Lightweight-CancerNet: a deep learning approach for brain tumor detection

Asif Raza and Muhammad Javed Iqbal

Department of Computer Science, University of Engineering and Technology Taxila, Taxila, Punjab, Pakistan

ABSTRACT

Detecting brain tumors in medical imaging is challenging, requiring precise and rapid diagnosis. Deep learning techniques have shown encouraging results in this field. However, current models require significant computer resources and are computationally demanding. To overcome these constraints, we suggested a new deep learning architecture named Lightweight-CancerNet, designed to detect brain tumors efficiently and accurately. The proposed framework utilizes MobileNet architecture as the backbone and NanoDet as the primary detection component, resulting in a notable mean average precision (mAP) of 93.8% and an accuracy of 98%. In addition, we implemented enhancements to minimize computing time without compromising accuracy, rendering our model appropriate for real-time object detection applications. The framework's ability to detect brain tumors with different image distortions has been demonstrated through extensive tests combining two magnetic resonance imaging (MRI) datasets. This research has shown that our framework is both resilient and reliable. The proposed model can improve patient outcomes and facilitate decision-making in brain surgery while contributing to the development of deep learning in medical imaging.

INTRODUCTION

A tumor, characterized by uncontrolled cell division and the formation of an abnormal mass, can harm the brain, which consists of billions of cells. Cancer can develop either within or outside of the brain. This disorder exhibits the highest rates of morbidity and mortality compared to all other forms of cancer that affect both adults and children. Diagnosing brain tumors is a complex task due to the elusive nature of their beginnings and the challenge of accurately determining their growth rates (*Kleihues et al., 2002*).

Brain tumors are a significant health issue, with their occurrence rates consistently increasing. The National Cancer Institute (NCI) confirms that brain tumors are most commonly diagnosed in adults aged 50 or older (*National Cancer Institute, 2024*). By gender, they are found more predominantly in males (*Popat & Patel, 2022*). Unfortunately, brain tumors have become the leading cause of death in individuals below the age of 20 (*DeAngelis, 2001*).

In medical literature, enhancing the accuracy and efficiency of brain tumor detection has become a crucial factor. Intelligent diagnostic methods and treatment strategies are

essential for addressing this issue. The utilization of deep learning techniques for detecting brain tumors has received considerable interest, especially in the application of edge computing for medical prioritization systems. Deep neural networks, a key component of deep learning, have garnered the attention of researchers, engineers, and healthcare professionals. For instance, convolutional neural networks (CNNs) are commonly employed in medical image analysis, yielding promising results (*Gupta, Gaurav & Arya, 2024*; *Qureshi et al., 2022a*; *Cen et al., 2019*; *Gao et al., 2019*).

Magnetic resonance imaging (MRI) stands out as the go-to non-invasive technique encouraged by radiologists for scanning (*Ardan & Indraswari, 2024*; *Kale & Gadicha, 2024*; *Nawaz et al., 2021*; *Asiri et al., 2024*). It is beneficial for detecting subtle structural changes that might be difficult to detect with computed tomography (CT) scans (*Simo et al., 2024*; *Aamir et al., 2022*). However, identifying the specific type of tumor can be challenging, particularly when time is limited, as is often the case during the prediction stage of artificial intelligence-based solutions.

Real-time detection could expedite the diagnostic process compared to traditional methods (*Mahmud, Mamun & Abdelgawad, 2023*). Moreover, real-time detection during surgery could provide more precise information about the tumor's boundaries and location, allowing for accurate and potentially less invasive resections (*Mohan et al., 2022*).

While there have been significant advances in deep learning computing applications and services, we still face challenges in achieving real-time human-computer interaction with efficiency and the ability to detect unknown knowledge. In particular, clear advancements in ultra-light deep learning frameworks have yet to be observed in medical brain tumor diagnosis.

These challenges become even more critical during intraoperative brain surgery, where quick decisions are crucial. We have developed an ultra-light deep learning framework to address these gaps and provide an effective solution. This framework is intended to streamline tumor identification and enhance decision-making during brain surgery, ultimately improving patient outcomes.

The study of ultra-light deep learning frameworks (*Qureshi et al., 2022b*; *Chinaev et al., 2024*) in the medical brain tumor diagnosis presents two major challenges. Firstly, resource efficiency, *i.e.*, Ultra-light deep learning frameworks, must handle large medical imaging datasets efficiently, balancing real-time processing with limited computational resources. Secondly, maintaining diagnostic accuracy while reducing model complexity is challenging due to trade-offs between lightweight architectures and superior performance in brain tumor diagnosis.

Furthermore, focusing solely on the intelligent diagnosis of brain tumors overlooks the broader significance of how such technology can be integrated into medical triage platforms or used by doctors on mobile or advanced Reduced Instruction Set Computer (RISC) machine (ARM) devices.

Most of the research on brain tumor diagnostics has mainly concentrated on CNNs within deep learning networks (*Lee, Chae & Cho, 2024*; *Vinod, Prakash & Salman, 2024*; *Tinu, Appathurai & Muthukumaran, 2024*). A widely used deep learning network can be employed to extract features from numerous medical images. Nevertheless, there is a lack

of research examining the realistic implementation of lightweight models to diagnose brain tumors (*Chengwei, 2019*; *Howard et al., 2017*; *Qiang et al., 2021*). Most of the deep learning frameworks require substantial computational power and resources. Furthermore, deep learning algorithms fail to achieve a balance between accuracy and speed. This research presents a lightweight deep learning model called Lightweight-CancerNet as a solution to address the existing challenges. This model is designed to be utilized by ARM devices.

It is suggested that the customized Lightweight-CancerNet be used with MobileNet as the backbone, which will help extract the distinguished features from MRI images. The NanoDet is used as a head for the detection of cancerous parts. We have demonstrated the strength and reliability of our proposed model by conducting extensive experiments on tumor samples.

The notable contributions of our work are as follows:

- An accurate method capable of computing reliable image features to enhance the tumor's detection is proposed.
- A robust framework is designed to improve the mAP and accuracy to detect the cancer, tumor, or lesions.
- A comprehensive evaluation of the proposed framework is performed on a complex dataset, and its effectiveness is confirmed through rigorous experimentation.

The remainder of the article is divided into four sections. "Literature Review" covers the existing work; "Methodology" discusses the proposed methodology. In "Experimental Setup and Results" and "Conclusion", the results and the conclusion are presented.

LITERATURE REVIEW

This section has examined previous studies on detecting and categorizing cancers from medical images. Additionally, we analyzed various object detection models, analyzing their performance, strengths, and weaknesses. We have evaluated each model based on mean average precision (mAP), precision, recall, and F1 score.

Brain tumor segmentation techniques based on machine learning typically rely on voxel-based features extracted from the image's volume of interest (VOI). Various segmentation methods have been evaluated, demonstrating diverse performances (*Kumar, 2023*). *Abdusalomov, Mukhiddinov & Whangbo (2023)* employed You Only Look Once (YOLO)v7 and transfer learning techniques to enhance brain tumor diagnosis in MRI scans, achieving an exceptional 99.5% accuracy in detecting prevalent brain tumor types like Glioma, Meningioma, and Pituitary. However, they recognize the need for further investigation, especially in determining minor tumors. The methods ResNet (*Aggarwal et al., 2023*) and CRF-RNN were employed to achieve precise segmentation outcomes.

The evaluation of performance was conducted utilizing the Multimodal Brain Tumor Segmentation Benchmark (BraTS) dataset (*BraTS, 2020*). Similarly, *Ghafourian et al. (2023)* integrated the results of SVM, Naive Bayes, and k-nearest neighbors (KNN). They combined the outcomes of multiple models by averaging their values, thereby enhancing tumor classification results.

The BraTS dataset has a significant impact on brain tumor imaging studies. *Menze et al. (2015)* introduced BraTS as a standardized dataset with MRI scans and tumor annotations. *Bakas et al. (2017)* augmented BraTS by incorporating expert annotations and radiometric characteristics. Further, *Bakas et al. (2019)* identified machine-learning algorithms for tumor segmentation and prediction.

Furthermore, segmentation and diagnosis studies using MR images were conducted. This research employed a search algorithm to apply a thresholding technique. Morphological operations and connected component analysis were employed to minimize image noise and enhance brain tumor identification. Comparisons with CNN algorithms revealed high success rates in the obtained results (*Aleid et al., 2023*).

The concept of lightweight models has emerged in recent years, aiming to address model size and speed issues. Unlike working with pre-trained models, designing lightweight models offers an alternative approach. *Shelatkar et al. (2022)* introduced a novel way of diagnosing brain tumors using a lightweight deep-learning model with a fine-tuning methodology. The dataset acquired from the RSNA-MICCAI brain tumor radiogenomic classification was utilized in this investigation. The preprocessed data is partitioned into several subsets for testing and training the model. The accuracy of the YOLOv5 model is reported to be 88% (*Shelatkar et al., 2022*).

