



ISSN: 0067-2904

Enhanced Brain Tumor Classification Using EfficientNetB0 and Hybrid Machine Learning Approaches

Abeer A. Alshiha, Abdulrahman R. Qubaa*, Rayan Ghazi Thannoun

Remote Sensing Center, University of Mosul, Mosul, Iraq.

Received: 23/9/2024

Accepted: 23/ 7/2025

Published: 30/ 7/2026

Abstract

The classification of brain tumors has remained a strong area of research in the medical field, and this has been caused by the high number of cases of brain tumors reported all over the world. The incidence rate of brain tumors in India in 2023 was found to be more than 38,000 new cases. This is one of the widespread cancers, with fewer than 34,000 deaths occurring. The opposing issues in the classification of brain tumors using machine learning (ML) include the small scale of the data, feature selection, and significant classification errors. Although deep learning (DL) methods are continually evolving, this research will provide solutions using a hybrid scheme that combines mixed ML and DL methods. The models examined in this paper are: Convolutional Neural Networks (CNN), CNN with Support Vector Machine (SVM), Random Forest (RF), Logistic Regression (LR), and EfficientNetB0. Four types of measures were taken to establish the accuracy of the results and to compare the five types of models applied in the study. The outcome indicated that EfficientNetB0 was the best model, with an accuracy of 99.80%, a precision of 99.90%, a recall of 99.80%, and an F1-score of 99.80%. The CNN, CNN-SVM, CNN-RF, and CNN-LR models, on the other hand, provided accuracies in the range of 89% to 91%. This outcome demonstrates that EfficientNetB0 outperforms other models in detecting brain tumors, as well as its ability to enhance diagnostic capabilities in clinical settings. The paper presented the performance of each tumour class (Glioma, Meningioma, Pituitary Tumor, No Tumor) in each model. It also employed statistical validation methods, including cross-validation and significance testing, to support the results.

Keywords: Classification of brain tumors; machine learning; deep learning; EfficientNetB0; MRI.

تصنيف محسن لأورام الدماغ باستخدام EfficientNetB0 وأساليب التعلم الآلي الهجينة

عبير عادل محمد الشبيحة , *عبدالرحمن رمزي قبيع , ريان غازي ذنون

مركز التحسس النائي, جامعة الموصل, نينوى, العراق

الخلاصة

لا يزال تصنيف أورام المخ مجالاً بحثياً قوياً في مجال الطب، ويعود ذلك إلى العدد الكبير من حالات أورام المخ المبلغ عنها في جميع أنحاء العالم. وقد وُجد أن معدل الإصابة بأورام المخ في الهند في عام 2023 تجاوز 38000 حالة جديدة. ويُعد هذا أحد أنواع السرطان المنتشرة مع حدوث أقل من 34000 حالة

*Email: abdqubaa@uomosul.edu.iq



وفاة. تشمل القضايا المتعارضة في تصنيف أورام المخ من خلال التعلم الآلي ML صغر حجم البيانات واختيار الميزات والأخطاء الكبيرة في التصنيف. وعلى الرغم من أن أساليب التعلم العميق DL قيد التطوير المستمر، إلا أن البحث سيوفر حلولاً باستخدام مخطط هجين يستخدم مزيجاً من أساليب التعلم الآلي والتعلم العميق المختلطة. النماذج التي تمت دراستها في هذه الورقة هي: الشبكات العصبية التلافيفية CNN، وشبكة CNN مع آلة المتجهات الداعمة SVM، والغابات العشوائية RF، والانحدار اللوجستي LR، وشبكة EfficientNetB0. وقد تم إجراء أربعة أنواع من المقاييس لتحديد دقة النتائج بالإضافة إلى مقارنة الأنواع الخمسة من النماذج المطبقة في الدراسة. أشارت النتيجة إلى أن نموذج EfficientNetB0 كان الأفضل بين جميع النماذج بدقة 99.80%، وجودة 99.90% واستدعاء 99.80% ودرجة F1-Score 99.80% من ناحية أخرى وفرت نماذج CNN و CNN-SVM و CNN-RF و CNN-LR دقة تتراوح بين 89 إلى 91% . توضح النتائج أن نموذج EfficientNetB0 يتميز بأداء أقوى من النماذج الأخرى في الكشف عن أورام المخ، وقدرته على زيادة إمكانيات التشخيص في البيانات السريرية. قدّم البحث أداء كل فئة من فئات الأورام (الورم الدبقي، والورم السحائي، وورم الغدة النخامية، وعدم وجود ورم) في كل نموذج. كما استخدم أساليب التحقق الإحصائي، مثل التحقق المتبادل، واختبار الدلالة، وغيرها، لدعم النتائج.

