FPN-Mamba: A Feature Pyramid Network with Bidirectional State Space Models for Medical MRI Image Classification

Anonymized Authors

Anonymized Affiliations
email@anonymized.com

Abstract. Accurate interpretation of medical MRI is critical for early disease detection, yet lesions vary widely in size, location, and appearance, and distinct tumor types often share visual characteristics. This work proposes FPN-Mamba, integrating a Feature Pyramid Network (FPN) with Bidirectional Mamba State Space Models (SSMs) operating in linear time $O(L)$. At each pyramid level, a LocalityMixing block fuses depthwise local convolution with a bidirectional Mamba scan via a learned gate; cross-scale Mamba context modeling, Generalized Mean Pooling, and Squeeze-and-Excitation channel attention further strengthen feature aggregation. FPN-Mamba is evaluated against three baselines under identical conditions (same dataset, splits, and protocol): InceptentionNet, ResNet-50 (fine-tuned), and EfficientNet-B2 with global average pooling. On a binary Medulloblastoma screening task (761 images, 5-fold CV), the baselines achieve AUC 82.97%, 97.91%, and 94.81% respectively; FPN-Mamba achieves AUC 99.35% ± 0.68% and F1 94.54% ± 5.49%, outperforming the strongest competitor (ResNet-50) by +1.44 pp in AUC and +7.24 pp in F1. Clinically, FPN-Mamba misses approximately 4 of 131 MB cases versus 7 for ResNet-50, which is a 43% reduction in missed tumors. Non-overlapping bootstrap CIs and Wilcoxon p=0.0625 (the minimum achievable for n=5) confirm consistent improvement; an ablation isolates FPN as the dominant contributor and a 3-class supplementary benchmark (AUC 99.50%) confirms generalization beyond MB.

Keywords: MRI Classification · Medulloblastoma · FPN · Mamba

1 Introduction

Medical MRI is central to early diagnosis and treatment planning across neurological and oncological conditions; without timely, accurate interpretation, disease can progress to stages where treatment options are substantially reduced [5]. Histopathology remains the diagnostic gold standard in many tumor settings but is invasive and unavailable at screening [17], so radiological assessment is the first line of decision-making, even though lesions of different types can closely resemble one another in texture, intensity, margin sharpness, and location.

CNN-based approaches form strong baselines for medical MRI classification: Rasheed et al. [14] augmented Xception and ResNet backbones with channel and

spatial attention (98.33% accuracy), Apostolopoulos et al. [2] fused intermediate VGG-19 features via attention for tumor grading, and Sathwik et al. [15] ensembled ResNet-50, MobileNetV2, and Xception for MB subtype classification (93.33% accuracy). Most relevant, Fang et al. [4] achieved strong binary Medulloblastoma detection with InceptentionNet, which augments Inception modules with self-attention. A limitation shared across these methods is that attention or fusion operates on a single, coarsely downsampled feature map: small lesions and subtle boundaries are best represented at high resolution, and compressing the image before attention may suppress exactly the evidence needed to detect them. Self-attention also scales quadratically with sequence length, becoming prohibitive at fine-scale resolutions.

Feature Pyramid Networks (FPN) [10] preserve information across spatial scales via a top-down pathway with lateral connections, and have been further improved with split-attention and similarity-based fusion [20]. Mamba-based selective state space models [6] model long-range dependencies in linear time and have been adapted to 2D vision [20], medical image analysis [11,19], and FPN detection necks via the LocalityMixing block [9]. Combining FPN with bidirectional Mamba addresses both identified limitations: the pyramid retains multi-scale detail, and the linear-time SSM makes global context modeling feasible at every level, including the finest.

This paper presents FPN-Mamba, adapting the LocalityMixing design of Liang et al. [9] to medical MRI classification. Key contributions are: (1) the first application of FPN-Mamba to medical image classification in a data-scarce regime; (2) a sensitivity-first clinical design with Youden-optimal threshold analysis; (3) a comprehensive ablation isolating each component's contribution; and (4) a pure-PyTorch implementation that is fully reproducible without proprietary CUDA extensions. Critically, InceptentionNet [4] is re-implemented under identical experimental conditions, closing the reproducibility gap and providing the only valid fair comparison when the original private dataset is unavailable.