MobileNetV1 introduces a significant advancement in depth-wise separable convolution. The network forgoes conventional standard convolutions. The tumor detection accuracy of MobileNet has been reported to be 82.61% (*Ullah et al., 2022*).

Single Shot Detector (SSD) is a scanner that detects a single event. The proposed network does not include a region proposal; instead, it predicts the bounding boxes and the classes directly from feature maps in a single run. SSD incorporates small convolutional filters to enhance precision and predict object classes and offsets to the default bounding boxes (*Cen et al., 2019*). The SSD has an accuracy rate of 82.7% in detecting cancer on endoscopic images.

Ronneberger, Fischer & Brox (2015) proposed the U-Net architecture, which has now become a standard for medical image segmentation tasks, specifically in brain tumor detection. The U-Net model accomplished a precision rate of 93.4% when evaluated on the BraTS 2013 dataset, as reported by *Menze et al. (2015)* in 2015. However, the system's computational complexity and memory constraints hinder its implementation on devices with limited resources.

To overcome this constraint, *Howard et al. (2017)* introduced MobileNet, a CNN design that is less resource-intensive yet achieves similar accuracy to U-Net. MobileNet employs depthwise separable convolutions to decrease computational complexity and memory consumption, resulting in a 91.2% accuracy rate on the BraTS 2017 dataset (*Howard et al., 2017*).

In 2018, *Zhang et al. (2017)* presented ShuffleNet, a lightweight CNN architecture with enhanced efficiency and accuracy. ShuffleNet employs group convolutions and channel shuffle operations to decrease computational complexity and memory consumption, resulting in an accuracy of 92.5% on the BraTS 2013 dataset.

In 2017, *Iandola et al. (2016)* introduced SqueezeNet, a very efficient CNN design that offers cutting-edge performance while demanding minimal computational resources. The SqueezeNet utilizes knowledge distillation to train a compact model that emulates the functionality of a pre-existing, more intricate model, resulting in 94.1% accuracy on the BraTS 2017 data.

Additional research has investigated the application of deep learning models for the identification of brain tumors, such as DeepMedic (*Kamnitsas et al., 2016*), 3D U-Net (*Çiçek et al., 2016*), U-Net++ (*Zhou et al., 2018*), and VGG-16 (*Simonyan & Zisserman, 2015*). These models have attained different levels of precision. However, it frequently requires substantial computer resources and memory.

Despite achieving a reasonable level of accuracy, there remains a possibility for further enhancement in detecting brain tumors. Although techniques have potential, their ability to analyze various datasets and clinical settings may need to be revised. Furthermore, further improvements are required to enhance accuracy and ensure reliable diagnosis in real-world situations.
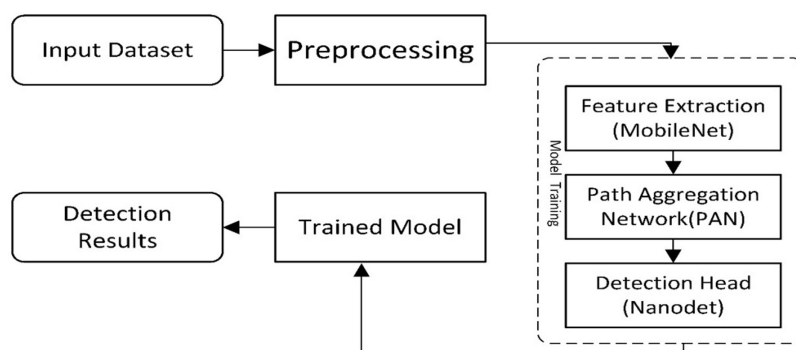
METHODOLOGY

Efficient and effective automated localization and detection of cancer from medical images, *i.e.*, CT, endoscopy, MRI X-rays, *etc.*, is still a complex task because of the vast slight variations in the size, color, and position of lesions. Moreover, medical investigations require lightweight models that can be used through mobile or arm devices. In this work, we tried to overcome the challenges above by fine-tuning an efficient model. The proposed method is robust and lightweight, providing better mAP detection for cancer and lesions in medical images.

The presented work is composed of two main components. The first is 'Data Preparation', and the second is a deep learning-based model for cancer detection. The first module assembles two datasets to locate the exact region of interest, while the second module of the new model presents a cancer/lesion detection model. This module consists of two sub-modules: the first is the study of different lightweight models used for cancer detection, and the second will be the training component, which performs training using the critical points computed from the model. Figure 1 illustrates the flow diagram of the proposed methodology, and Fig. 2 presents the overview of the proposed work.

The steps for a lightweight cancer detection model are depicted in the flow diagram. It starts by taking an input image dataset, cleaning and preparing it, and then extracts key feature vectors using a MobileNet backbone. These features are further processed by a path aggregation network (PAN) before reaching the NanoDet detection head, which identifies and locates objects in the image. Finally, the trained model is used to analyze detection results for inference.

Datasets

For training the model, we have used two different publicly available datasets. One is the Multimodal Brain Tumor Segmentation Benchmark (BraTS) (*Bakas, 2018*) with 1,140 images. It is a publicaly available brain tumor dataset that contains four MRI modalities,

Figure 1 A visual overview of the Lightweight-CancerNet framework's methodological flow, illustrating each stage from data preprocessing to model training and evaluation.
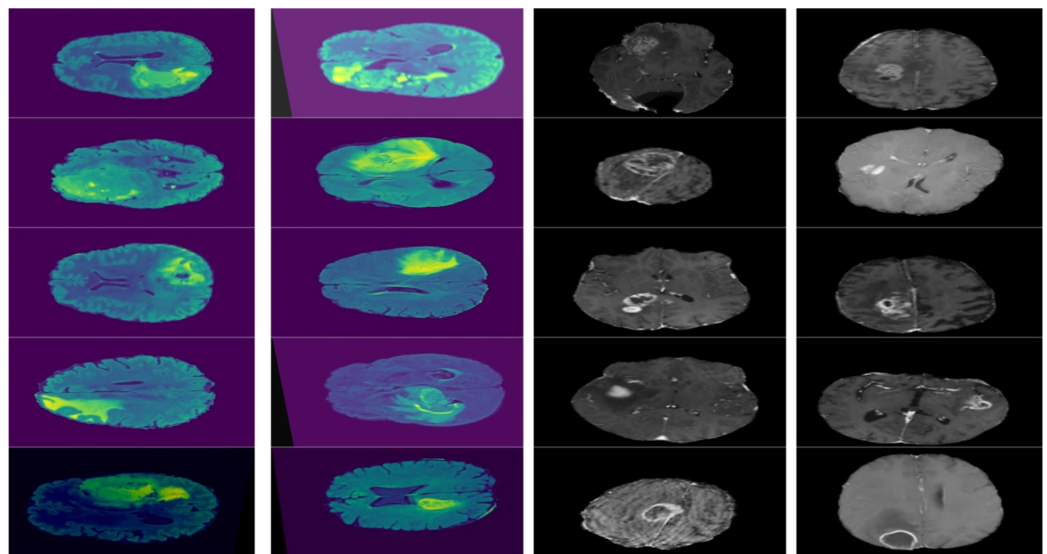
Full-size DOI: 10.7717/peerj-cs.2670/fig-1
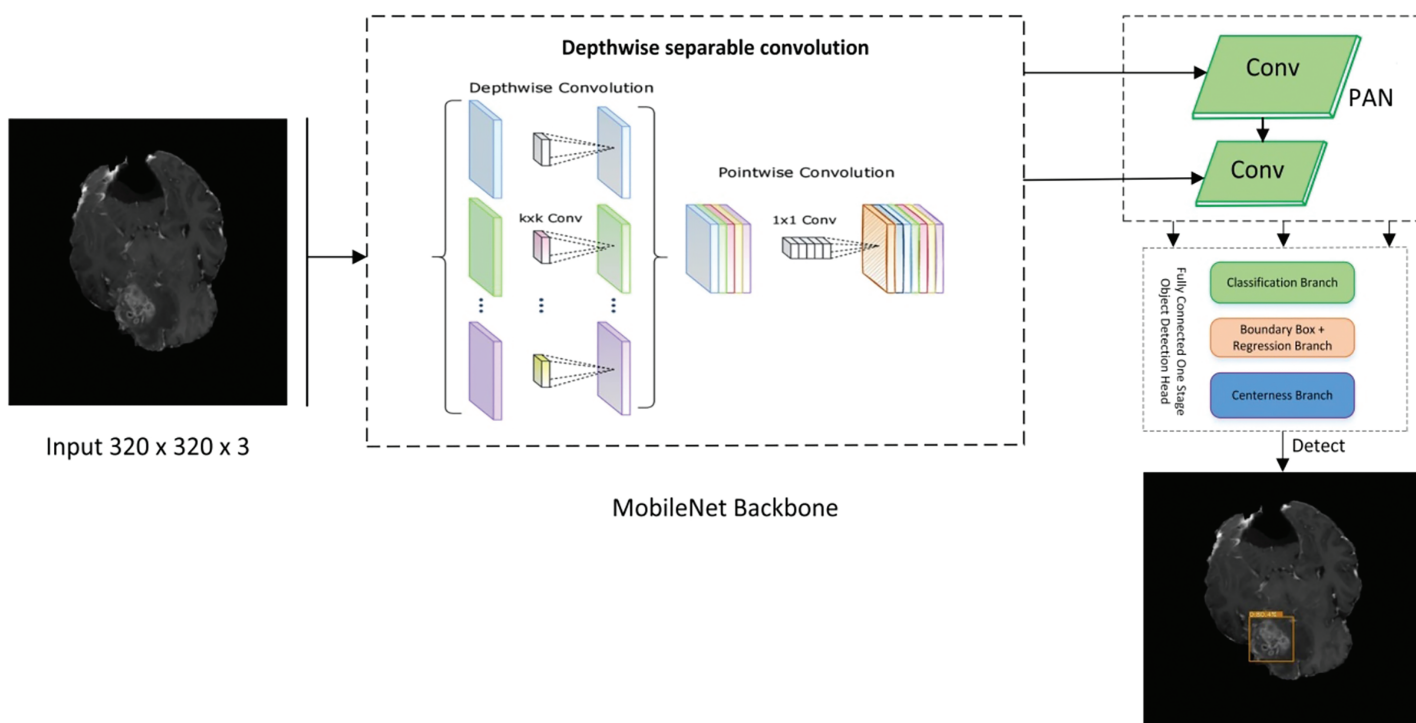
Figure 2 Proposed lightweight cancernet with three modules feature extraction by using MobileNet backbone, PAN for the fusion of feature maps, and NanoDet head as prediction module.