1. Introduction

Based on the IARC reports, in 2023, brain tumors were the tenth most common cancer in India, with more than 38,000 incidences every year, and the estimated number of deaths associated with the condition was 34,000. Magnetic Resonance Imaging (MRI) plays a crucial role in understanding the human body's anatomy, offering high spatial resolution and contrast of soft tissues with minimal radiation exposure [1]. Brain tumors can also be divided into two types, namely, primary and secondary nerve tumors, and malignant and benign ones. This is more applicable in clinical pathology, although it cannot be directly compared to the classification of MRI images used in the diagnosis of brain tumors. What is truly challenging is distinguishing between these types of tumors and their corresponding MRI images, and this is what this study aims to address. Such a system, being an automated one, which can accurately read tumor images and locate their specific locations, is very essential for early diagnosis. A specific definition of the location and type of the tumor plays a critical role in planning effective treatment procedures, including surgery, radiation therapy, and chemotherapy. The fact that the margins of the tumor can be accurately delineated assists the surgeon in establishing the limits of the surgery to be done and in the removal of the tumor without destroying the healthy brain tissue that surrounds it [2].

Over the past few years, numerous approaches have been proposed for brain tumor classification, which can be categorized into machine learning (ML) and deep learning (DL) techniques [3]. Although deep learning (DL) methods, especially Convolutional Neural Networks (CNNs), are very accurate and widely applied in the field of medical image analysis, they also have their setbacks. DL approaches are more likely to be trained on large datasets, and therefore, this aspect is time- and resource-intensive when large datasets cannot be obtained in the context of interest [4]. Additionally, most DL approaches require expensive computing systems, such as GPUs, which further increase the cost for healthcare organizations [5]. Finally, the randomness of the DL framework, as well as the training parameters, may complicate its application in the clinical setting. On the other hand, the traditional ML approach also has several advantages, including the ability to work with small data sets, reduced training requirements, and more straightforward interpretation. However, the need has been raised to develop new and improved machine learning (ML) models that can be effective in labeling brain tumor cases, particularly in instances where deep learning (DL) models are not applicable or cannot be used [6].

In this paper, the authors aimed to address these problems and developed new machine learning (ML) tools with the same level of classification ability as deep learning (DL) algorithms. The size of the data and the time consumed, as well as training skills, will not influence the techniques. In this way, the research will lead to the better automation of brain tumor classification, overcoming the drawbacks that commonly occur when using DL methods.

2. Related works

In the classification of brain tumors, many studies have been suggested with many approaches on how to do it. Deep learning (DL) and machine learning (ML). Such studies are typically not critically reviewed or compared, and they often show little interest in addressing the gaps present in the literature, which is the gap that this study aims to fill. A review of relevant works on this topic is presented below, with a focus on identifying gaps and constraints in the literature that the current research aims to address. A feasible method for brain tumor classification, presented by Cheng et al. (2015), was introduced. Feature extraction techniques, including grey-level co-occurrence matrix, and the bag-of-words Model, as well as raster intensity histograms and methods. The methodology was 91% precise. The method's effectiveness is worthwhile, despite its success. There was no research conducted on data mining issues related to small datasets and feature selection, which are particularly significant when classifying brain tumors [7].

Cherukuri et al. (2020) introduced the multi-level attention network (MANet) for classifying brain tumors with 96% accuracy on the T1W-CE MRI dataset, in conjunction with the Xception model. Although the technique is quite accurate, it is, however, computationally costly, and this method may not always be applicable in clinics, especially when it involves data based on big data. Furthermore, the study fails to disclose any details about the computational cost of the model [8]. Instead, Guan et al. (2019) enhanced the contrast of the images and employed a nonlinear method for segmenting and classifying brain tumors, achieving an accuracy of 98% through fivefold cross-validation of the T1W-CE MRI dataset. Notwithstanding the method, a heavy dependence on image enhancement remains, and it cannot be efficient enough in cases of poor image quality or scarce data [9]. Bad et al. (2019) proposed a 22-layer convolutional neural network (CNN) for classifying meningioma, glioma, and pituitary tumors with an accuracy of 96%. Although the performance is optimal, the use of deep CNNs raises concerns regarding computational cost and training time, particularly for small datasets. It failed to address the issue of class imbalance, which is a common problem in medical imaging [10]. Deepak et al. (2019) optimized GoogleNet and utilized SVM and KNN as classifiers, achieving accuracies of 92%, 97%, and 98% for T1-weighted contrast-enhanced (T1W-CE) MRI. The strategy is hindered by the fact that it requires large amounts of data to perform well, yet it does not adequately examine the feature selection procedure [11]. Díaz-Pernas and other researchers proposed a three-path multiscale convolutional neural network (CNN) for brain tumor segmentation and classification, achieving an accuracy of 97.3%. Nonetheless, the sensitivity of the method applied to meningioma was low, indicating a problem with low contrast between the tumor and tissues in some instances [12].

