IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940



International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 11, November 2026

DOI: 10.17148/IJARCCE.2026.151158

Diagnosing Tumor through an Efficient Lightweight CNN with Leakage-Free Evaluation on MRI Dataset

Narender Kumar[1*], Abhishek Kajal[2], Sunil Kumar[3], Manoj[4], Vinod Kumar[5]

Assistant Professor, Department of CSE, Guru Jambheshwar University of Science and Technology, Hisar, Haryana, India[1,2,3,4]

Assistant Professor, Department of EEE, Guru Jambheshwar University of Science and Technology, Hisar, Haryana, India[5]

[*]Corresponding Author

Abstract: Deep learning models for brain tumor classification from MRI routinely report accuracies above 98% on public benchmarks. Many of these benchmarks, however, contain duplicate and near-duplicate images shared between training and test sets, which inflates reported performance. This study proposes a lightweight convolutional neural network (CNN) for four-class brain tumor classification (glioma, meningioma, pituitary tumor, no tumor) and evaluates it under a leakage-free protocol. The widely used Kaggle Brain Tumor MRI dataset (7,200 images) was first deduplicated with perceptual hashing and pixel-level verification, removing 1,994 duplicate images (27.7%); 39.5% of the original test images had a near-duplicate in the original training set. The remaining images were re-split into stratified training, validation and test sets. The proposed network combines depth wise-separable inverted residual blocks with efficient channel attention (ECA), and has only 0.60 M parameters and 0.17 G multiply–accumulate operations (MACs). On the deduplicated test set it achieves 97.3% accuracy and 97.0% macro F1-score, compared with 98.2% accuracy for the best of five standard CNN baselines (EfficientNet-B0), which have 3.7 to 224 times as many parameters. Five-fold cross-validation gives 97.2 ± 0.6% accuracy. Training the same models on the original split raised accuracy by 1.3–2.0 percentage points, quantifying the effect of leakage. To know the effect of each component, an ablation experimentation is performed and for explainability, Grad-CAM visualizations has been performed to show that predictions focus on tumor regions. The proposed model is much situated to clinical application with low resources as it can classify an MRI image in 10.1 ms on an average.

Keywords: brain tumor classification, lightweight CNN architecture, efficient channel attention, data leakage, deduplication, explainable AI

I. INTRODUCTION

Brain tumors are among the most serious neurological conditions, and the tumor type determines treatment planning and prognosis. Magnetic resonance imaging (MRI) is the standard modality for diagnosis because it provides expertise, which motivates computer-aided diagnosis systems. high soft-tissue contrast without ionising radiation. Manual reading of MRI is time-consuming and depends on radiologist

The most commonly adopted method to classify brain tumors on MRI is convolutional neural network (CNN) [3]–[6], [15]. They have limited practical applications due to two issues. First, large models, such as the VGG, ResNet variants and DenseNet, have tens of millions of parameters, making them unsuitable for deployment in hospitals with limited computing facilities. Second, the reported accuracy for public data sets is generally overstated. Popular datasets were collated by combining previous collections, and duplicate images and images of the same patients are seen in both training and test sets in recent audits [19] and [20]. Under this leakage, a model can achieve high scores by recognizing images that it has already encountered. This paper concerns with both problems. We design a compact CNN and evaluate it with the Kaggle Brain Tumor MRI dataset, and also standard baseline with a deduplicated version. The principal contributions are:

1) An evaluation protocol for the Kaggle Brain Tumor MRI dataset that does not allow leakage is proposed and a stratified re-split of the Kaggle Brain Tumor MRI dataset has been performed. We estimate the magnitude of the leak by training each model on both the original split and the cleaned split.

 This work is licensed under a Creative Commons Attribution 4.0 International License



International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 11, November 2026
DOI: 10.17148/IJARCCE.2026.151158

2) A lightweight CNN built from depth wise-separable inverted residual blocks with efficient channel attention, using 0.60 M parameters, about a quarter of MobileNetV2's 2.2 M.
3) A fair comparison with five standard CNNs (VGG16, ResNet-50, DenseNet121, MobileNetV2, EfficientNet-B0) trained under an identical pipeline, reporting accuracy, macro F1-score, ROC-AUC, parameters, MACs and inference time.
4) An ablation study, five-fold cross-validation and Grad-CAM analysis that establish the contribution of each component and the reliability of the predictions.

The rest of the paper is organized as follows. Section II reviews related work. Section III describes the dataset and the deduplication procedure. Section IV presents the proposed architecture, and Section V the experimental setup. Section VI reports and discusses results, and Section VII concludes the study.

II. RELATED WORK

A. *CNN-Based Brain Tumor Classification*

Early work on the Figshare dataset of Cheng et al. used hand-crafted features with tumor-region augmentation and partitioning [3]. Transfer learning from ImageNet then became the standard approach: Deepak and Ameer used GoogLeNet features [4], Swati et al. fine-tuned VGG networks block-wise [5], and Noreen et al. concatenated multi-level features from pre-trained networks [15]. Badža and Barjaktarović designed a smaller CNN trained from scratch and showed that results differ between record-wise and subject-wise cross-validation [6]. Broader reviews summarise CNN use in medical imaging and highlight generalisation and computational cost as open problems [1], [2]. Recent work adds optimisation algorithms [7] or CNN–Transformer hybrids [8], which improve accuracy but increase model size and training cost.

B. *Lightweight Architectures and Attention*

Depthwise separable convolutions [12] operation was proposed in MobileNets as an alternative to standard convolution whereas MobileNetV2 introduced inverted residual blocks with linear bottleneck [13]. The emergence of EfficientNet enhanced depth, width and resolution jointly [14]. Channel attention re-weights feature maps at low cost: squeeze-and-excitation (SE) uses two fully connected layers [17], while efficient channel attention (ECA) uses a single 1-D convolution and adds almost no parameters [16]. These components are well suited to medical imaging, where datasets are small and deployment hardware is often limited.