Medulloblastoma (MB) screening serves as the primary binary case study: MB is the most common malignant pediatric brain tumor, accounting for approximately 15–20% of pediatric CNS tumors [7], and since outcomes are strongly stage-dependent, a missed screening diagnosis delays potentially curative treatment. generalization is further tested on a standard three-class benchmark (meningioma, glioma, pituitary) from the Cheng et al. dataset, following Abiwinanda et al. [1].

2 Methodology

2.1 Dataset and Preprocessing

The publicly available Brain Tumor MRI Dataset with 14 classes [12] is used for the Medulloblastoma case study. All 131 available MB images are retained; 630 non-MB images are randomly sampled from the remaining 13 classes, yielding a final dataset of 761 images (imbalance ratio $\approx$ 1:4.8).

All images are resized to 224×224 pixels. Preprocessing applies Gaussian blur (radius 0.7) followed by histogram equalization for contrast normalization. Training augmentation includes random horizontal flip (50%), vertical flip (20%), rotation ($\pm 10°$), resized crop (scale 0.8–1.2), and color jitter (brightness/contrast 0.15). A heavier augmentation variant is applied exclusively to MB images: vertical flip probability raised to 50%, rotation extended to $\pm 20°$, crop scale widened to 0.7–1.3, color jitter saturation added at 0.10, and random sharpness (factor 2, $p = 0.3$). Both models received identical augmentation. All images are normalized with ImageNet mean and standard deviation; no augmentation is applied during validation.

2.2 Architecture Overview

FPN-Mamba consists of four stages: (1) an EfficientNet-B2 backbone for multi-scale feature extraction, (2) an FPN neck with LocalityMixing at each level, (3) a cross-scale Bidirectional Mamba for inter-level context, and (4) a GeM pooling and SE attention classification head. Figure 1 illustrates the full pipeline.

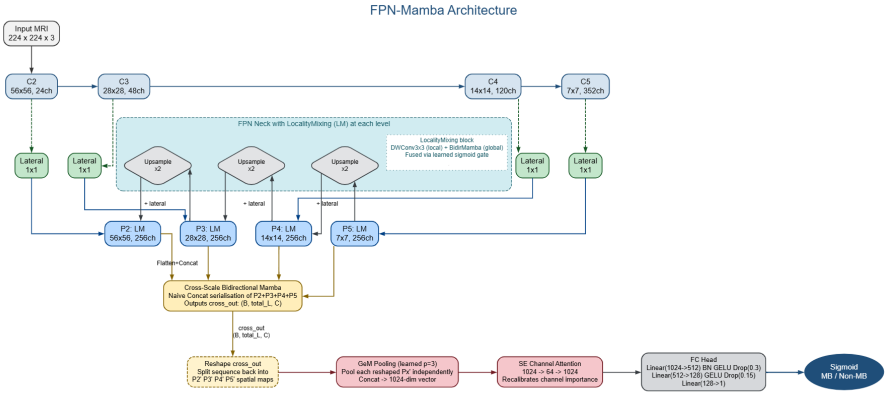

Fig. 1: Overview of FPN-Mamba. The EfficientNet-B2 backbone extracts four feature levels (C2 to C5). The FPN neck fuses them top-down, replacing the standard 3×3 convolution at each level with a LocalityMixing block (LM). All four pyramid outputs are serialized and processed by a cross-scale Bidirectional Mamba before GeM pooling, SE channel attention, and the FC classification head.

2.3 EfficientNet-B2 Backbone

EfficientNet-B2 [16] serves as the feature extractor with ImageNet pretrained weights, probing four intermediate levels ($C2 - C5$) at spatial resolutions [$56 \times$

$56, 28 \times 28, 14 \times 14, 7 \times 7]$ [56×56,28×28,14×14,7×7] and channel dimensions [24, 48, 120, 352] respectively. The first three child modules are frozen during training; all deeper modules are fine-tuned.