A hard-swish-based ReLU CNN, which achieved 98% accuracy, was used in Alhassan et al. (2020). The model has incorporated the concepts of normalization and HOG descriptors but has failed to address the issues of dataset size and computational cost, which should be of great concern when applying the model in a practical field, such as real medical imaging [13]. Ghassemi et al. (2019) utilized a DCGAN for feature extraction, achieving an accuracy of

95.6% in brain tumor classification. Despite generating such effective feature extraction, the operation of DCGANs is computationally unattainable and could prove unfeasible in a low-resource clinical setting [14]. Noreen et al. (2020) employed deep and machine learning hybrids, as well as pre-trained models such as Inception-v3 and Xception, to extract features. They then utilized SVM, KNN, and ensemble classifiers. Inception-v3-Ensemble had reported an accuracy of 94.34% on the T1W-CE MRI dataset. Still, the comparison between the workings of different kinds of classifiers was not provided in the research, and issues related to small datasets and class imbalance were not discussed [15]. Gumaei et al. (2019) proposed a hybrid feature extraction method based on both regularized extreme learning machine (RELM) and PCA-NGIST, achieving an accuracy of 94.23% under five-fold cross-validation. The given approach enhanced feature extraction, but it did not affect the size of the dataset or the computational load, which are the challenges in real-world practice [16]. Haq et al. (2020) proposed the use of a deep convolutional neural network (DCNN) approach for brain tumor screening and classification, employing various preprocessing methods and data augmentation techniques. Nevertheless, the study achieved relatively poor accuracy in testing certain tumors, including meningioma (91%) and glioma (1%), which suggests possible issues with sensitivity and generalizability for various tumor types [17].

The above studies provide valuable details on the categorization of brain tumors; however, some of their shortcomings are also revealed. These include susceptibility to extensive collections, computational expense, and a lack of methods for dealing with coarse inequality. Most studies, particularly those, did not pay sufficient attention to this latter aspect. The proposed study aims to fill these gaps by creating machine learning models with high classification accuracy and applicability, even on small datasets, without requiring costly computing resources. The issue of class imbalance, which was adequately addressed in the majority of previous works, will be discussed in this study. The ongoing study aims to make brain cancer categorization more feasible and practical by developing effective machine learning (ML) models.

3. Methodology

As shown in Figure 1, the methodology adopted for classifying brain tumors follows the principles of deep learning. The first step was to import all necessary libraries and select the appropriate color schemes for data visualizations. The pictures in the dataset are stratified into four classes: no Tumor, Glioma, Meningioma, and Pituitary Tumor. The images were then loaded and resized to some sensible values (e.g., 250) and stored in arrays for training and testing purposes. It was, however, incomplete in the description of the preprocessing steps. In particular, the researcher does not provide detailed information about the preprocessing operations applied to the images prior to training. It is necessary to identify actions such as normalization, dimension change, and margin exclusion, as well as other techniques that can be used to enhance the model's functioning. After loading and resizing the images, they were shuffled to create a random order and preserved as categorical labels. The dataset was then divided into two sets: the training set and the testing set. The percentage of data allocated to training and testing should be specified; for example, 80% of the data used for training and 20% for testing, or any other suitable combination. A model selected for this study is EfficientNetB0, which is fine-tuned using the dataset. Initially, the pre-trained EfficientNetB0 model, trained on ImageNet data, was applied. Then, this model was adapted by introducing additional layers that are specific to the brain tumor classification task. The model was trained on the training dataset, and validation of the data was performed to check its performance during training. Observing the model's activity, adjusting the learning rate,

and avoiding overfitting are key steps. TensorBoard and Model Checkpoint were the tools employed to save the best-performing model.

3.1 Dataset Description

This dataset comprises 7,023 MRIs of the human brain, categorized into four distinct classes: glioma, meningioma, pituitary tumor, and no tumor. These pictures are taken from a collection of three datasets: figshare, the SARTAJ dataset, and Br35H. Particularly, the pictures that the Br35H was used to extract the No Tumor class. It was noted, however, that the SARTAJ dataset contains incorrectly annotated Glioma images, which I obtained from the results of other investigators' and individual researchers' works. These pictures were misplaced, and I decided to relocate them to eliminate them from the dataset and use the Glioma images obtained from Figshare instead. The dataset was used to classify them correctly. The decision to use such data in this research was made to ensure its effective utilization. They are valid and dependable. The pictures in this dataset have varying sizes. After preprocessing, the images should be resized to a standard size, removing any additional margins. This resizing is crucial to ensure the images have uniformity, which enables the models to work effectively during training. The data consists of a vast number of MRI brain images and thus serves as a valuable resource for performing tasks in brain tumor classification. This dataset aims to aid in inferring, classifying, and localizing brain tumors in medical scans. The images are applicable in training deep learning, which is particularly useful for tasks involving the identification and classification of different types of brain tumors.

3.2 Used Models

The model used in the current research is a Convolutional Neural Network (CNN). It is an image classification algorithm optimized for input images of size 150 pixels by 150 pixels, with three channels (RGB). The architecture initially comprises two convolutional layers. The first convolutional layer has 32 filters, and the second convolutional layer has 64 filters, both with a 3×3 kernel and ReLU activation. These convolutional layers were then followed by the Max Pooling operation, which decreases the spatial dimensions of an image. Additionally, a Dropout layer is introduced to set 30% of neurons to zero, thereby avoiding overfitting and enhancing the model's generalization during the learning process.