Full-size DOI: 10.7717/peerj-cs.2670/fig-2

having T1, T1ce, T2, and Flair. BraTS is used extensively by researchers (*Ardan & Indraswari, 2024*; *Bakas et al., 2019*; *Kumar, 2023*). The other one is RSNA-MICCAI competition data (*Roberts, 2021*) having 400 images. The dataset is also publically available and has images in various resolutions, colors and modalities. The dataset has been expanded 1,131 images using horizontal and vertical flip augmentation. We ensured that the augmented images would not be part of test data so that they could not effect the test

Raza and Iqbal (2025), *PeerJ Comput. Sci.*, DOI 10.7717/peerj-cs.2670

6/22

Figure 3 Displays example MRI slices with cancerous regions, offering insight into the dataset and visual characteristics used for training and testing the model.

Full-size 🖼 DOI: 10.7717/peerj-cs.2670/fig-3

results of the model. Both datasets were merged to enhance the accuracy and robustness of the findings.

Furthermore, combining datasets increases the diversity and sample size, leading to reliable conclusions. The images in the collection have a resolution of 320 × 320. Figure 3 shows the sample images from the dataset.

Annotation

In this work, first, we used the dataset's images as input images and annotated the regions of interest using the coordinates given with the datasets. Annotation is done manually using the annotation tool available at https://www.makesense.ai/. In the annotation process, positive labels and bounding box coordinates were indicated. After annotation, the images were forwarded to the MobileNet backbone for feature extraction.

MobileNet

The MobileNet architecture is commonly employed as the basis for object detection algorithms because of its exceptional efficiency and accuracy. Therefore, it is used as the backbone of the proposed framework. The backbone network of MobileNet is composed of fully connected (FC) layers followed by many depth-wise separable convolution layers. According to Eq. (1), the MobileNet backbone network can accept an input image, W × H × D as input size.

$$Imgsize = W \times H \times D \tag{1}$$

The MobileNet backbone network produces a feature map with dimensions H × W × D, where W is the width, H is height, and D is the number of feature channels. The Region Proposal Network (RPN) and object detector utilize this feature map as input to recognize

and accurately identify items in the picture. The RPN produces a sequence of region suggestions. Then, the feature map obtained from the MobileNet backbone network is passed *via* Region of Interest (ROI) pooling to extract a feature vector of a consistent size for each suggestion. The ROI pooling method involves applying a max-pooling operation to each rectangular region of the feature map that corresponds to the region suggestions.

The rationale behind the selection of MobileNet as the backbone is the selection of MobileNet as the backbone network is a well-reasoned choice that balances the need for computational efficiency, model performance, and adaptability to various mobile and embedded vision applications (*Howard et al., 2017*).

Path aggregation network (PAN)

PAN is used to fuse different features received from different paths from the MobileNet backbone. The fusion is done using the concept of element-wise addition. Equation (2) shows the simplified working of PAN.

$$Ffused = Fi + Fi^{th} \in \mathrm{R}H \times W \times C \tag{2}$$

$Fi \in \mathrm{R}\ H \times W \times C$ are the feature maps from the ith path, where H, W, C are the height, width, and number of channels, respectively. The final output of the PAN module is the fused feature map F_{fused}, which is then passed to the next network stage.

Then, the fine-tuning of hyperparameters like learning rate, hidden layers, neurons, and batch sizes is carried out. After that, we trained the NanoDet model and tested the final lightweight model against the hold-out dataset.

NanoDet

NanoDet is an advanced object detection model that does not require anchors and has various remarkable features. As an anchor-free model in the Fully Convolutional One-Stage Object Detection (FCOS) technique (*RangiLyu, 2021*). FCOS consists of three main branches. The three branches of FCOS are:

Classification Branch: It consists of a series of convolutional layers responsible for predicting the class object at each location on the feature map. The output layer employs softmax or sigmoid function to predict the class probabilities. Furthermore, it does not only predict a single class but also provides probability scores for the classes that allow it to handle multiclass problems.

Bounding Box Branch: It is responsible for the offset prediction for the bounding box coordinates relative to each location of the feature map. It predicts the distances to the sides of the bounding box that encapsulates the detected object. These distances are later transformed into absolute bounding box coordinates in the original image space.

Center-Ness Branch: This cognitive function differentiates the position of an object and its central point. The model aids in reducing the impact of low-quality bounding boxes that are distant from the object's center by calculating a centeredness score for each feature level. Greater normalized values indicate closer proximity to the center of the object. The center-ness score is high if the location is near the object's centre and low if it is near the edges.

In addition, NanoDet is exceptionally lightweight, with a model file size of either 980 KB (INT8) or 1.8 MB (FP16). Its lightning-fast speed of 97 fps (10.23 ms) on mobile ARM CPU, in addition to this, makes it an excellent option for real-time object identification apps (*RangiLyu, 2021*). The current article features a straightforward network layout and reduced network parameters, facilitating its portability and deployment.

NanoDet is used as a head, and convolutional layers for classification and regression were used. Each convolution layer comprises a Convolutional Layer, Norm Layer, and Activation Layer. These are applied to the output of the backbone, *i.e.*, input feature maps in parallel. Then, the output is generated using Eqs. (3) and (4).

$$cls\_feat = \sigma((Wcls * x) + bcls) \tag{3}$$
$$reg\_feat = \sigma((Wreg * x) + breg) \tag{4}$$

where x is the input feature map, *Wcls* and *Wreg* are the weights of the convolutional layers for classification and regression, respectively. bcls and breg are the convolutional layers biases for classification and regression. Here, σ is the activation function (LeakyReLU) and * denotes the convolution operation.

The modified NanoDet Head maintains the fundamental architecture of the initial model, consisting of a Convolutional Module (ConvModule) followed by two distinct convolutional layers (conv_cls and conv_reg) for classification and regression tasks, respectively. However, we present the subsequent enhancements:

1) **Reduced number of channels:** We have decreased the number of channels in the ConvModule from 256 to 128, reducing computational costs while maintaining essential features.

2) **Depthwise separable convolutions:** We replace the standard convolutional layers with depthwise separable convolutions, which factorize standard convolutions into depthwise and pointwise convolutions, reducing the number of parameters and computations.

The optimized convolutional layer can be represented as:

$$Y = Conv(F, K, P, C, G) = \delta(\Sigma(Gi * F)) \tag{5}$$

Let F be the input feature map, K be the kernel size, P be the padding, C be the number of channels, and G be the number of groups. Y is the output feature map, Gi is the group convolutional kernel, δ is the activation function (ReLU), and Σ denotes the summation over the groups.

The model significantly reduces computational time by introducing these optimizations, making NanoDet more suitable for real-time object detection applications while maintaining accuracy.

The pseudo-code of the proposed Lightweight-CancerNet algorithm is listed in Algorithm 1.