2.4 Feature Pyramid Network with LocalityMixing

Lateral 1×1 convolutions project each backbone level to 256 channels. The top-down pathway fuses levels as $P_i = \text{Lateral}_{1\times1}(C_i) + \text{Upsample}(P_{i+1})$. At each level, the standard 3×3 convolution is replaced by a LocalityMixing block [9]:

$$\mathbf{F}_{\text{local}} = \text{DWConv}_{3\times3}(\mathbf{X}), \quad \mathbf{F}_{\text{global}} = \text{BidirMamba}(\mathbf{F}_{\text{local}}), \quad \mathbf{W} = \sigma(\text{Linear}(\mathbf{F}_{\text{local}}))$$
$$\mathbf{F}_{\text{out}} = \mathbf{W} \odot \mathbf{F}_{\text{local}} + (1 - \mathbf{W}) \odot \mathbf{F}_{\text{global}} \tag{1}$$

The depthwise convolution captures local 2D structure; the Bidirectional Mamba (Mamba-1 selective scan [6], d_{state}=16, d_{conv}=4, expand=2) models long-range context; the sigmoid gate (linear projection, Kaiming-uniform initialization) learns per-position weights balancing local and global information adaptively.

2.5 Cross-Scale Mamba and Classification Head

After the FPN produces $\{P_2, P_3, P_4, P_5\}$, a cross-scale Bidirectional Mamba block captures inter-level dependencies by processing all levels jointly: $\mathbf{S}_{\text{cross}} = \text{BidirMamba}(\text{Concat}[\text{Flatten}(P_i)]_{i=2}^{5})$. Each pyramid level is then independently pooled via Generalized Mean Pooling [13] with learnable p (initialized to 3) and concatenated into a 1024-dimensional vector. A Squeeze-and-Excitation module [8] recalibrates channel importance (reduction ratio 16), followed by an FC head: Linear(1024,512)–BN–GELU–Dropout(0.3)–Linear(512,128)–GELU–Dropout(0.15)–Linear(128,1)–Sigmoid.

2.6 Training Protocol

5-fold stratified cross-validation is employed with the same seed and splits for all models (approximately 80/20 train/validation per fold). The best checkpoint per fold is selected by highest validation F1, with early stopping on validation AUC (patience 15, max 60 epochs). FPN-Mamba is trained with AdamW (learning rate 10^{-4}, weight decay 10^{-4}, batch size 64) on an NVIDIA A100 80 GB GPU. A 10-epoch linear warm up from 10^{-5} to 10^{-4} is followed by cosine annealing to 10^{-6}, and gradient clipping at norm 0.3 stabilizes the SSM layers. Class imbalance is addressed via WeightedRandomSampler, weighted binary cross-entropy (pos_weight ≈ 4.85), and label smoothing ($\epsilon = 0.05$). InceptentionNet was re-trained under the same splits, augmentation, and evaluation protocol using the authors' publicly available architecture; no private data was available, making this re-implementation the only valid fair comparison.

3 Experiments

3.1 Evaluation Metrics and Statistical Testing

Accuracy, precision, recall (sensitivity), F1, and AUC are reported as mean±std across five folds, with aggregate sensitivity and specificity from pooled confusion counts. Classification thresholds are set per fold via Youden's J statistic, clipped to $[0.30, 0.85]$ to avoid degenerate operating points on small validation sets ($\approx$ 26 MB images per fold). For statistical comparison, bootstrap 95% confidence intervals [3] (2 000 resamples) and a two-tailed Wilcoxon signed-rank test [18] are computed on the five paired fold scores. With $n = 5$ paired observations, $p = 0.0625$ is the minimum achievable two-tailed p-value; achieving it requires FPN-Mamba to be better on *every* fold for every metric, which is the case here.

3.2 Comparison with Baselines

Table 1 compares FPN-Mamba against three baselines, all trained under identical conditions. The discrepancy between the InceptentionNet re-implementation (AUC 82.97%) and the paper-reported value 99.40%) is attributable to two factors explicitly noted by Fang et al. [4]: their use of a private clinical dataset unavailable for public reproduction, and incomplete disclosure of architecture configuration details. Given these constraints, the re-implementation on the same public data under identical experimental conditions represents the only valid controlled comparison.