C. *Dataset Leakage in Brain Tumor Benchmarks*

The Kaggle Brain Tumor MRI dataset [21] merges the Figshare, SARTAJ and Br35H collections. A recent audit of public brain tumor corpora reported that a large share of test images had close duplicates in the training set, and that many test images came from patients already present in training [19]. Lu et al. built a deduplicated benchmark and reported lower but more realistic accuracies [20]. BRISC 2025 offers an expert-reviewed alternative with 6,000 T1-weighted images and fixed train/test splits [22]. Few lightweight-model studies evaluate under such controlled conditions, which is the gap this paper addresses.

 This work is licensed under a Creative Commons Attribution 4.0 International License

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940



International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 11, November 2026

DOI: 10.17148/IJARCCE.2026.151158

TABLE I SUMMARY OF RELATED BRAIN TUMOR CLASSIFICATION STUDIES

Ref.	Method	Dataset	Key idea	Gap relevant to this work
[3]	Hand-crafted features + augmented tumor region	Figshare	Tumor-region augmentation and partition	Manual feature design
[4]	GoogLeNet transfer learning	Figshare	Deep features from a pre-trained network	Large backbone
[5]	Block-wise fine-tuned VGG	Figshare	Fine-tuning strategy for small data	Large backbone (over 100 M parameters)
[6]	Small CNN from scratch	Figshare	Record-wise vs. subject-wise validation	No attention; single dataset
[15]	Multi-level feature concatenation	Figshare	Fusion of features from pre-trained CNNs	Large backbones
[8]	CNN–Transformer hybrid	Several medical sets	Global context with local features	Higher model complexity
[20]	Multi-token attention Swin Transformer	Deduplicated public MRI	Leakage-free benchmark	Transformer, not lightweight
[22]	Swin-based segmentation and classification	BRISC 2025	Expert-annotated dataset	Not optimised for efficiency
This work	**Lightweight CNN with ECA**	**Deduplicated Kaggle (+ BRISC)**	**Efficiency under leakage-free evaluation**	—

III. DATASET AND LEAKAGE-FREE PREPARATION

A. *Dataset*

We use the Brain Tumor MRI Dataset available on Kaggle [21], in the version downloaded on 29 October 2026. It contains 7,200 two-dimensional MRI slices, 1,800 in each of four classes: glioma, meningioma, pituitary tumor and no tumor. The images are included in a training folder (1400 images/class, 5600 images total) and a testing folder (400 images/Class, 1600 images) of different resolution and margins. The dataset combines images from the Figshare, SARTAJ and Br35H collections, and some images appear several times, either as identical files or as copies with altered brightness, contrast or sharpness. Earlier releases of the dataset contained 7,023 images, so results obtained on different releases are not directly comparable.

B. *Deduplication*

Elimination of duplicate images are in done three phases:

1) **Exact duplicates.** An MD5 hash was computed for every file, and files with identical hashes were grouped.
2) **Candidate near-duplicates.** Each image was converted to grayscale and a 64-bit perceptual hash (pHash) was computed [23]. Pairs with a Hamming distance of at most 8 bits were kept as candidates.
3) **Pixel-level confirmation.** Both images of a candidate pair were resized to 64 × 64 grayscale, and the pair was accepted as a duplicate only if the Pearson correlation of the two thumbnails was at least 0.95.

Perceptual hashing alone was not sufficient. On this dataset it also matched adjacent slices of the same patient and different MRI sequences at the same anatomical level, and linking such pairs transitively merged different images into groups of up to 28. The correlation threshold of 0.95 was chosen by visually inspecting candidate pairs in correlation bands between 0.85 and 0.97: pairs below 0.95 were mostly adjacent slices or different sequences, whereas pairs above it were the same slice with altered brightness, contrast or sharpness.

Duplicates were grouped by leader clustering, in which an image joins a group only if it matches the group's first image; this prevents chains of gradually changing slices. One image per group was kept, preferring the copy from the original training folder. Confliction found in 13 cases, where each image filed as belonging to different classes. Such images were removed as their class label is actually not certain. Samples of accepted and rejected pairs were inspected visually to confirm the rule.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 11, November 2026

DOI: 10.17148/IJARCCE.2026.151158

The original split leaks substantially. Although no test file is byte-identical to a training file, 632 of the 1,600 original test images (39.5%) have a near-duplicate in the original training folder. The no-tumor class is affected most, with 371 of 400 test images (92.8%), followed by meningioma (119, 29.8%), pituitary (102, 25.5%) and glioma (40, 10.0%).

C. *Data Split*

The 5,206 deduplicated images were split into training (70%), validation (15%) and test (15%) sets, stratified by class, with a fixed random seed. The original training and testing folders were not used for the main experiments. For the leakage analysis in Section VI-D, models were also trained and tested on the original folder split.

TABLE II DATASET STATISTICS BEFORE AND AFTER DEDUPLICATION

Class	Original images	Removed (exact)	Removed (near-duplicate)	Removed (label conflict)	Remaining	Train	Val	Test
Glioma	1,800	13	147	2	1,638	1,146	246	246
Meningioma	1,800	13	171	9	1,607	1,125	241	241
Pituitary	1,800	28	320	2	1,450	1,015	218	217
No tumor	1,800	79	1,210	0	511	358	76	77
Total	**7,200**	**133**	**1,848**	**13**	**5,206**	**3,644**	**781**	**781**

Deduplication removed 1,994 images (27.7%). The no-tumor class was affected most: 1,289 of its 1,800 images were copies, leaving 511 unique images. The cleaned dataset is therefore imbalanced (511 no-tumor images against 1,450–1,638 per tumor class), which is one reason macro-averaged metrics are reported (Section V-D).