Algorithm 1 Lightweight-CancerNet (Proposed).

Input:

   1. D = $\{(\mathbf{X}i, \mathbf{y}i)\}_{i=1}^{N}$, where $\mathbf{X}i \in \mathrm{R}^{w \times h \times 3}$ is the ith input image and $\mathbf{y}i \in \{0, 1\}$ is the corresponding label

   A, **Steps:**

  2. **Feature Extraction**

     • $F \leftarrow MobileNet(D)$, Extract features from the annotated images in D using MobileNet to form F

  3. **Path Aggregation Network (PAN)**

     • $F' \leftarrow PAN(F)$, Apply the Path Aggregation Network (PAN) to the features F to form F'

  4. **NanoDet Model**

     • $M \leftarrow NanoDet(F', optimized = True)$, Train the NanoDet model with the optimized parameter settings using the features F' and the annotated dataset D^* to form the trained model M

  5. **Training**

     • $M \leftarrow Train(M, D^*, hyperparameters)$, Train the model M with the hyper parameters and the annotated dataset D^*

Output:

M, a trained Lightweight-CancerNet model

EXPERIMENTAL SETUP AND RESULTS

The experiments utilize the Pytorch deep learning framework, built in Python. The algorithm training environment is a Windows 10 operating system, running on an Intel(R) Core (TM) i5-1235U@ 2.64 GHz and an NVIDIA UHD Graphics 630 Ti 08 GB. Based on the 7:2:1 ratio, the dataset is randomly partitioned into the training, test, and validation sets. The training begins with an initial learning rate of 0.11 and is set at 0.14 after some attempts. Before training, warm-up training is conducted with a step size of 5. Simultaneously, one-dimensional linear interpolation (*Hung & Hung, 2014*) is employed. The cosine annealing algorithm (*Zhang et al., 2023*) adjusts the learning rate dynamically during the training process.

Moreover, a training step count of 300 and a batch size of 32 based on the specific computer configuration were employed for the training process. The model's input image size during training is 320 × 320. The detection model's robustness to object size is enhanced to a certain degree by training on photos of varying scales. It took 7 h for training. Table 1 exhibits the training parameters used in this study.

Figure 4 depicts the accuracy and learning curve of the model throughout 300 epochs. It exhibits an initial drop in the first 20 epochs. It is possibly due to initialization. It then shows an increase throughout the training process, which achieved an accuracy of 98%.

Evaluation parameters

We employed the precision and recall rates as fundamental metrics regarding recognition accuracy. The precision rate assesses the correctness of the prediction, while the recall rate examines the completeness of the search. Equations (6) to (9) calculate the precision rate, recall rate, accuracy, and F1 score, respectively. The mAP is subsequently computed as the

Table 1 Training parameters for Lightweight-CancerNet model. Detailed parameter settings used during the training phase of the Lightweight-CancerNet model, including batch size, learning rate, and number of epochs, offering insight into the experimental setup for reproducibility.

Framework parameter	Value
Epochs	300
Batch size	32
Learning rate	0.14
Image size	320 × 320
Optimization function	AdamW
Loss function	Huber loss (IOU loss)
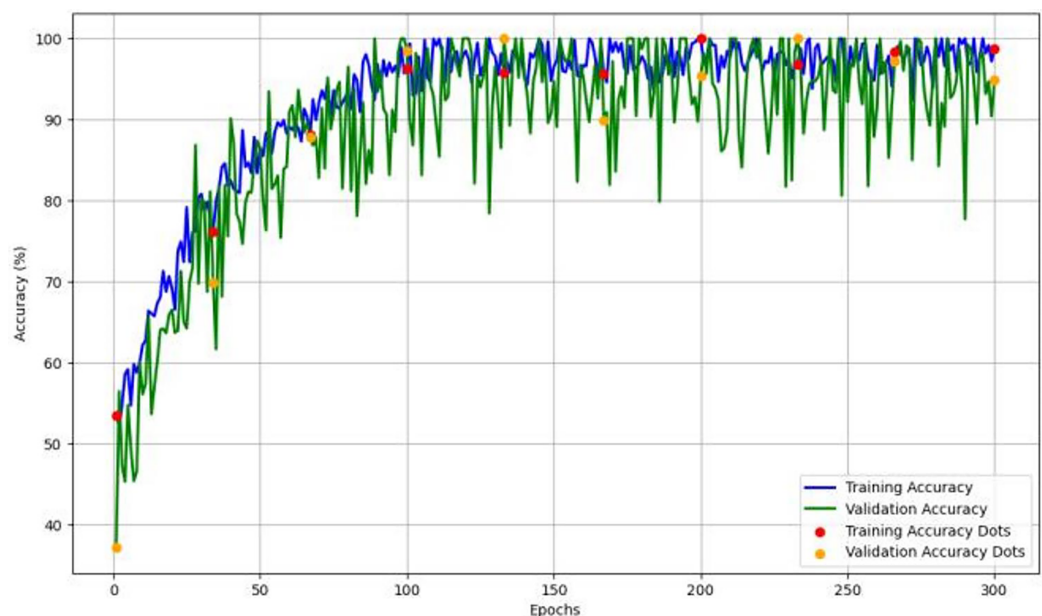
Figure 4 The model's training and validation accuracy over 300 epochs, highlighting trends that indicate the model's learning stability and generalization capability across epochs.
Full-size ⤓ DOI: 10.7717/peerj-cs.2670/fig-4

final evaluation index of accuracy, relying on the precision rate and recall rate. The mAP metric is employed to assess the overall performance of the trained model across all categories. The threshold for determining detection success is set at 0.5, determined by the intersection and union ratio.

$$Precision = \quad TP/(TP + FP) \tag{6}$$
$$Recall = TP/(TP + FN) \tag{7}$$
$$Accuracy = \quad (TP + TN)/(TP + TN + FP + FN) \tag{8}$$
$$F1 = (2 \times TP)/(2 \times TP + FN + FP) \tag{9}$$

In the formula, true positive (TP) denotes the total of predicted positive and actual positive samples, false negative (FN) denotes the total of predicted negative and actual

positive samples, true negative (TN) refers to the total number of instances that were correctly predicted as negative, and false positive (FP) denotes the total of projected positive and actual negative samples.

The equation for calculating the mAP is presented in Eq. (10), where AP represents the average accuracy of each group, t refers to the analyzed image, and T represents the total number of test photographs.

$$mAP := \sum_{i=0}^{T} AP(t)/T \tag{10}$$

Detection results on BraTS dataset

The proposed framework is compared with other state-of-the-art models on the BarTS dataset. Table 2 shows that the proposed Lightweight-CancerNet model demonstrates higher performance than state-of-the-art models on the BraTS dataset, obtaining an accuracy of 97.3% on BraTS. Significantly, this surpasses the 93.4% accuracy attained by U-Net, the 91.2% accuracy of MobileNet, the 92.5% accuracy of ShuffleNet, and the 94.1% accuracy of SqueezeNet. The substantial enhancement in precision showcases the effectiveness of the proposed model on the BraTS dataset. Figure 5 illustrates the detection results of BraTS dataset.

Comparison of the lightweight CancerNet with other state-of-the-art the art deep learning models

An accurate and nominative set of features is required for effective object detection. Therefore, we evaluated the results of Lightweight-Cancernet with other deep learning feature extractor frameworks including YoloV5 (*Shelatkar et al., 2022*), MovileNet (*Ullah et al., 2022*), SSD (*Cen et al., 2019*), U-Net (*Ronneberger, Fischer & Brox, 2015*), ShuffleNet (*Zhang et al., 2017*), SqueezNet (*Iandola et al., 2016*), DeepMedic (*Kamnitsas et al., 2016*), 3D-UNet (*Çiçek et al., 2016*) and VGG-16 (*Simonyan & Zisserman, 2015*). The results of the merged dataset are used for comparison to enhance the generalizability and robustness. The proposed model was inference by using the same conditions and hardware, *i.e.*, used for the training. Figure 6 represents cancer detection results using the proposed framework Lightweight-Cancernet framework. Moreover, Table 3 presents the analyses of various object detection models, including their respective accuracy, mAP ratings, and time taken to detect a single image.

YOLO earns an accuracy of 88% and takes 0.05 s, demonstrating its efficacy in detecting objects in real time. ShuffleNet, recognized for its effectiveness on mobile devices, attains a marginally superior accuracy of 92.5%. The MobileNet, with one instance achieving an accuracy of 82.61% and another earning a far higher score of 91.2%. This highlights how specific implementation details or dataset characteristics influence the performance. The U-Net demonstrated an accuracy of 93.4%, suggesting there may be compromises between accuracy and speed.

Table 2 The proposed Lightweight-CancerNet framework's performance with other state-of-the-art models, highlighting accuracy, to illustrate its competitive efficacy.