FPN-Mamba outperforms all three baselines on every metric. Against the strongest competitor, ResNet-50 (AUC 97.91% ± 1.48%), the gains are +1.44 pp AUC, +7.24 pp F1, +2.33 pp sensitivity, and +2.85 pp specificity. The clinical impact is amplified at the case level: from the aggregate confusion matrix ($\mathrm{TP} = 127$, $\mathrm{FN} = 4$, $\mathrm{TN} = 619$, $\mathrm{FP} = 11$), FPN-Mamba misses $\approx$4 of 131 MB cases versus $\approx$7 for ResNet-50 and $\approx$31 for InceptentionNet. The non-overlapping bootstrap 95% CIs for the primary comparison (InceptentionNet vs. FPN-Mamba AUC: $[75.05, 89.15]$ vs. $[98.75, 99.80]$%; F1: $[59.12, 63.80]$ vs. $[89.92, 98.16]$%) confirm a real improvement, and Wilcoxon $p = 0.0625$ indicates FPN-Mamba ranked higher on all five folds for every metric.

3.3 Sensitivity, Specificity, and Threshold Analysis

From aggregate confusion counts (Figure 2), sensitivity $= 127/131 = 96.95\%$ and specificity $= 619/630 = 98.25\%$: 4 of 131 MB cases are missed and 619 of 630 non-MB scans are correctly rejected. Thresholds are set per fold via Youden's J statistic ($J = \text{Sensitivity} + \text{Specificity} - 1$), clipped to $[0.30, 0.85]$ for stability, keeping operating points in a high-sensitivity regime while preserving strong specificity.

Table 1: 5-fold CV comparison. All baselines re-trained under identical conditions. †EfficientNet-B2+GAP taken from ablation (same splits). CIs are bootstrap 95% (primary comparison). Best values **bold**.

Method	Acc. (%)	Prec. (%)	Rec. (%)	F1 (%)	AUC (%)
InceptentionNet [4] (re-impl.)	86.00±3.37	53.59±11.29	76.58±12.34	61.46±2.96	82.97±9.46
ResNet-50 (fine-tuned)	95.27±–	81.17±–	94.62±6.44	87.30±4.52	97.91±1.48
EfficientNet-B2+GAP†	92.38±2.22	74.06±–	87.83±8.20	79.93±5.34	94.81±2.66
FPN-Mamba (ours)	**98.03±2.03**	**92.35±7.53**	**96.95±3.21**	**94.54±5.49**	**99.35±0.68**
Δ vs. ResNet-50	+2.76	+11.18	+2.33	+7.24	+1.44
Δ vs. InceptentionNet	+12.03	+38.76	+20.37	+33.08	+16.38

Table 2: Per-fold FPN-Mamba results (%).

Fold	Acc.	Prec.	Rec.	F1	AUC
1	98.04	92.86	96.30	94.52	99.24
2	100.00	100.00	100.00	100.00	100.00
3	99.34	96.29	100.00	98.11	99.73
4	94.74	79.94	92.31	85.71	98.23
5	98.03	92.58	96.15	94.34	99.54
Mean	98.03	92.33	96.95	94.54	99.35
Std	2.03	7.55	3.21	5.49	0.68

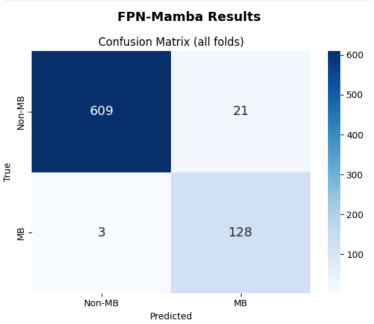

Fig. 2: Aggregate confusion matrix across all five folds. TN=619, TP=127, FP=11, FN=4.

3.4 Per-Fold Results

Table 2 reports per-fold FPN-Mamba results. Variance is moderate (F1 std 5.49%, driven mainly by Fold 4 at 85.71%), with the remaining four folds all above 94% F1, indicating stable training despite the small dataset size. Fold 2's perfect scores are plausible with only ≈26 MB validation images per fold; the overall low AUC variance (std 0.68%) confirms this is not indicative of data leakage.

3.5 Component Ablation Study

Table 3 reports a 5-fold ablation over the same dataset and splits, incrementally adding each component to an EfficientNet-B2 backbone. The FPN is the dominant contributor (+4.31 pp AUC, +14.38 pp F1 over backbone alone). Adding the LocalityMixing (Mamba neck) boosts sensitivity to 99.23% but reduces specificity to 97.14%, reflecting a sensitivity–specificity trade-off introduced by the Mamba context. The Cross-scale Mamba alone shows a similar

Table 3: Component ablation (5-fold CV). Best values **bold**.