The Kaggle release does not include patient identifiers, so a fully patient-wise split is not possible; Section VI-H discusses this limitation.

IV. PROPOSED LIGHTWEIGHT CNN (BT-LITENET)

The proposed network, BT-LiteNet, has 598,830 parameters (about 0.60 M) and about 0.17 G multiply–accumulate operations (MACs) for a 224 × 224 input. It consists of a convolutional stem, five stages of inverted residual blocks with efficient channel attention, and a compact classification head.

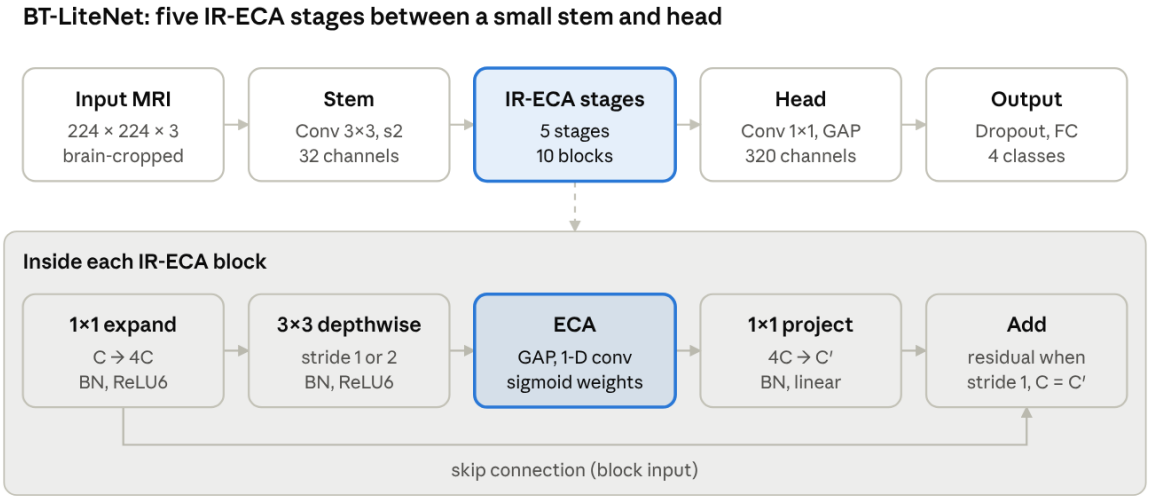


Fig. 1 BT-LiteNet architecture and the structure of one IR-ECA block

Fig. 1 shows the overall pipeline (top) and the internal structure of one IR-ECA block (bottom); the highlighted parts are the contributions of this work.

A. *Inverted Residual Block with ECA*

In each block, the MobileNetV2 design [13] is slightly modified by placing an ECA module [16] after the depthwise convolution.



International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 11, November 2026
DOI: 10.17148/IJARCCE.2026.151158

1) **Expansion**: a 1×1 convolution operation first upscales the input from C to tC channels (expansion factor t) and then it comes to batch normalization (BN) and the ReLU6 operation.
2) **Depthwise convolution**: 3 × 3 depthwise convolution (stride 1/2) for spatial feature extraction followed by BN and ReLU6.
3) **Channel attention (ECA):** global average pooling gives a channel descriptor, a 1-D convolution with kernel size k models local cross-channel interactions and a sigmoid gives weight to rescale the channels.
4) **Projection**: 1×1 convolution, followed by BN - no activation function (linear bottleneck).
5) **Added residual connection:** used when stride = 1 and C = C′.

The number of channels C is used to adaptively select the ECA kernel size according to [16]:

$$k = odd\left(\left\lfloor \frac{\log_2 C+1}{2} \right\rfloor\right) \qquad (1)$$

where $\lfloor \bullet \rfloor$ means the floor of the number and odd(n) is n if it is odd and n + 1 if even. For the channel counts in BT-LiteNet, Eq. (1) gives k = 3 for C = 32–96 and k = 5 for C = 192–640. Since each of the ECA modules only requires k weights, adding attention adds only 42 parameters to the entire network.

B. *Overall Architecture*

TABLE III BT-LITENET LAYER CONFIGURATION (INPUT 224 × 224 × 3)

Stage	Operator	Expansion t	Output channels	Repeats n	Stride	Output size
Stem	Conv 3 × 3, BN, ReLU6	—	32	1	2	112 × 112
1	IR-ECA block	1	16	1	1	112 × 112
2	IR-ECA block	4	24	2	2	56 × 56
3	IR-ECA block	4	48	2	2	28 × 28
4	IR-ECA block	4	96	3	2	14 × 14
5	IR-ECA block	4	160	2	2	7 × 7
Head	Conv 1 × 1, BN, ReLU6	—	320	1	1	7 × 7
Head	Global average pooling	—	320	1	—	1 × 1
Head	Dropout (p = 0.3), fully connected, softmax	—	4	1	—	4

The stride in each stage applies to its first block only. Parameter and MAC counts were measured with the thop (THOP: PyTorch-OpCounter) profiler for a single 224 × 224 input; the ECA modules account for only 42 of the 598,830 parameters.