Model	Dataset	Accuracy
U-Net	BraTS	93.4%
MobileNet	BraTS	91.2%
ShuffleNet	BraTS	92.5%
SqueezeNet	BraTS	94.1%
Proposed	BraTS	97.7%
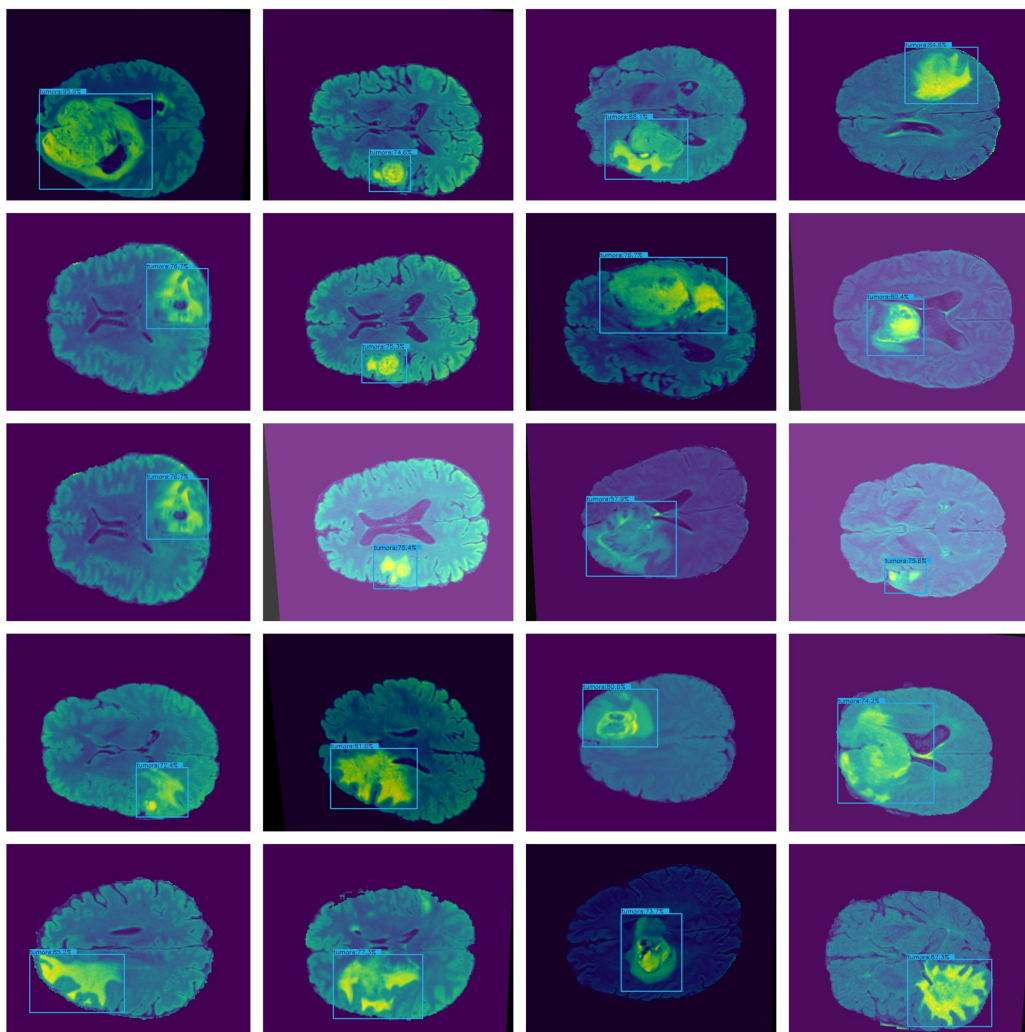
Figure 5 Lightweight-CancerNet framework on BraTS. Full-size 🖼 DOI: 10.7717/peerj-cs.2670/fig-5

Nevertheless, the Lightweight-CancerNet stands out as the top performer in the chart, presenting an amazing mAP of 93.8%, indicating notable progress in precision and effectiveness, resolving the compromises observed in earlier models.
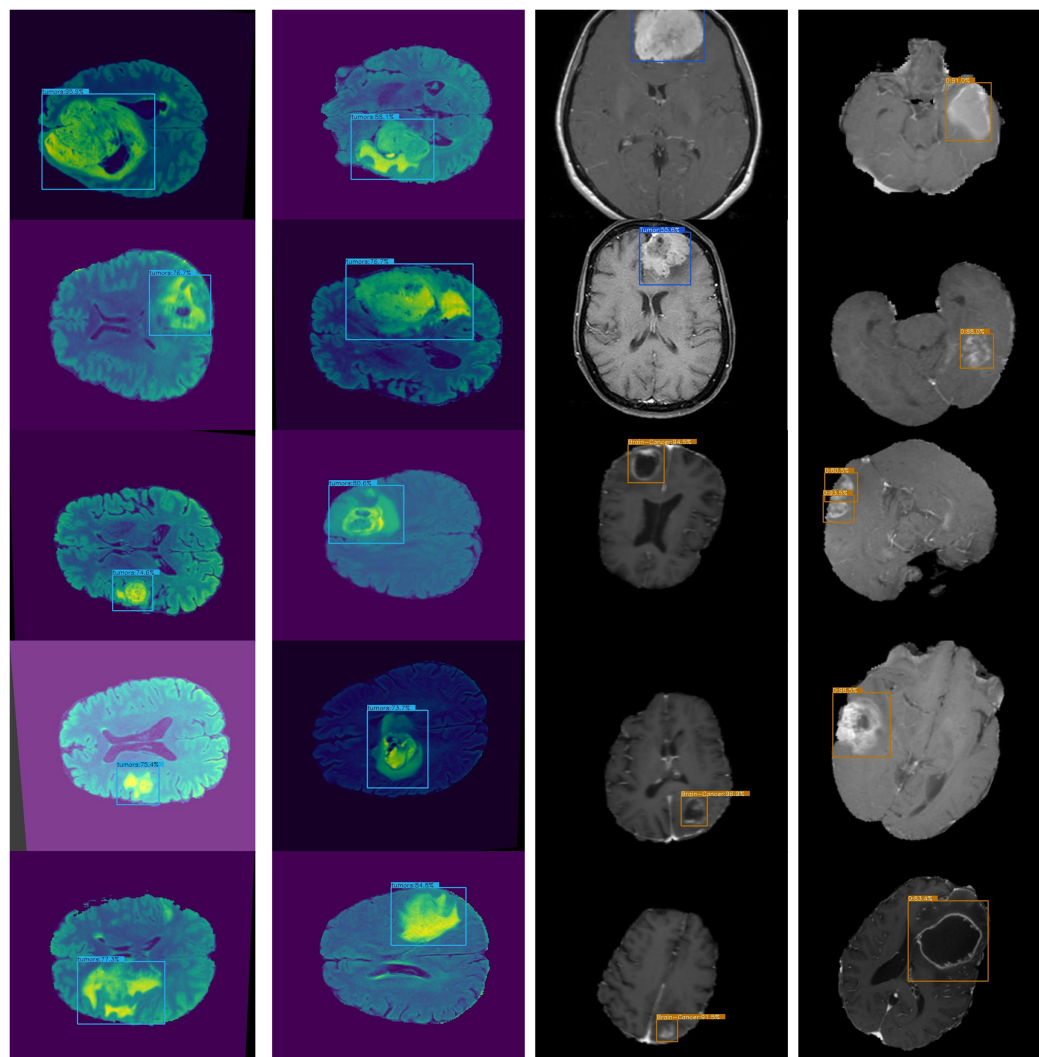
Figure 6 Presents sample outputs showing successful detection of cancerous regions.
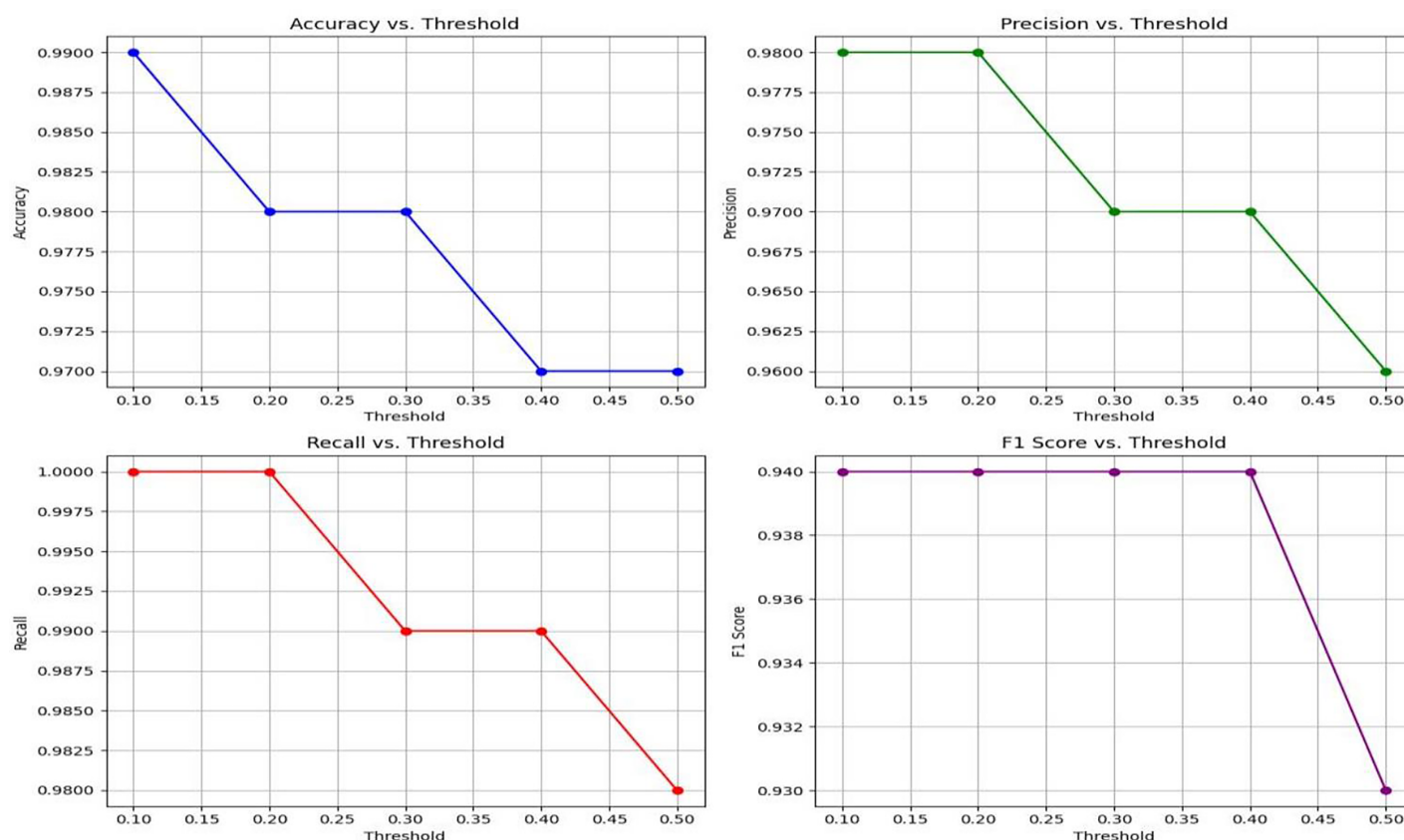Full-size 🖼 DOI: 10.7717/peerj-cs.2670/fig-6