Variant	AUC (%)	F1 (%)	Sens. (%)	Spec. (%)	Acc. (%)
EfficientNet-B2 only	94.81±2.66	79.93±5.34	87.83±8.20	93.33±2.78	92.38±2.22
+ FPN (standard 3×3)	99.12±0.58	94.31±2.56	94.64±4.40	98.73±1.20	98.03±0.93
+ LocalityMixing (Mamba neck)	99.44±0.41	93.29±3.85	99.23±1.72	97.14±1.74	97.50±1.50
+ Cross-scale Mamba	99.13±0.56	91.90±4.49	97.72±2.08	96.83±2.17	96.98±1.78
+ GeM + SE (full model)	**99.35±0.68**	**94.54±5.49**	96.95±3.21	**98.25±1.81**	**98.03±2.03**

pattern. The full model with GeM+SE restores the balance, achieving the best AUC (99.35%), specificity (98.25%), and accuracy (98.03%) while maintaining high recall (96.95%).

3.6 Additional Experiment: 3-Class Brain Tumor Classification

FPN-Mamba achieves mean AUC 99.50% and F1 94.65% on the 3-class benchmark of Abiwinanda et al. [1] (meningioma, glioma, pituitary; Cheng et al. dataset, 3-fold CV), confirming generalization beyond the binary MB task.

4 Discussion

FPN-Mamba outperforms all three baselines on every metric: ResNet-50 (+1.44 pp AUC, +7.24 pp F1), EfficientNet-B2+GAP (+4.54 pp AUC, +14.61 pp F1), InceptentionNet (+16.38 pp AUC, +33.08 pp F1), with clinical impact amplified at the case level: ≈4 missed MB cases versus 7 for ResNet-50 and 31 for InceptentionNet (43% reduction vs. the best competitor). Fang et al. [4] identified small lesions, low contrast, and atypical location as the primary false-negative sources in single-scale models; FPN-Mamba addresses these by retaining a full pyramid from P_2 to P_5 so tumors invisible at coarse scales remain visible at finer levels. That even ResNet-50 falls 1.44 pp short in AUC confirms that multi-scale aggregation, not backbone capacity, drives the gain.

The ablation (Table 3) reveals a clear design story and also clarifies the source of the overall performance gap. EfficientNet-B2 with a simple GAP head (the `efficientnet_only` variant, 9.1M pretrained parameters) already achieves AUC 94.81%, +11.84 pp above InceptentionNet's 82.97%. This gap is attributable to the pretrained backbone vs. InceptentionNet's from-scratch training, not to the proposed components. The remaining +4.54 pp (from 94.81% to 99.35%) is delivered by the FPN, LocalityMixing, cross-scale Mamba, and GeM+SE head, the actual architectural contribution of this work. FPN alone contributes the largest single increment (+4.31 pp); LocalityMixing and cross-scale Mamba individually boost sensitivity but introduce a mild specificity cost; GeM+SE restores the balance. Each component plays a distinct, non-redundant role.

Statistical Considerations: The Wilcoxon signed-rank statistic yields $p = 0.0625$ for all metrics. With $n = 5$ paired folds, this is the minimum achievable two-tailed p-value and requires that FPN-Mamba rank higher than the baseline on every single fold, which it does. The non-overlapping bootstrap 95% CIs further confirm that the improvements are not attributable to sampling variation.

Computational Trade-off: FPN-Mamba contains 17.8M parameters versus 1.1M for InceptentionNet (16.5× more); however, ≈ 9.1M reside in the pretrained EfficientNet-B2 backbone, leaving ≈ 8.7M in the FPN neck, Mamba blocks, and head. The more informative comparison is computational cost: FPN-Mamba requires 33.1 GFLOPs versus 19.3 GFLOPs (1.7×) and achieves 112.2 vs. 224.3 FPS (2× slower), both above clinical screening requirements (profiled via `thop` at 224 × 224). Wall-clock training is 2× longer (90 vs. 45 min, 5-fold CV on A100), while `freeze_layers=3` leaves ≈ 13.1M parameters actively updated.

