degradation. World Journal of Advanced Engineering Technology and Sciences. 2025 Sep 30;16(3):277–91. doi:10.30574/wjaets.2025.16.3.1335

[22] Siddiqui MIH, Khan S, Limon ZH, Rahman H, Khan MA, Al Sakib A, et al. Accelerated and accurate cervical cancer diagnosis using a novel stacking ensemble method with explainable AI. Inform Med Unlocked. 2025 Jan 1;56(2):101657. doi:10.1016/j.imu.2025.101657

[23] Haque R, Laskar SH, Khushbu KG, Hasan MJ, Uddin J. Data-Driven Solution to Identify Sentiments from Online Drug Reviews. Computers 2023, Vol 12,. 2023 Apr 21;12(4). doi:10.3390/computers12040087

[24] Haque R, Al Sakib A, Hossain MF, Islam F, Ibne Aziz F, Ahmed MR, et al. Advancing Early Leukemia Diagnostics: A Comprehensive Study Incorporating Image Processing and Transfer Learning. BioMedInformatics 2024, Vol 4, Pages 966-991. 2024 Apr 1;4(2):966–91. doi:10.3390/biomedinformatics4020054

[25] Haque R, Islam N, Islam M, Ahsan MM. A Comparative Analysis on Suicidal Ideation Detection Using NLP, Machine, and Deep Learning. Technologies 2022, Vol 10,. 2022 Apr 29;10(3). doi:10.3390/technologies10030057

[26] Haque R, Islam N, Tasneem M, Das AK. Multi-class sentiment classification on Bengali social media comments using machine learning. International Journal of Cognitive Computing in Engineering. 2023 Jun 1;4:21–35. doi:10.1016/j.ijcce.2023.01.001

[27] Noman A Al, Hossain A, Sakib A, Debnath J, Fardin H, Sakib A Al, et al. ViX-MangoEFormer: An Enhanced Vision Transformer–EfficientFormer and Stacking Ensemble Approach for Mango Leaf Disease Recognition with Explainable Artificial Intelligence. Computers 2025, Vol 14,. 2025 May 2;14(5). doi:10.3390/computers14050171

[28] Pranta ASUK, Fardin H, Debnath J, Hossain A, Sakib AH, Ahmed MR, et al. A Novel MaxViT Model for Accelerated and Precise Soybean Leaf and Seed Disease Identification. Computers 2025, Vol 14,. 2025 May 18;14(5). doi:10.3390/computers14050197

[29] Haque R, Khan MA, Rahman H, Khan S, Siddiqui MIH, Limon ZH, et al. Explainable deep stacking ensemble model for accurate and transparent brain tumor diagnosis. Comput Biol Med. 2025 Jun 1;191:110166. doi:10.1016/j.compbiomed.2025.110166 PubMed PMID: 40249992.

[30] Ahmed MR, Rahman H, Limon ZH, Siddiqui MIH, Khan MA, Pranta ASUK, et al. Hierarchical Swin Transformer Ensemble with Explainable AI for Robust and Decentralized Breast Cancer Diagnosis. Bioengineering 2025, Vol 12,. 2025 Jun 13;12(6). doi:10.3390/bioengineering12060651

[31] Debnath J, Uddin Khondakar Pranta AS, Hossain A, Sakib A, Rahman H, Haque R, et al. LMVT: A hybrid vision transformer with attention mechanisms for efficient and explainable lung cancer diagnosis. Inform Med Unlocked. 2025 Jan 1;57(1):101669. doi:10.1016/j.imu.2025.101669

[32] Ahmed MR, Haque R, Rahman SMA, Reza AW, Siddique N, Wang H. Vision-audio multimodal object recognition using hybrid and tensor fusion techniques. Information Fusion. 2026 Feb 1;126(1):103667. doi:10.1016/j.inffus.2025.103667

[33] Rahman Swapno SMM, Sakib A, Uddin Khondakar Pranta AS, Hossain A, Debnath J, Al Noman A, et al. Explainable transformer framework for fast cotton leaf diagnostics and fabric defect detection. iScience. 2026 Feb 20;29(2):114411. doi:10.1016/j.isci.2025.114411

[34] Islam S, Haque R, Khan MA, Mohiuddin A Bin, Hossain Siddiqui MI, Limon ZH, et al. Ensemble Transformer with Post-hoc Explanations for Depression Emotion and Severity Detection. iScience. 2026 Feb 20;29(2):114605. doi:10.1016/j.isci.2025.114605

[35] Haque R, Sultana S, Rafy A, Babul Islam M, Arafat MA, Bhattacharya P, et al. A Transfer Learning-Based Computer-Aided Lung Cancer Detection System in Smart Healthcare. IET Conference Proceedings. 2024;2024(37):594–601. doi:10.1049/icp.2025.0858

[36] Khan S, Rahman H, Hossain Siddiqui MI, Hossain Limon Z, Khan MA, Haque R, et al. Ensemble-Based Explainable Approach for Rare Medicinal Plant Recognition and Conservation. 2025 10th International Conference on Information and Network Technologies, ICINT 2025. 2025;88–93. doi:10.1109/ICINT65528.2025.11030872

[37] Haque R, Mamun MA Al, Ratul MH, Aziz A, Mittra T. A Machine Learning Based Approach to Analyze Food Reviews from Bengali Text. 12th International Conference on Electrical and Computer Engineering, ICECE 2022. Published online 2022:80-83. doi:10.1109/ICECE57408.2022.10088971

[38] Nobel SMN, Swapno SMMR, Islam MR, Safran M, Alfarhood S, Mridha MF. A machine learning approach for vocal fold segmentation and disorder classification based on ensemble method. Scientific Reports 2024 14:1. 2024;14(1):1-25. doi:10.1038/s41598-024-64987-5

[39] Khushubu KG, Masum A Al, Rahman MH, et al. TransUNetB: An advanced Transformer–UNet framework for efficient and explainable brain tumor segmentation. Inform Med Unlocked. 2025;59(10):101706. doi:10.1016/j.imu.2025.101706

[40] Al Masum A, Limon ZH, Islam MA, et al. Web Application-Based Enhanced Esophageal Disease Diagnosis in Low-Resource Settings. 2024 IEEE International Conference on Biomedical Engineering, Computer and Information Technology for Health, BECITHCON 2024. Published online 2024:153-158. doi:10.1109/BECITHCON64160.2024.10962580

[41] Rahman S, Parameshachari BD, Haque R, Masfequier Rahman Swapno SM, Babul Islam M, Nobel SN. Deep Learning-Based Left Ventricular Ejection Fraction Estimation from Echocardiographic Videos. 2023 International Conference on Evolutionary Algorithms and Soft Computing Techniques, EASCT 2023. Published online 2023. doi:10.1109/EASCT59475.2023.10392607