However, the methodology does not provide a clear explanation for why specific default hyperparameters were chosen (i.e., why choose the optimizer and why use a specific number of epochs, such as 10 epochs). For complex tasks such as brain tumor classification, the number of epochs is usually a subject of careful analysis, and more epochs are required for deep networks to perform better. Without an explanation for these choices, the reader can be left wondering what thoughts are behind these decisions.

Additionally, the problem of class imbalance in the dataset remains unresolved. It is worth discussing whether the dataset was balanced in terms of the different classes (Glioma, Meningioma, Pituitary Tumor, No Tumor), as this significantly influences the model's performance. For cases of class imbalance, approaches such as class weighting, oversampling, or undersampling should be employed to ensure that the model effectively learns to classify all classes. The lack of such a discussion creates a gap in understanding how the model handles potentially skewed data distribution, which may lead to biased results, especially for underrepresented classes.

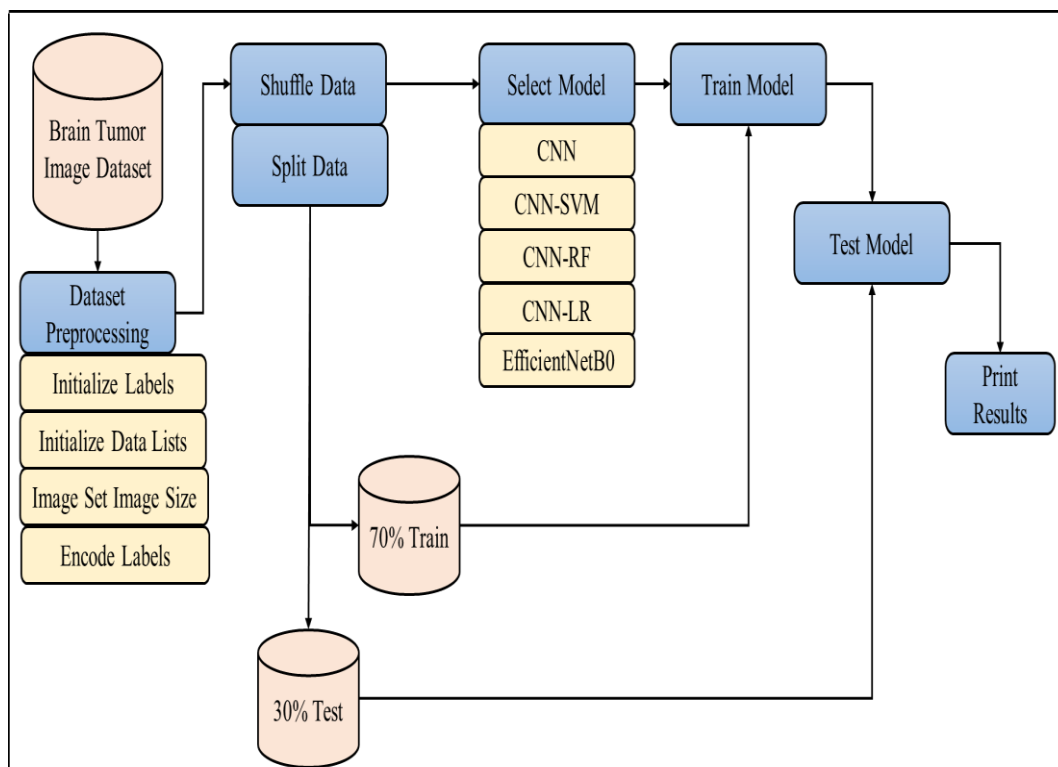


Figure 1: Model Methodology

The deeper the convolutional network gets, the more layers of convolution are incorporated in order to obtain more complex features. Each layer of convolutions employs 64 filters with a 3×3 kernel size, followed by a ReLU activation function. One can find the same pattern repeated all over the architecture. A second Max Pooling layer was then applied to reduce the spatial dimensions of the feature maps. A Dropout layer was used next, setting 30% of the neurons to zero to prevent overfitting.

In the subsequent layers, there were three convolutional layers with 128 filters, a 3×3 kernel size, and ReLU activation. The subsequent layers are the Max Pooling and Dropout layers, following the established pattern once again. However, the description of the CNN architecture is redundant, especially for describing Convolution, Max Pooling, and Dropout layers. Such layers and their functions are explained similarly at different phases of the network and could be redundant. A less wordy explanation of the overall architecture, without repeating the exact details over and over, would be more appropriate to make it clear and easier to read. Additionally, instead of iterating the same layer types, it would be better to emphasize the difference in the number of filters or the purpose of each layer stage as the network grows deeper, allowing for a better insight into how the network changes to detect more complex features.