Moreover, the experimental results also demonstrate the exceptional effectiveness of the proposed Lightweight-CancerNet model for segmenting brain tumors. The suggested model surpasses previous state-of-the-art models, such as YoloV5, MobileNet, SSD, U-Net, ShuffleNet, SqueezNet, DeepMedic, 3D-UNet, and VGG-16, with an accuracy of 98% and an mAP of 93.8%. The suggested model attains remarkable accuracy and mAP while using considerably less computational effort, averaging around 0.003 s per image. This model is far more efficient than previous models, such as U-Net, which has a processing time of 3.33 s per image. Figure 7 represents a graphical view of performance evaluation concerning different threshold values.

Lightweight-CancerNet performance: threshold analysis

This section presents a comprehensive evaluation of the performance metrics of Lightweight-CancerNet for brain tumor detection. The results illustrate the method's

Table 3 Detailed performance comparison of Lightweight-CancerNet with alternative models.
Expands on performance metrics across different state-of-the-art models, presenting accuracy, and computational efficiency to assess Lightweight-CancerNet's suitability for efficient and accurate cancer detection.

| Method	MAP	Accuracy	Inference time (S) per image
YoloV5	–	88%	0.05
MobileNet	–	82.61%	0.007
SSD	77.2%	82.7%	0.016
U-Net	–	93.4%	3.33
MobileNet	–	91.2%	0.01
ShuffleNet	–	92.5%	0.0074
SqueezNet	–	94.1%	0.0071
DeepMedic	74.4%	–	–
3D-UNet	75.1%	–	–
VGG-16	71.8%	–	–
Lightweight-CancerNet	93.8%	98%	0.003

Figure 7 A visual representation of how different classification thresholds affect performance metrics.
Full-size 🖾 DOI: 10.7717/peerj-cs.2670/fig-7

Table 4 **Performance metrics for brain tumor segmentation at various thresholds.** Presents segmentation performance across multiple thresholds, showing variations in accuracy, recall, precision and F1 score to assess the model's robustness in different scenarios.

Threshold	Accuracy	Recall	Precision	F1 score
0.1	0.99	1.00	0.98	0.94
0.2	0.98	1.00	0.98	0.94
0.3	0.97	0.99	0.97	0.94
0.4	0.97	0.99	0.97	0.94
0.5	0.96	0.98	0.96	0.93
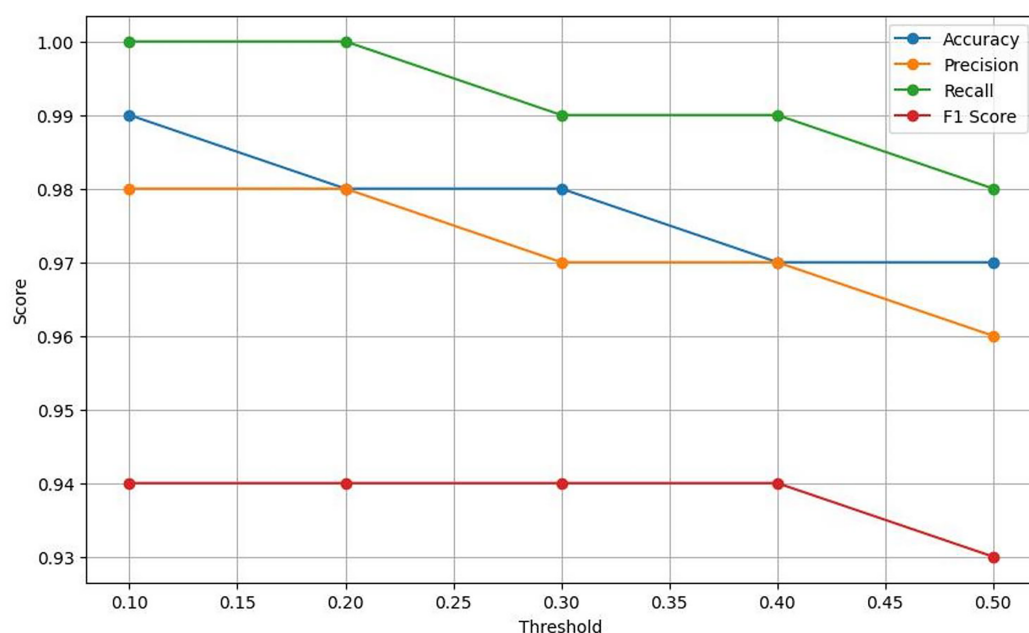
Figure 8 **Plots the four key performance metrics against threshold values, helping visualize trade-offs between metrics and guiding threshold selection for the most balanced performance.**
Full-size 🖼 DOI: 10.7717/peerj-cs.2670/fig-8

reliability in detecting brain tumors, considering the balance between accuracy and the choice of threshold (*Kumar, 2018*). Here the threshold is the value or boundary that used to determine that the input will be categorized in positive or negative with respect to defined value.

Table 4 exhibits that Lightweight-CancerNet attains a high level of accuracy (between 97% and 99%) consistently over a range of thresholds, proving that it is a reliable method for identifying brain cancers. Additionally, there is a trade-off between precision and threshold, beginning at a high of 98% for a 0.1 threshold and dropping to 96% for a 0.5 threshold. This indicates that CancerNet is better at detecting cancers at low levels, although it may miss some crucial instances due to this. Meanwhile, recall remains high against all threshold values, decreasing between 98% and 1.0. Therefore, CancerNet can

find all pertinent brain tumor instances. A plot with precision, recall, accuracy, and F1 score curves plotted against thresholds is depicted in Fig. 8.

The F1 score shows a slight drop as the threshold increases. Starting at a threshold of 0.1, it approaches 94%; at 0.5, it hits 93%. CancerNet has great potential in accuracy and recall, implying that it can accurately identify and capture many brain cancers. Setting the threshold correctly requires examining the context, as there is a trade-off between accuracy and threshold. If the aim is to minimize false positives, then reducing the threshold would be the most effective approach. However, opting for a more significant threshold may be the best option if the main objective is to detect all potential cancers.