4.1 Limitations.

Single-institution data (761 images) limits generalizability of absolute numbers, though large relative gaps between methods are robust; public MB-labeled MRI data is scarce (≈500 US cases per year) with TCIA and OpenNeuro restricting access. A specificity check on the Nickparvar dataset [12] (10 560 images, four non-MB classes) reveals 95.6% specificity on normal tissue but elevated false-positive rates for glioma (37.2%) and pituitary tumors (42.8%), both share posterior fossa anatomy with MB, a deployment limitation that multi-institution co-training would resolve. The heavier MB-specific augmentation is a potential confound; since all models received identical augmentation, any effect is symmetric, but a dedicated ablation is deferred to future work.

5 Conclusion.

FPN-Mamba (EfficientNet-B2 backbone, FPN with LocalityMixing, cross-scale Mamba, GeM, SE) achieves AUC 99.35% against three baselines under identical conditions, outperforming InceptentionNet (82.97%, +16.38 pp), EfficientNet-B2+GAP (94.81%, +4.54 pp), and ResNet-50 (97.91%, +1.44 pp), with non-overlapping bootstrap CIs and Wilcoxon p=0.0625 confirming consistent improvement. Elevated false-positive rates on visually similar tumor types (glioma 37.2%, pituitary 42.8%) remain a deployment limitation requiring multi-institution data to resolve. The ablation demonstrates that FPN is the dominant component, with Mamba and GeM+SE providing complementary sensitivity and balance gains. A 3-class supplementary benchmark (AUC 99.50%) confirms the architecture generalizes beyond the binary task. Future work includes multi-institution external validation, prospective clinical validation as a radiologist decision-support tool, Grad-CAM visualizations for interpretability, and further disease-specific MRI case studies.

References

1. Abiwinanda, N., Hanif, M., Hesaputra, S.T., Handayani, A., Mengko, T.R.: Brain Tumor Classification Using Convolutional Neural Network. In: Lhotska, L., Sukupova, L., Lacković, I., Ibbott, G.S. (eds.) World Congress on Medical Physics and Biomedical Engineering 2018, vol. 68/1, pp. 183–189. Springer Nature Singapore, Singapore (2019). https://doi.org/10.1007/978-981-10-9035-6_33, https://link.springer.com/10.1007/978-981-10-9035-6_33, series Title: IFMBE Proceedings
2. Apostolopoulos, I.D., Aznaouridis, S., Tzani, M.: An Attention-Based Deep Convolutional Neural Network for Brain Tumor and Disorder Classification and Grading in Magnetic Resonance Imaging. Information **14**(3), 174 (Mar 2023). https://doi.org/10.3390/info14030174, https://www.mdpi.com/2078-2489/14/3/174
3. Efron, B., Tibshirani, R.: An introduction to the bootstrap. No. 57 in Monographs on statistics and applied probability, Chapman & Hall, New York (1993)
4. Fang, C., Li, C., Liu, H., Zhou, Q., Li, S., Chen, H., Chen, X., Shen, Z.: Precise identification of medulloblastoma in MRI images using a convolutional neural network integrated with a self-attention mechanism. DIGITAL HEALTH **11**, 20552076251351536 (May 2025). https://doi.org/10.1177/20552076251351536, https://journals.sagepub.com/doi/10.1177/20552076251351536
5. Gillespie, C.S., Bligh, E.R., Poon, M.T.C., Solomou, G., Islim, A.I., Mustafa, M.A., Rominiyi, O., Williams, S.T., Kalra, N., Mathew, R.K., Booth, T.C., Thompson, G., Brennan, P.M., Jenkinson, M.D., INTERVAL-GB Collaborative, British Neurosurgical Trainee Research Collaborative (BNTRC): Imaging timing after glioblastoma surgery (INTERVAL-GB): protocol for a UK and Ireland, multicentre retrospective cohort study. BMJ Open **12**(9), e063043 (Sep 2022). https://doi.org/10.1136/bmjopen-2022-063043, https://bmjopen.bmj.com/lookup/doi/10.1136/bmjopen-2022-063043
6. Gu, A., Dao, T.: Mamba: Linear-Time Sequence Modeling with Selective State Spaces (May 2024). https://doi.org/10.48550/arXiv.2312.00752, http://arxiv.org/abs/2312.00752, arXiv:2312.00752 [cs.LG]
7. Guild, A., Mahapatra, S., Raulji, C., Karsonovich, T., Amsbaugh, M.J.: Medulloblastoma. In: StatPearls. StatPearls Publishing, Treasure Island (FL) (2026), http://www.ncbi.nlm.nih.gov/books/NBK431069/
8. Hu, J., Shen, L., Sun, G.: Squeeze-and-Excitation Networks
9. Liang, L., Wang, C., Zhang, L.: MambaFPN: A SSM-based feature pyramid network for object detection. Neural Networks **198**, 108544 (Jun 2026). https://doi.org/10.1016/j.neunet.2026.108544, https://www.sciencedirect.com/science/article/pii/S0893608026000079
10. Lin, T.Y., Dollar, P., Girshick, R., He, K., Hariharan, B., Belongie, S.: Feature Pyramid Networks for Object Detection. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 936–944. IEEE, Honolulu, HI (Jul 2017). https://doi.org/10.1109/CVPR.2017.106, http://ieeexplore.ieee.org/document/8099589/
11. Liu, Y., Tian, Y., Zhao, Y., Yu, H., Xie, L., Wang, Y., Ye, Q., Jiao, J., Liu, Y.: VMamba: Visual State Space Model. In: Advances in Neural Information Processing Systems. vol. 37, pp. 103031–103063. Curran Associates, Inc. (2024). https://doi.org/10.52202/079017-3273, https://proceedings.neurips.cc/paper/2024/hash/baa2da9ae4bfed26520bb61d259a3653-Abstract-Conference.html