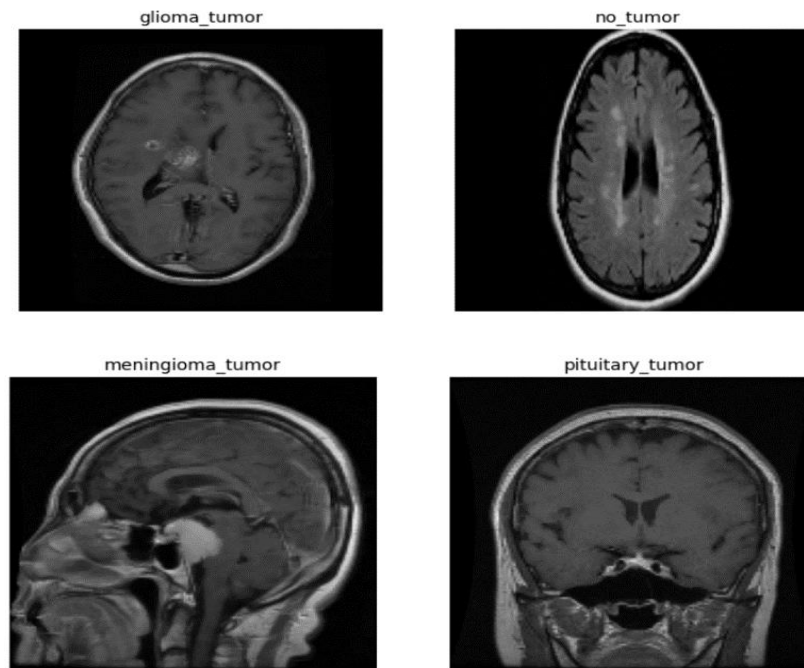


Figure 2: Dataset Sample

The last layer in the Convolutional Block is of size 128, followed by 256 with rectified linear unit activation and 3×3 kernels. The model concludes with a Max Pooling layer, followed by a Dropout layer, which transitions from the convolutional layer to the dense layer. They were flattened to a vector in the XY plane and passed through two fully connected layers of 512 neurons, each with ReLU activation. Elaborately, a Dropout layer was employed to minimize the overfitting candidacy. The last layer of the proposed model was the output layer, which consisted of 4 neurons that utilized a softmax function activation to perform well in the multi-class classification with four classes. This architecture presents the idea of depth and introduces more structure to the process, aiming to identify intricate relations without overfitting.

CNN-SVM Model: This approach utilizes a CNN for feature extraction, while the final layers employ SVM to classify the features extracted by the CNN. The details regarding the CNN model are as follows: the images processed through this network must be of a specific size (referred to as `image_size`) with three color channels. It was initiated with a convolutional layer of 32 filters, each with a 3×3 kernel size, followed by a Max Pooling layer to downsample the spatial dimensions. The pattern was repeated with a second convolutional layer with 64 filtration zones, another Max Pooling layer, the third convolutional layer with 128 filtration zones, and the final Max Pooling layer. Finally, after the above layers, the network flattens the 3D feature maps into a 1D vector, which features the fundamental attributes of the input images [18].

After that, the CNN model obtained through training was utilized to extract features from the training feature set (`X_train`) and the test feature set (`X_test`). These extracted features provide information input to a baseline delegate, a traditional machine learning classifier. In particular, SVM, which has a linear kernel, was used for this purpose. The SVM classifier was then built using the feature vectors derived from the training data, and the resulting classifier was used to classify the test sets. The proposed model utilizes the CNNs' efficient feature extraction and achieves high classification performance by incorporating SVMs to provide a proficient solution for image classification problems [19].

CNN-RF Model: The model applied is based on a CNN to separate the features of an image and an RF—Random Forest classifier for classification. The CNN is composed of three CNN layers, similar to the standard CNN, which includes three convolutional layers. The number of filters for each layer is 32, 64, and 128, respectively. Between such layers, Max Pooling layers are found. Additionally, the steps used to shrink the feature maps are incorporated into it. The last feature maps were then flattened into a 1D array of vectors and used as input to a Random Forest classifier with 100 estimators. This type of integration benefits from the fact that CNN extracts significant features from images, and Random Forest is both powerful and flexible in classification, forming an efficient image classification system [20].

CNN-LR: This model performs feature extraction from images using a CNN, while classification utilizes a logistic regression classifier. The CNN comprises three convolutional layers with 32 filters, 64 filters, and 128 filters, followed by a Max Pooling layer to downsample the spatial dimensions. For each feature map, after being flattened to a vector of size 1024, a Logistic Regression classifier was trained with a maximum of 1000 iterations. This integration leverages the strength of CNN, which can extract features of interest from images, and the weakness of the Logistic Regression model, which excels at making linear classifications, thereby enhancing the image classification pipeline [21].