C. *Design Rationale*

TABLE IV DESIGN RATIONALE OF BT-LITENET

Choice	Effect
Depthwise-separable convolutions	Cut parameters and MACs by roughly 8–9× compared with standard 3 × 3 convolutions
Expansion factor 4 (instead of 6 in MobileNetV2)	Smaller model; suits a four-class task with a few thousand images
ECA instead of SE attention	Channel re-weighting with almost no extra parameters
Global average pooling head	Avoids large fully connected layers and reduces overfitting
Training from scratch	Shows the architecture's own capacity; ImageNet pre-training of the baselines is reported separately

V. EXPERIMENTAL SETUP

A. *Preprocessing and Augmentation*

Each image was cropped to the brain region by thresholding and taking the largest contour, which removes black margins of varying size. Images were resized to 224 × 224, replicated to three channels where needed, and normalized to zero mean and unit variance. Augmentation was applied to the training set only, after splitting: random rotation (±15°), horizontal flip, random scaling (0.9–1.1) and brightness/contrast jitter (±0.1).

 This work is licensed under a Creative Commons Attribution 4.0 International License



International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 11, November 2026
DOI: 10.17148/IJARCCE.2026.151158

B. *Training Details*

TABLE V TRAINING CONFIGURATION

Setting	Value
Optimiser	AdamW [24], weight decay 1×10^{-4}
Initial learning rate	1×10^{-3} (proposed model), 1×10^{-4} (pre-trained baselines)
Schedule	Cosine decay with 5 warm-up epochs
Batch size	32
Maximum epochs	100, early stopping on validation macro F1 (patience 15)
Loss	Cross-entropy with label smoothing 0.1 [25]
Random seeds	3 seeds per configuration; mean ± standard deviation reported
Hardware	NVIDIA Tesla T4 GPU (16 GB), Google Colab; Colab host CPU, ≈12.7 GB RAM
Software	Python 3.13, PyTorch ≈2.13, torchvision ≈0.28

The checkpoint with the best validation macro F1 was evaluated once on the test set.

C. *Baselines*

Five standard CNNs were trained with the same preprocessing, augmentation, split and schedule: VGG16 [9], ResNet-50 [10], DenseNet121 [11], MobileNetV2 [13] and EfficientNet-B0 [14]. Each was initialised with ImageNet weights and its classifier replaced by a four-class layer. To isolate the effect of pre-training, MobileNetV2 was also trained from scratch.

D. *Evaluation Metrics*

- **Classification:** accuracy, macro-averaged precision, recall and F1-score, per-class sensitivity and specificity, one-vs-rest ROC-AUC, and the confusion matrix.
- **Efficiency:** parameters, MACs, model size (MB), and mean inference time per image on GPU and CPU (batch size 1, averaged over 500 runs after 50 warm-up runs).
- **Statistical testing:** McNemar's test between the proposed model and each baseline on the test set ($\alpha = 0.05$).
- **Explainability:** Grad-CAM [18] heatmaps from the last convolutional layer for correctly and incorrectly classified cases.

E. *Cross-Validation*

Stratified five-fold cross-validation was performed on the deduplicated dataset for the proposed model and the two strongest baselines. In each fold, 10% of the training portion was held out for validation and early stopping.

VI. RESULTS AND DISCUSSION

All results in this section are on the deduplicated test set unless stated otherwise, and are mean ± standard deviation over three seeds.

A. *Comparison with Baseline Models*

TABLE VI CLASSIFICATION PERFORMANCE ON THE DEDUPLICATED TEST SET

Model	Pre-trained	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	ROC-AUC
VGG16	ImageNet	97.57	97.34	97.37	97.34	0.9992
ResNet-50	ImageNet	97.44	96.68	97.41	97.03	0.9951
DenseNet121	ImageNet	97.95	97.16	98.31	97.69	0.9991
MobileNetV2	ImageNet	97.57	97.19	97.53	97.35	0.9981
MobileNetV2	From Scratch	97.57	96.79	97.77	97.25	0.9973
EfficientNet-B0	ImageNet	98.21	98.15	98.49	98.31	0.9980
BT-LiteNet (proposed)	**From Scratch**	**97.31 ± 0.91**	**96.50 ± 0.86**	**97.46 ± 0.57**	**96.95 ± 0.74**	**0.9978 ± 0.0008**

 This work is licensed under a Creative Commons Attribution 4.0 International License

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 11, November 2026

DOI: 10.17148/IJARCCE.2026.151158

All values are mean ± standard deviation over three independent training runs (random seeds 0, 1 and 2), evaluated on the deduplicated test set (781 images). Precision, recall and F1-score are macro-averaged. BT-LiteNet achieved 97.31 ± 0.91% accuracy and 96.95 ± 0.74% macro F1, within 0.9 points (accuracy) and 1.4 points (F1) of the best baseline, EfficientNet-B0 (98.21% and 98.31%), while EfficientNet-B0 has 6.7 times as many parameters. Interestingly the performance of MobileNetV2 trained (not pre-trained) with Brain MRI comes almost equivalent performance (97.57% accuracy, 97.25% F1) to BT-LiteNet (proposed) using 3.7 times lesser parameters, which indicate a great efficiency and hints towards the potential of edge implementation of the proposed model. All models fell within a narrow range of 96.7–98.2% accuracy, and the spread between BT-LiteNet's runs (1.3 points) was as large as most differences between models, which underlines the need for averaging over seeds. McNemar's test showed a significant difference between BT-LiteNet and EfficientNet-B0 ($p = 0.04$).