CONCLUSION

This article presents Lightweight-CancerNet, a new deep-learning framework that provides excellent accuracy and efficiency in the diagnosis of brain tumor. Our model surpasses existing models in terms of accuracy and efficiency, utilizing the capabilities of MobileNet and NanoDet. The proposed Lightweight-CancerNet is designed to manage extensive medical imaging datasets using minimal computer resources, making it highly suitable for instantaneous processing. This study has significant implications for improving patient outcomes and decision-making in brain surgery. It also illustrates the potential of using lightweight models to detect objects in medical imaging quickly.

The proposed model's performance heavily depends on the training dataset's diversity and quality. Biased or limited datasets may result in suboptimal performance. Another limitation of the model is that it may not generalize well to unseen data and require more parameter training.

In our future research, we will explore more improvements and extensions to our framework and apply it to other medical imaging applications. Furthermore, we intend to evaluate the application of our system in real-world clinical settings by examining its effectiveness in practical scenarios and its potential to improve patient outcomes.

ACKNOWLEDGEMENTS

ADDITIONAL INFORMATION AND DECLARATIONS

Competing Interests
The authors declare that they have no competing interests.

Author Contributions
- Asif Raza conceived and designed the experiments, performed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.

- Muhammad Javed Iqbal conceived and designed the experiments, analyzed the data, authored or reviewed drafts of the article, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The BraTS (Brain Tumor Segmentation) dataset is a widely used benchmark in the field of medical image analysis, particularly for brain tumor detection and segmentation tasks. It was introduced by the Medical Image Computing and Computer-Assisted Intervention (MICCAI) community and is updated regularly to reflect advancements in brain imaging and tumor segmentation techniques. It is available at https://www.med.upenn.edu/cbica/brats2020/data.html. The BraTS dataset is freely accessible for research purposes through the MICCAI official site, with a license for academic use only.

Access to the BraTS dataset can be requested here https://www.med.upenn.edu/cbica/brats2020/registration.html and questions directed to brats2020@cbica.upenn.edu

The code is available in the Supplemental Files.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj-cs.2670#supplemental-information.

REFERENCES

Aamir M, Rahman Z, Dayo ZA, Abro WA, Uddin MI, Khan I, Imran AS, Ali Z, Ishfaq M, Guan Y, Hu Z, Hu Z. 2022. A deep learning approach for brain tumor classification using MRI images. *Computers and Electrical Engineering* **101**:108105 DOI 10.1016/j.compeleceng.2022.108105.

Abdusalomov AB, Mukhiddinov M, Whangbo TK. 2023. Brain tumor detection based on deep learning approaches and magnetic resonance imaging. *Cancers* **15(16)**:4172 DOI 10.3390/cancers15164172.

Aggarwal M, Tiwari AK, Sarathi MP, Bijalwan A. 2023. An early detection and segmentation of Brain Tumor using Deep Neural Network. *BMC Medical Informatics and Decision Making* **23**:78 DOI 10.1186/s12911-023-02174-8.

Aleid A, Alhussaini K, Alanazi R, Altwaimi M, Altwijri O, Saad AS. 2023. Artificial intelligence approach for early detection of brain tumors using MRI images. *Applied Sciences* **13(6)**:3803 DOI 10.3390/app13063808.

Ardan IS, Indraswari R. 2024. Design of brain tumor detection system on MRI image using CNN. In: *2024 ASU International Conference in Emerging Technologies for Sustainability and Intelligent Systems (ICETSIS)*. Piscataway: IEEE, 1388–1393.

Asiri AA, Soomro TA, Shah AA, Pogrebna G. 2024. Optimized brain tumor detection: a dual-module approach for MRI image enhancement and tumor classification. *IEEE Access* **12**:42868–42887 DOI 10.1109/ACCESS.2024.3379136.

Bakas S. 2018. Multimodal brain tumor segmentation Benchmark. In: *Center for Biomedical Image Computing & Analytics*.

Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, Freymann JB, Farahani K, Davatzikos C. 2017. Advancing the cancer genome atlas glioma MRI collections with expert

segmentation labels and radiomic features. *Scientific Data* **4**:170117 DOI 10.1038/sdata.2017.117.

Bakas S, Reyes M, Jakab A, Bauer S, Rempfler M, Crimi A, Shinohara RT, Berger C, Ha SM, Rozycki M, Prastawa M, Alberts E, Lipkova J, Freymann J, Kirby J, Bilello M, Fathallah-Shaykh H, Wiest R, Kirschke J, Wiestler B, Colen R, Kotrotsou A, Lamontagne P, Marcus D, Milchenko M, Nazeri A, Weber MA, Mahajan A, Baid U, Gerstner E, Kwon D, Acharya G, Agarwal M, Alam M, Albiol A, Albiol A, Albiol FJ, Alex V, Allinson N, Amorim PHA, Amrutkar A, Anand G, Andermatt S, Arbel T, Arbelaez P, Avery A, Azmat M, B P, Bai W, Banerjee S, Barth B, Batchelder T, Batmanghelich K, Battistella E, Beers A, Belyaev M, Bendszus M, Benson E, Bernal J, Bharath HN, Biros G, Bisdas S, Brown J, Cabezas M, Cao S, Cardoso JM, Carver EN, Casamitjana A, Castillo LS, Catà M, Cattin P, Cerigues A, Chagas VS, Chandra S, Chang YJ, Chang S, Chang K, Chazalon J, Chen S, Chen W, Chen JW, Chen Z, Cheng K, Choudhury AR, Chylla R, Clérigues A, Colleman S, Colmeiro RGR, Combalia M, Costa A, Cui X, Dai Z, Dai L, Daza LA, Deutsch E, Ding C, Dong C, Dong S, Dudzik W, Eaton-Rosen Z, Egan G, Escudero G, Estienne T, Everson R, Fabrizio J, Fan Y, Fang L, Feng X, Ferrante E, Fidon L, Fischer M, French AP, Fridman N, Fu H, Fuentes D, Gao Y, Gates E, Gering D, Gholami A, Gierke W, Glocker B, Gong M, González-Villá S, Grosges T, Guan Y, Guo S, Gupta S, Han WS, Han S II, Harmuth K, He H, Hernández-Sabaté A, Herrmann E, Himthani N, Hsu W, Hsu C, Hu X, Hu X, Hu Y, Hu Y, Hua R, Huang TY, Huang W, Huffel SV, Huo Q, HV V, Iftekharuddin KM, Isensee F, Islam M, Jackson AS, Jambawalikar SR, Jesson A, Jian W, Jin P, Jose VJM, Jungo A, Kainz B, Kamnitsas K, Kao PY, Karnawat A, Kellermeier T, Kermi A, Keutzer K, Khadir MT, Khened M, Kickingereder P, Kim G, King N, Knapp H, Knecht U, Kohli L, Kong D, Kong X, Koppers S, Kori A, Krishnamurthi G, Krivov E, Kumar P, Kushibar K, Lachinov D, Lambrou T, Lee J, Lee C, Lee Y, Lee M, Lefkovits S, Lefkovits L, Levitt J, Li T, Li H, Li W, Li H, Li X, Li Y, Li H, Li Z, Li X, Li Z, Li XG, Li W, et al. 2019. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge. ArXiv DOI 10.48550/arXiv.1811.02629.

BraTS. 2020. *Multimodal brain tumor segmentation challenge 2020: data. Available at* https://www.med.upenn.edu/cbica/brats2020/data.html.

Cen Q, Pan Z, Li Y, Ding H. 2019. Laryngeal tumor detection in endoscopic images based on convolutional neural network. In: *IEEE 2nd International Conference on Electronic Information and Communication Technology.* Piscataway: IEEE.

Chengwei. 2019. How to run SSD Mobilenet V2 object detection on Jetson Nano at 20+ FPS. DLology. *Available at* https://www.dlology.com/blog/how-to-run-ssd-mobilenet-v2-object-detection-on-jetson-nano-at-20-fps/.

Chinaev A, Spitz T, Thaleiser S, Enzner G. 2024. Matrix study of feature compression types and instrumental speech quality metrics in ultra-light DNN-based spectral speech enhancement. In: *2024 18th International Workshop on Acoustic Signal Enhancement (IWAENC)*, 11–15.

Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 2016. 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention.* Cham: Springer, 424–432.

DeAngelis LM. 2001. Brain tumors. *The New England Journal of Medicine* **344**:114–123 DOI 10.1056/NEJM20010111344020.

Gao Y, Zhang ZD, Li S, Guo YT, Wu QY, Liu SH, Yang SJ, Ding L, Zhao BC, Li S, Lu Y. 2019. Deep neural network-assisted computed tomography diagnosis of metastatic lymph nodes from gastric cancer. *Chinese Medical Journal (Engl)* **132(23)**:2804–2811 DOI 10.1097/CM9.0000000000000532.