EfficientNet B0 is one of the architectures from a series of CNN models intended to expand the network's performance and practical usage compared to basic CNNs. EfficientNet is a group of models developed with the assistance of Google AI, based on a baseline network known as EfficientNet B0. All of these models surpass the baseline by utilizing depth, width, and increased resolution.

EfficientNet B0 begins with the base network formed by the sequence of go-throughs containing convolutional operators, batch normalization, and the activation function (Swish), which were grouped into several blocks. It presents a compound scaling approach that scales the dimensions of depth, width, and resolution equally to determine the best trade-off between model complexity and accuracy.

EfficientNet B0 comprises convolutional layers with 32 filters at the initial stage, but gradually increases the number of filters and depth of the network as the stages progress. Each was followed by downsampling using a stride-convolution or pooling layer to decrease the field of view and increase the stride.

The EfficientNet B0 is exceptional in terms of efficiency, as it can operate and perform well in mobile and edge devices with limited computation power. The three design principles used in MobileNetv2, including compound scaling and depth-wise separable convolutions, enhance the model's capability to achieve top-of-the-range accuracy in various image classifications while simultaneously being efficient in terms of model size and computational complexity [22].

3.3 Training parameters

Table 1 gives the setup of the parameters for both machine learning and deep learning models to be utilized during the experiment. Such numbers were determined to obtain the optimal sets of hyperparameters that enabled the best model performance using a specific dataset and task. The table provides detailed data on the principal parameters of the research work, including the type of optimizer, learning rate, type of loss function, and other parameters specified by the model. However, it is unreasonable that the number of epochs is not specified. For example, 10 epochs were used during the training of the CNN model, and 15 were used during the training of EfficientNetB0. It is a minimal value, especially considering

that deep networks, by general rule, need more epochs to converge efficiently. The choice of using such a small number of epochs should be explained, as more epochs are typically required for the good performance of deep neural networks, especially with complex datasets such as brain tumor classification. The logic behind the selection of these particular numbers of epochs should be examined to provide a more comprehensive understanding of the model's performance and the fact that the models are trained sufficiently to yield accurate results.

Table 1: Training Parameters

Convolutional Neural Network	
Optimizer	Adam
Learning Rate	Default
Loss Function	Sparse Categorical Crossentropy
Metrics	Accuracy
Epochs	10
Batch Size	Default (typically 32 or 64)
Support Vector Machine	
Kernel	Linear
C	1
Random Forest	
n_estimators	100
random_state	42
Logistic Regression	
Solver	Liblinear
max_iter	1000
random_state	42
Efficient Net B0	
Optimizer	Default (Adam optimizer)
Loss Function	Default (Sparse Categorical Crossentropy)
Metrics	Default (Accuracy)
Epochs	15
Batch Size	32
Validation Split	0.1
Callbacks	TensorBoard, ModelCheckpoint, ReduceLROnPlateau
TensorBoard Log Directory	'logs'
Model Checkpoint	'effnet.h5', monitor='val_accuracy', save_best_only=True
Reduce LR on Plateau	monitor='val_accuracy', factor=0.3, patience=2, min_delta=0.001

3.4 Evaluation Metrics

When training a classifier, choosing an evaluation metric was pivotal in achieving optimal classifier accuracy. The selection of the appropriate rating scale is crucial for differentiation and ensuring top-notch performance. In this context, a comprehensive analysis of relevant evaluation metrics was conducted to serve as effective discriminators to enhance the generative classifier.

Precision was a commonly utilized metric by many generative classifiers to identify the optimal solution during training. While accuracy is valuable, it may have limitations, including giving less importance to minority classes, perpetuating discrimination, and potentially biasing data from the predominant class.

This section briefly introduces other metrics that explicitly delineate the ideal solution, as shown in Table 2 [23]. The binary classifier utilizes a Confusion Matrix (Figure 3) to assess its performance. Assessments regarding the possible outcomes of classification models were derived from the terms TP (True Positive), TN (True Negative), FP (False Positive), and FN (False Negative), all of which were elements present within the confusion matrix [24].

Table 2: The Elements of the Evaluation Process (Variables, Definitions, and Equations).

Variable	Definition	Equation
Accuracy	The accuracy of predictions from a set of tests can be readily calculated by dividing the number of correct forecasts by the total number of predictions made.	$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$
Precision	Another important metric involves determining the ratio of correctly identified instances from a specific class to all instances predicted to belong to that class.	$Precision = \frac{TP}{TP + FP}$
Recall	When evaluating model performance, it is essential to consider the relationship between the total number of occurrences and the proportion of instances expected to belong to a specific class.	$Recall = \frac{TP}{TP + FN}$
F1-Score	The term used to characterize the precision of a test is the F1-score. The F1-score can range from 0, indicating low recall and precision, to 1, which signifies exceptional performance in terms of recall and precision.	$F1 - Score = 2 \times \frac{precision \times recall}{precision + recall}$

The generic definition applicable to the binary problem can be applied to a multi-class problem. However, it must be described broadly because True/False binary definitions are unreliable [25].