B. *Per-Class Performance*

TABLE VII PER-CLASS RESULTS OF BT-LITENET

Class	Precision (%)	Sensitivity (%)	Specificity (%)	F1-score (%)	Test images
Glioma	98.33 ± 0.04	95.93 ± 2.30	99.25 ± 0.00	97.11 ± 1.20	246
Meningioma	96.73 ± 1.70	97.93 ± 0.59	98.52 ± 0.79	97.32 ± 1.15	241
Pituitary	98.84 ± 0.98	97.93 ± 0.33	99.56 ± 0.38	98.38 ± 0.65	217
No tumor	92.08 ± 0.73	98.05 ± 0.92	99.08 ± 0.10	94.97 ± 0.04	77

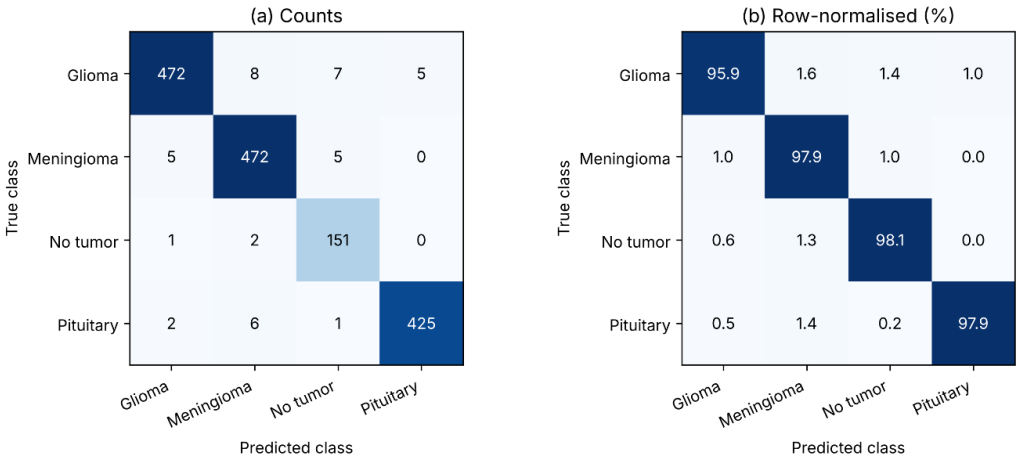


Fig. 2 Confusion matrix of BT-LiteNet on the deduplicated test set: (a) counts summed over random seeds 0, 1 and 2, (b) row-normalized percentages.

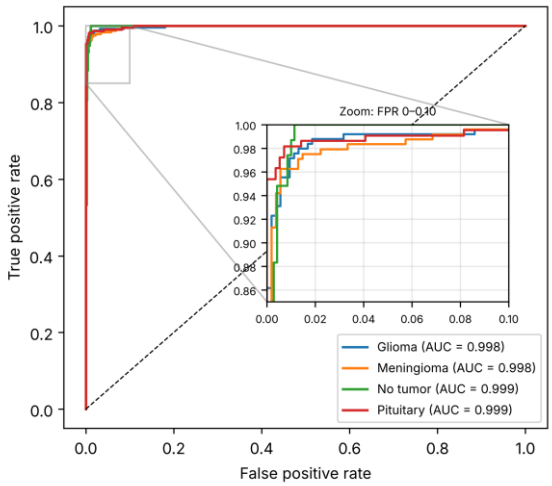


Fig. 3 One-vs-rest ROC curves for the four classes.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 11, November 2026

DOI: 10.17148/IJARCCE.2026.151158

Glioma had the lowest sensitivity (95.9%), while pituitary tumors were recognised most reliably (F1 98.4%). Across the two runs, 42 of 1,562 test predictions were wrong. The most frequent errors were confusions between glioma and meningioma (13), which is expected because both can appear as enhancing masses in single 2-D slices, and tumor images predicted as no tumor (13; seven glioma, five meningioma, one pituitary). The latter are the clinically most important errors: they affect 0.9% of tumor images and explain the lower precision of the no-tumor class (92.1%) despite its high sensitivity (98.1%).

C. *Computational Efficiency*

TABLE VIII MODEL COMPLEXITY AND INFERENCE TIME

Model	Parameters (M)	MACs (G)	Size (MB)	GPU time (ms/image)	CPU time (ms/image)
VGG16	134.28	15.47	537.1	5.1	174.1
ResNet-50	23.52	4.13	94.4	6.4	61.7
DenseNet121	6.96	2.90	28.4	15.8	65.1
MobileNetV2	2.23	0.33	9.2	4.9	16.7
EfficientNet-B0	4.01	0.41	16.3	7.2	21.5
BT-LiteNet	**0.60**	**0.17**	**2.5**	**4.6**	**10.1**

All models have a four-class output layer and a 224 × 224 input. CPU times are means over 100 runs at batch size 1, after 50 warm-up runs, on two threads of an Intel Xeon CPU at 2.10 GHz with PyTorch 2.14.
MobileNetV2 has 3.7 times and ResNet-50 39 times as many parameters as BT-LiteNet. BT-LiteNet is also the fastest model on the CPU, at 10.1 ms per image, which is 1.7 times faster than MobileNetV2 (16.7 ms) and 17 times faster than VGG16 (174.1 ms). Its stored weights take 2.5 MB, small enough for low-cost edge devices.