Ghafourian E, Samadifam F, Fadavian H, Canatalay PJ, Tajally A, Channumsin S. 2023. An ensemble model for the diagnosis of brain tumors through MRIs. *Diagnostics* **13(3)**:561 DOI 10.3390/diagnostics13030561.

Gupta BB, Gaurav A, Arya V. 2024. Deep CNN based brain tumor detection in intelligent systems. *International Journal of Intelligent Networks* 30–37.

Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, Andreetto M, Adam H. 2017. MobileNets: efficient convolutional neural networks for mobile vision applications. ArXiv preprint DOI 10.48550/arXiv.1704.04861.

Hung BL, Hung HA. 2014. Using one-dimensional linear interpolation method to check over-fitting in neural network with multi-dimensional inputs. *Applied Mechanics and Materials* **496–500**:2228–2232 DOI 10.4028/www.scientific.net/AMM.496-500.2228.

Iandola FN, Han S, Moskewicz MW, Ashraf K, Dally WJ, Keutzer K. 2016. SqueezeNet: a simple and efficient CNN for image classification. In: *IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE, 3511–3519.

Kale PV, Gadicha AB. 2024. Deep CNN-based MRI imaging for brain tumor detection and classification. In: *2024 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)*. Piscataway: IEEE, 1–6.

Kamnitsas K, Ferrante E, Parisot S, Ledig C, Nori AV, Criminisi A, Rueckert D, Glocker B. 2016. DeepMedic for brain tumor segmentation. In: Menze B, Reyes M, Crimi A, Maier O, Winzeck S, Handels H, eds. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Cham: Springer Verlag, 138–149.

Kleihues P, Louis DN, Scheithauer BW, Rorke LB, Reifenberger G, Burger PC, Cavenee WK. 2002. The WHO classification of tumors of the nervous system. *Journal of Neuropathology & Experimental Neurology* **61(3)**:215–225; discussion 226–229 DOI 10.1093/jnen/61.3.215.

Kumar A. 2023. Study and analysis of different segmentation methods for brain tumor MRI application. *Multimedia Tools and Applications* **82(5)**:7117–7139 DOI 10.1007/s11042-022-13636-.

Kumar N. 2018. Thresholding in salient object detection: a survey. *Multimedia Tools and Applications* **77**:19139–19170 DOI 10.1007/s11042-017-5329-y.

Lee J-H, Chae J-W, Cho H-C. 2024. Improved classification of different brain tumors in MRI scans using patterned-GridMask. *IEEE Access* **12**:40204–40212 DOI 10.1109/ACCESS.2024.3377105.

Mahmud MI, Mamun M, Abdelgawad A. 2023. A deep analysis of brain tumor detection from MR images using deep learning networks. *Algorithms* **16(4)**:176 DOI 10.3390/a16040176.

Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, Burren Y, Porz N, Slotboom J, Wiest R, Lanczi L, Gerstner E, Weber MA, Arbel T, Avants BB, Ayache N, Buendia P, Collins DL, Cordier N, Corso JJ, Criminisi A, Das T, Delingette H, Demiralp Ç, Durst CR, Dojat M, Doyle S, Festa J, Forbes F, Geremia E, Glocker B, Golland P, Guo X, Hamamci A, Iftekharuddin KM, Jena R, John NM, Konukoglu E, Lashkari D, Mariz JA, Meier R, Pereira S, Precup D, Price SJ, Raviv TR, Reza SMS, Ryan M, Sarikaya D, Schwartz L, Shin H-C, Shotton J, Silva CA, Sousa N, Subbanna NK, Szekely G, Taylor TJ, Thomas OM, Tustison NJ, Unal G, Vasseur F, Wintermark M, Ye DH, Zhao L, Zhao B, Zikic D, Prastawa M, Reyes M, Van Leemput K. 2015. The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Transactions on Medical Imaging* **34(10)**:1993–2024 DOI 10.1109/TMI.2014.2377694.

Mohan P, Easwaramoorthy SV, Subramani N, Subramanian M, Meckanzi S. 2022. Handcrafted deep-feature-based brain tumor detection and classification using MRI images. *Electronics* **11(24)**:4178 DOI 10.3390/electronics11244178.

National Cancer Institute. 2024. *Brain tumors—health professional version*. Retrieved from brain tumors—health professional version. *Available at* https://www.cancer.gov/types/brain/hp.

Nawaz M, Nazir T, Masood M, Mehmood A, Mahum R, Khan MA, Kadry S, Thinnukool O. 2021. Analysis of brain MRI images using improved CornerNet approach. *Diagnostics* **11(10)**:1856 DOI 10.3390/diagnostics11101856.

Popat M, Patel S. 2022. Research perspective and review towards brain tumour segmentation and classification using different image modalities. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization* **11(5)**:1579–1597 DOI 10.1080/21681163.2022.2124546.

Qiang B, Zhai Y, Zhou M, Yang X, Peng B, Wang Y. 2021. SqueezeNet and fusion network-based accurate fast fully convolutional network for hand detection and gesture recognition. *IEEE Access* **9**:77661–77674 DOI 10.1109/ACCESS.2021.3079337.

Qureshi MA, Qureshi KN, Jeon G, Piccialli F. 2022a. Deep learning-based ambient assisted living for self-management of cardiovascular conditions. *Neural Computing and Applications* **34**:10449–10467 DOI 10.1007/s00521-020-05678-w.

Qureshi SA, Raza SEA, Hussain L, Malibari AA, Nour MK, Rehman AU, Al-Wesabi FN, Hilal AM. 2022b. Intelligent ultra-light deep learning model for multi-class brain tumor detection. *Applied Sciences* **12(8)**:3715 DOI 10.3390/app12083715.

RangiLyu. 2021. NanoDet-Plus: super fast and high accuracy lightweight anchor-free object detection model. *Available at* https://github.com/RangiLyu/nanodet.

Roberts D. 2021. Brain tumor object detection datasets. Retrieved 08 2023, from brain tumor object detection datasets. *Available at* https://www.kaggle.com/datasets/davidbroberts/brain-tumor-object-detection-datasets/data.

Ronneberger O, Fischer P, Brox T. 2015. U-Net: convolutional networks for biomedical image segmentation. In: *Medical Image Computing and Computer-Assisted Intervention*, 234–241.

Shelatkar T, Urvashi D, Shorfuzzaman M, Alsufyani A, Lakshmanna K. 2022. Diagnosis of brain tumor using light weight deep learning model with fine-tuning approach. *Computational and Mathematical Methods in Medicine* **2022**:1–9 DOI 10.1155/2022/2858845.

Simo AM, Kouanou AT, Monthe V, Nana MK, Lonla BM. 2024. Introducing a deep learning method for brain tumor classification using MRI data towards better performance. *Informatics in Medicine Unlocked* **44(16)**:101423 DOI 10.1016/j.imu.2023.101423.

Simonyan K, Zisserman A. 2015. Very deep convolutional networks for large-scale image recognition. In: *Computer Vision and Pattern Recognition* ArXiv preprint DOI 10.48550/arXiv.1409.1556.

Tinu MA, Appathurai A, Muthukumaran N. 2024. Detection of brain tumour via reversing hexagonal feature pattern for classifying double-modal brain images. *IETE Journal of Research* 1–12.

Ullah N, Khan JA, Khan MS, Khan W, Hassan I, Obayya M, Negm N, Salama AS. 2022. An effective approach to detect and identify brain tumors using transfer learning. *Applied Sciences* **12(11)**:5645 DOI 10.3390/app12115645.

Vinod DS, Prakash SP, Salman HA. 2024. Ensemble technique for brain tumor patient survival prediction. *IEEE Access* **12**:19285–19298 DOI 10.1109/ACCESS.2024.3360086.

Zhang C, Shao Y, Sun H, Xing L, Zhao Q, Zhang L. 2023. The WuC-Adam algorithm based on joint improvement of Warmup andcosine annealing algorithms. *Mathmatical Bioscience and Engineering* **21(1)**:1270 DOI 10.3934/mbe.2024054.

Zhang X, Zhou X, Lin M, Sun J. 2017. ShuffleNet: an extremely efficient convolutional neural network for mobile devices. In: *Computer Vision and Pattern Recognition.* Piscataway: IEEE DOI 10.1109/CVPR.2018.00716.

Zhou Z, Siddiquee MM, Tajbakhsh N, Liang J. 2018. UNet++: a nested U-Net architecture for medical image segmentation. In: *Computer Vision and Pattern Recognition* ArXiv preprint DOI 10.48550/arXiv.1807.10165.

Raza and Iqbal (2025), *PeerJ Comput. Sci.,* **DOI 10.7717/peerj-cs.2670**

22/22