$$MultiClass Accuracy = (y_i, z_i) = \frac{\sum_{i=1}^N TP(C_i)}{\sum_{i=1}^N \sum_{j=1}^N C_{i,j}} \quad \dots\dots\dots (1)$$

Where C is the class number, and N is the number of classes

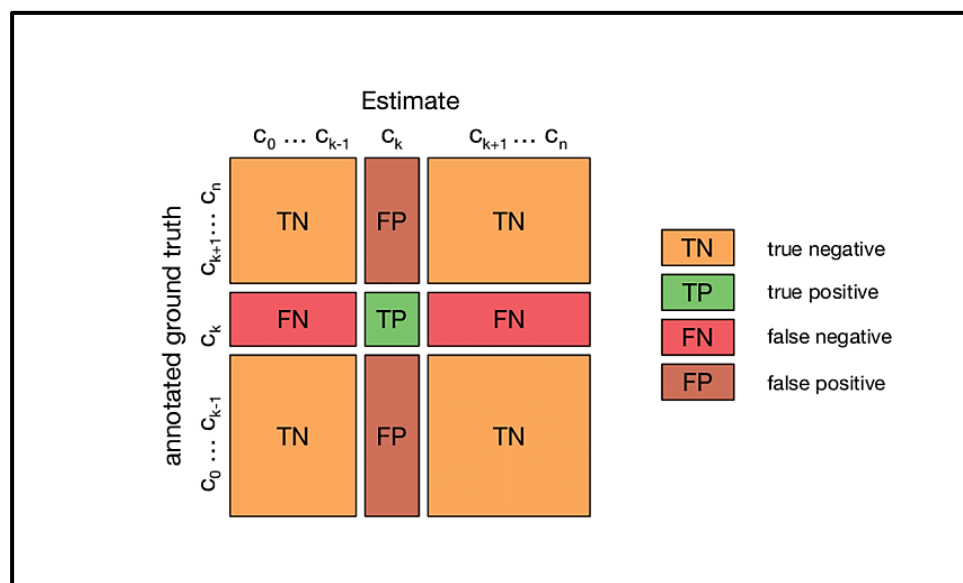


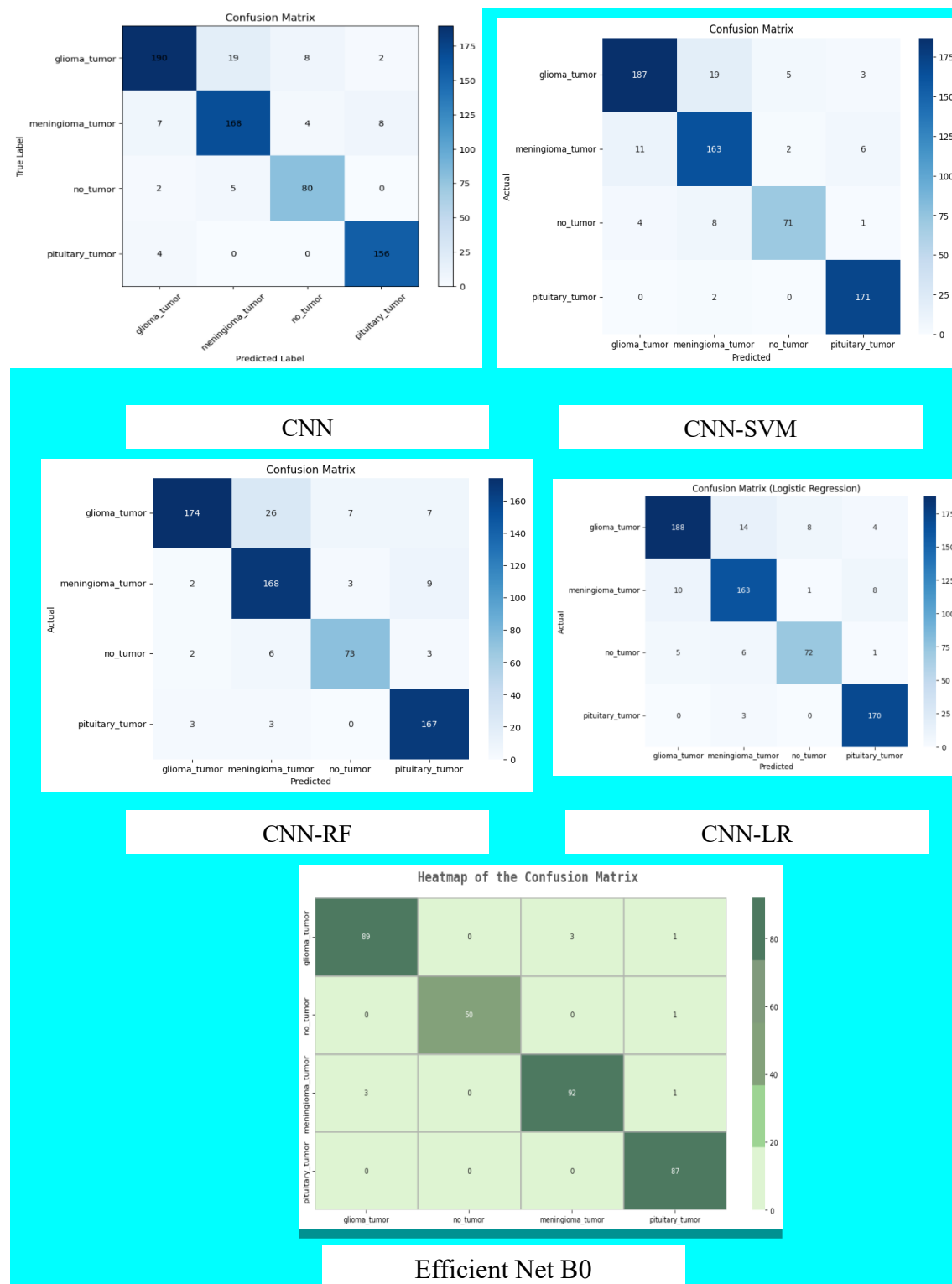
Figure 3: Multi-Class Classification Confusion Matrix [26].