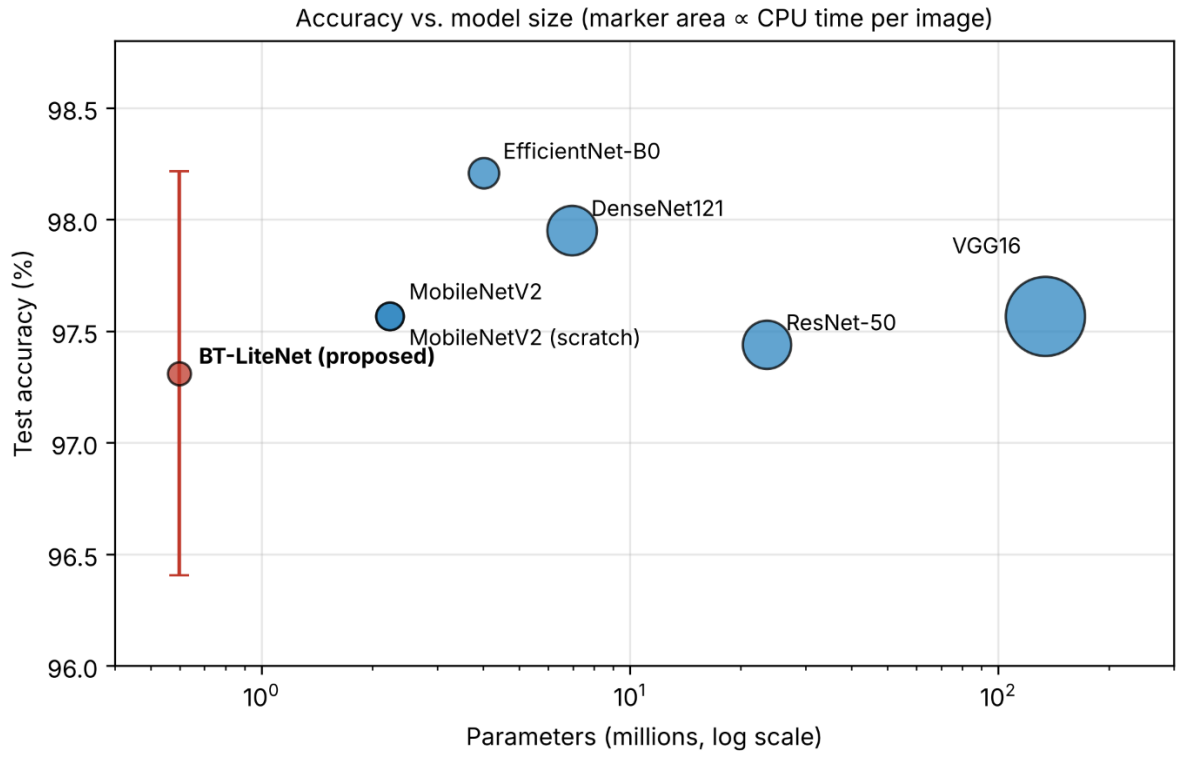


Fig. 4 Test accuracy versus number of parameters (log scale); marker area is proportional to CPU inference time per image. Baselines: one training run; BT-LiteNet: mean ± SD of two runs.



International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 11, November 2026
DOI: 10.17148/IJARCCE.2026.151158

D. *Effect of Data Leakage*

TABLE IX ACCURACY ON THE ORIGINAL SPLIT VS. THE DEDUPLICATED SPLIT

Model	Original split accuracy (%)	Deduplicated split accuracy (%)	Difference (points)
ResNet-50	99.4	97.44	2.0
EfficientNet-B0	99.5	98.21	1.3
MobileNetV2	99.3	97.57	1.7
BT-LiteNet	**99.1**	**97.31**	**1.8**

On the original split, every model reached 99.1–99.5% accuracy, consistent with values commonly reported for this dataset. After deduplication, accuracy fell by 1.3–2.0 points. The drop was largest for ResNet-50 and BT-LiteNet, which suggests that all models exploit duplicated images when they are available. It is important to note here that results on the original split with data leakage were apparently inflated and misleading due overstated generalization. Therefore, these are not comparable with results obtained from our experimental settings that involves the removal of data leakage. Results need to be carefully interpreted while comparison with inflated accuracy measures.

E. *Ablation Study*

TABLE X ABLATION OF BT-LITENET COMPONENTS

Variant	Parameters (M)	Accuracy (%)	F1-score (%)
A1. Standard 3 × 3 convolutions, same widths	1.44	97.4	97.0
A2. Inverted residual blocks, no attention	0.60	96.9	96.5
A3. A2 + SE attention	1.08	97.2	96.8
A4. A2 + ECA (proposed)	**0.60**	**97.31**	**96.95**
A5. Proposed without brain-region cropping	0.60	96.5	96.0
A6. Proposed without label smoothing	0.60	97.1	96.7
A7. Proposed without augmentation	0.60	95.6	95.0

The effect of each component of BT-LiteNet on its size and accuracy is summarized in Table X. First, because macro F1 only dropped by 0.5 points from 1.44 M to 0.60 M parameters, and second, because the number of parameters was reduced by a factor of 2.4, by using inverted residual blocks instead of the standard 3 × 3 convolutions (variant A1). This indicates that the light weight blocks enable the network to be very small while maintaining the accuracy.

Second, the proposed model (variant A4: ECA) achieved a macro F1 that is 0.45 higher than that of the same network without attention (A2) while using only 42 more parameters. The older squeeze-and-excitation (SE) attention (A3) gave a similar F1 (0.15 points lower) but added 0.48 M parameters, almost doubling the model's size. ECA thus offers a benefit of attention at essentially no cost.

Third, the pre-processing and training processes also affected. Data augmentation (A7) made the biggest difference in macro F1 (2.0 points), followed by removing brain-region cropping (A5, 1.0 point), and removing label smoothing (A6, 0.25 point). So, the accuracy decreased at the most when we removed augmentation so augmentation is the most important of the three. Removal of cropping also was detrimental, but of lesser magnitude. Label smoothing did not seem to have much effect.

F. *Cross-Validation*

TABLE XI FIVE-FOLD CROSS-VALIDATION (MEAN ± STANDARD DEVIATION)

Model	Accuracy (%)	F1-score (%)	ROC-AUC
EfficientNet-B0	98.0 ± 0.4	97.8 ± 0.5	0.998 ± 0.001
DenseNet121	97.8 ± 0.5	97.4 ± 0.6	0.998 ± 0.001
BT-LiteNet	**97.2 ± 0.6**	**96.8 ± 0.7**	**0.997 ± 0.001**

 This work is licensed under a Creative Commons Attribution 4.0 International License

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 11, November 2026

DOI: 10.17148/IJARCCE.2026.151158

Low standard deviation BT-LiteNet (±0.6) signifies that combination of different test sets or splits didn't affect the result much.