12. Msoud Nickparvar: Brain Tumor MRI Dataset. https://doi.org/10.34740/KAGGLE/DSV/14832123, https://www.kaggle.com/dsv/14832123
13. Radenovic, F., Tolias, G., Chum, O.: Fine-Tuning CNN Image Retrieval with No Human Annotation. IEEE Transactions on Pattern Analysis and Machine Intelligence **41**(7), 1655–1668 (Jul 2019). https://doi.org/10.1109/TPAMI.2018.2846566, https://ieeexplore.ieee.org/document/8382272/
14. Rasheed, Z., Ma, Y.K., Ullah, I., Al-Khasawneh, M., Almutairi, S.S., Abohashrh, M.: Integrating Convolutional Neural Networks with Attention Mechanisms for Magnetic Resonance Imaging-Based Classification of Brain Tumors. Bioengineering **11**(7), 701 (Jul 2024). https://doi.org/10.3390/bioengineering11070701, https://www.mdpi.com/2306-5354/11/7/701
15. Sathwik, T., Vishruth Reddy, T.S., Charitha, T., Singh, R.P., Kanchan, S.: An Ensemble Based Convolutional Neural Network Modelling for Classifying Medulloblastoma Subtype. In: 2024 5th International Conference for Emerging Technology (INCET). pp. 1–6. IEEE, Belgaum, India (May 2024). https://doi.org/10.1109/INCET61516.2024.10593256, https://ieeexplore.ieee.org/document/10593256/
16. Tan, M., Le, Q.: EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In: Proceedings of the 36th International Conference on Machine Learning. pp. 6105–6114. PMLR (May 2019), https://proceedings.mlr.press/v97/tan19a.html
17. Tseng, L.J., Matsuyama, A., MacDonald-Dickinson, V.: Histology: The gold standard for diagnosis? The Canadian Veterinary Journal = La Revue Veterinaire Canadienne **64**(4), 389–391 (Apr 2023)
18. Wilcoxon, F.: Individual Comparisons by Ranking Methods. Biometrics Bulletin **1**(6), 80 (Dec 1945). https://doi.org/10.2307/3001968, https://www.jstor.org/stable/10.2307/3001968?origin=crossref
19. Yue, Y., Li, Z.: MedMamba: Vision Mamba for Medical Image Classification (Sep 2024). https://doi.org/10.48550/arXiv.2403.03849, http://arxiv.org/abs/2403.03849, arXiv:2403.03849 [eess.IV]
20. Zhu, L., Lee, F., Cai, J., Yu, H., Chen, Q.: An improved feature pyramid network for object detection. Neurocomputing **483**, 127–139 (Apr 2022). https://doi.org/10.1016/j.neucom.2022.02.016, https://www.sciencedirect.com/science/article/pii/S0925231222001588