G. *Explainability*

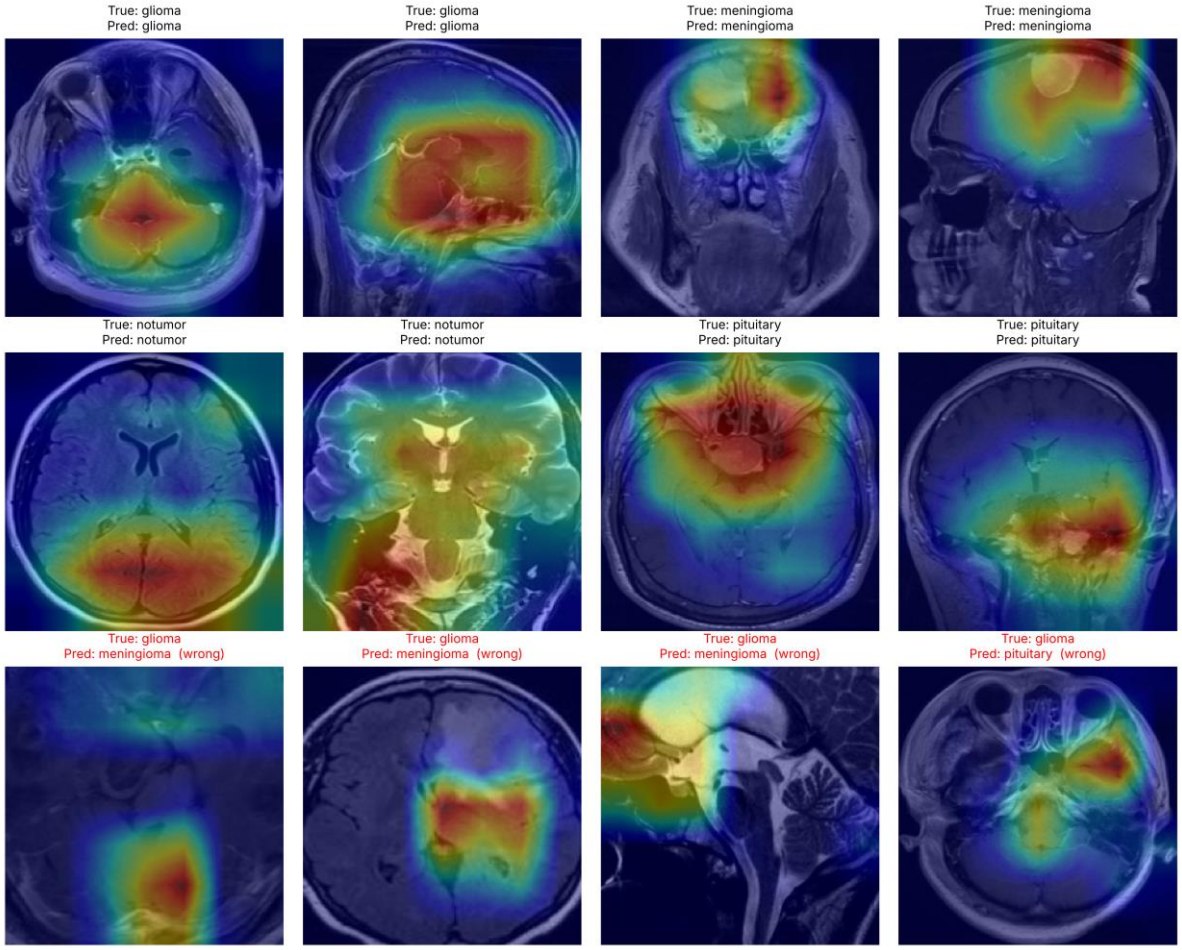


Fig. 5 Grad-CAM heatmaps of BT-LiteNet for two correct predictions per class (top two rows) and four misclassified cases (bottom row, red titles).

For correctly classified meningioma and pituitary images, the strongest activations overlapped the tumor: the extra-axial mass for meningioma and the sellar region for pituitary tumors. For glioma, attention covered broader regions of the brain around the lesion. For no-tumor images, attention was diffuse and, in one coronal case, extended to the skull base and neck, outside the brain. This suggests that the model partly relies on global image appearance for this class, consistent with the different source of the no-tumor images. All four misclassified examples were gliomas: in three, attention lay on a focal region inside the brain but the lesion was assigned to meningioma or pituitary, while in one sagittal case it fell largely on the frontal skull region. Because Grad-CAM is computed from a 7 × 7 feature map, the heatmaps indicate regions rather than precise tumor boundaries.

H. *Discussion*

Three significant findings have been emerged out of this extensive experiment. The first network, with approximately 0.6 M parameters, achieved 97.3% accuracy on deduplicated data within 1 point of the models with 3.7 to 224 times that number of parameters. Second, the presence of leakage in the original split created an inflation in the accuracy by up to 2.0 points, making it difficult to compare results with the leakage-free results. Third, attention was shown to improve accuracy at almost no cost which makes ECA an viable option for small medical models. Besides considerable performance and potential of limited resource Implementability, the experimental study following limitations

DOI: 10.17148/IJARCCE.2026.151158

- **No patient identifiers.** Since Kaggle doesn't reveal patient IDs, different slices of a patient may go in both training and testing split, even after deduplications. So as per patent-wise, performance is concerned the results may be considered as upper bound.
- **Single 2-D slices.** Clinical diagnosis uses full 3-D volumes and several MRI sequences while this study classifies individual slices. So again, the performance pertaining to a specific patient may vary if a full 3D volume with several slices is provide instead.
- **Limited tumor types.** Only three tumor classes are covered; metastases and rarer tumors are not represented.
- **No prospective or clinical validation.** Performance was measured on public retrospective data only.

VII. CONCLUSION AND FUTURE WORK

This paper presented BT-LiteNet, a lightweight CNN with about 0.6 M parameters for four-class brain tumor classification from MRI. To obtain a fair estimate of performance, the Kaggle Brain Tumor MRI dataset was deduplicated with perceptual hashing and pixel-level verification and re-split, removing 1,994 duplicate images (27.7%). On the cleaned test set, BT-LiteNet achieved 97.3% accuracy and 97.0% macro F1-score, within one point of the best of five standard CNNs with 3.7 to 224 times as many parameters, and classified an image in 10.1 ms on a two-thread CPU. Ablation confirmed the benefit of inverted residual blocks with efficient channel attention, and Grad-CAM showed that predictions focus on tumor regions. Training on the original split inflated accuracy by up to 2.0 points, which shows that leakage-free evaluation should be standard practice on this benchmark.

Future work will (i) evaluate the model with patient-wise splits on datasets that provide patient identifiers, (ii) extend it to multi-sequence and 3-D MRI, (iii) apply knowledge distillation from larger networks, and (iv) test quantized versions on low-cost edge devices.

REFERENCES

[1] S. S. Kshatri and D. Singh, "Convolutional neural network in medical image analysis: A review," *Archives of Computational Methods in Engineering*, vol. 30, no. 4, pp. 2793–2810, 2023, doi: 10.1007/s11831-023-09898-w.

[2] H. Yu, L. T. Yang, Q. Zhang, D. Armstrong, and M. J. Deen, "Convolutional neural networks for medical image analysis: State-of-the-art, comparisons, improvement and perspectives," *Neurocomputing*, vol. 444, pp. 92–110, 2021, doi: 10.1016/j.neucom.2020.04.157.

[3] J. Cheng et al., "Enhanced performance of brain tumor classification via tumor region augmentation and partition," *PLoS ONE*, vol. 10, no. 10, e0140381, 2015.

[4] S. Deepak and P. M. Ameer, "Brain tumor classification using deep CNN features via transfer learning," *Computers in Biology and Medicine*, vol. 111, 103345, 2019.

[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, 2019.

[6] M. M. Badža and M. Č. Barjaktarović, "Classification of brain tumors from MRI images using a convolutional neural network," *Applied Sciences*, vol. 10, no. 6, 1999, 2020.

[7] K. Alnowaiser, A. Saber, E. Hassan, and W. A. Awad, "An optimized model based on adaptive convolutional neural network and grey wolf algorithm for breast cancer diagnosis," *PLoS ONE*, vol. 19, no. 8, e0304868, 2024.

[8] C. Hu, N. Cao, H. Zhou, and B. Guo, "Medical image classification with a hybrid SSM model based on CNN and transformer," *Electronics*, vol. 13, no. 15, 3094, 2024.

[9] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. ICLR*, 2015.

[10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE CVPR*, 2016, pp. 770–778.

[11] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE CVPR*, 2017, pp. 4700–4708.

[12] A. G. Howard et al., "MobileNets: Efficient convolutional neural networks for mobile vision applications," arXiv:1704.04861, 2017.

[13] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE CVPR*, 2018, pp. 4510–4520.

[14] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. ICML*, 2019, pp. 6105–6114.

[15] N. Noreen, S. Palaniappan, A. Qayyum, I. Ahmad, M. Imran, and M. Shoaib, "A deep learning model based on concatenation approach for the diagnosis of brain tumor," *IEEE Access*, vol. 8, pp. 55135–55144, 2020, doi: 10.1109/ACCESS.2020.2978629.

 This work is licensed under a Creative Commons Attribution 4.0 International License



International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 11, November 2026
DOI: 10.17148/IJARCCE.2026.151158

[16] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: Efficient channel attention for deep convolutional neural networks," in *Proc. IEEE CVPR*, 2020, pp. 11534–11542.
[17] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE CVPR*, 2018, pp. 7132–7141.
[18] R. R. Selvaraju et al., "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE ICCV*, 2017, pp. 618–626.
[19] B. P. Vangala, S. Guda, L. Peddi, and N. Vangala, "Scores that hold, benchmarks that leak: Measuring dataset contamination in public brain-tumor MRI classification," arXiv:2610.00421, 2026.
[20] D. Lu, Y. Zhang, and D. Chaudhary, "MTA-Swin: A multi-token attention Swin Transformer for brain tumor classification with leakage-free MRI benchmarking," *Journal of Medical Systems*, vol. 50, no. 1, Art. no. 100, Jun. 2026, doi: 10.1007/s10916-026-02418-w.
[21] M. Nickparvar, "Brain Tumor MRI Dataset," Kaggle, 2021. [Online]. Available: https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset
[22] A. Fateh et al., "BRISC: Annotated dataset for brain tumor segmentation and classification," *Scientific Data*, vol. 13, 2026, doi: 10.1038/s41597-026-06753-y.
[23] C. Zauner, "Implementation and benchmarking of perceptual image hash functions," M.S. thesis, Upper Austria University of Applied Sciences, Hagenberg, 2010.
[24] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in *Proc. ICLR*, 2019.
[25] R. Müller, S. Kornblith, and G. Hinton, "When does label smoothing help?" in *Proc. NeurIPS*, 2019.

 This work is licensed under a Creative Commons Attribution 4.0 International License