4. Results

The assessment of five different machine learning models — namely, CNN, CNN-SVM, CNN-RF, CNN-LR, and EfficientNet B0 — was conducted based on their confusion matrices (Figure 4). Confusion matrices are used to provide an in-depth explanation of the classification findings by displaying the counts of true positives, false positives, false negatives, and true negatives for each tumor category. Using the analysis of these matrices, the capability of each model to accurately distinguish between a glioma, meningioma, no tumor, or pituitary tumor case is evaluated. Strengths and shortcomings are identified, general trends of misclassification are revealed, and the overall effectiveness in processing complex tumor imaging data is evaluated. This analysis is considered essential for directing developments in classification methodologies and enhancing the accuracy of diagnostics in clinical environments.

The CNN model exhibits excellent overall classification performance, with a high accuracy rate in classifying a large number of samples across all tumor classes. Glioma tumors are thoroughly distinguished, with 190 correct classifications, although there is an overlap with meningioma and no tumor classes. Meningioma tumors are correctly identified, although with low confusion, mostly towards glioma and pituitary tumors. The 'no tumor' category has a correct count that is lower, indicating some difficulty in distinguishing it distinctly from meningioma and glioma. Pituitary tumors exhibit one of the highest correct identification rates, displaying distinct characteristics that CNN can easily recognize. Overall, this model is relatively balanced in terms of sensitivity and specificity; however, it exhibits typical confusion between glioma and meningioma, which were previously shown to display imaging similarities.

The hybrid CNN-SVM achieves a higher classification accuracy compared to the exclusive CNN model, albeit with a slight difference. There is less accuracy in the classification of glioma tumors compared to that achieved by the sole use of the CNN, with fewer misclassifications in the 'no tumor' and 'pituitary tumor' categories. Classification between meningiomas is often confused with gliomas and pituitary tumors due to features shared that impact SVM decision boundaries. The 'no tumor' class suffers the most, with the lowest correct identifications and the highest misclassifications, indicating difficulties in discriminating between normal tissue and tumor classes using this hybrid approach. Pituitary tumor classification remains robust, with the highest accuracy for each class. The model infers the compromises introduced through the use of CNN feature extraction and SVM classification.

**Figure 4:** Models Confusion Metrics.

In the CNN-RF model, Random Forest, combined with features extracted by CNN, offers balanced performance, though slightly less than optimal, for certain classes. The detection of glioma tumors is afflicted by greater misclassification as meningioma, demonstrating the inability to distinguish between these two similar tumors with RF classifiers. The detection of meningioma tumors remains strong, with fewer mistakes, perhaps indicating the RF's adequate distinction between distinctive features in this class. The no tumor class has moderate correct identifications but is still confused with pituitary tumors and meningioma.

Pituitary tumor classification remains confident, with highly correct identification, reiterating its distinctive characteristics. The increased confusion rate between glioma and meningioma may be attributed to the ensemble nature of RF, which can suppress fine feature differences in challenging medical images.

The CNN-LR model exhibits good overall performance, with the best correct classification of glioma tumors among all the models. Some gliomas are misclassified as meningiomas or as having no tumor, indicating overlapping features in some feature spaces that cannot be fully resolved through logistic regression. Meningioma tumors are primarily well discriminated but with significant misclassification to pituitary tumors, as it can be due to similarities of features or a restriction in the linear decision boundary of logistic regression. The 'no tumor' class is discriminated equally well, but still indicates some mistakes with the glioma and meningioma classes. Pituitary tumors are well-discriminated with little overlap, highlighting the model's capacity to encode their distinctive features. In general, CNN-LR is a compromise between competitive accuracy and interpretability; however, the linearity of logistic regression constrains its capabilities.

EfficientNet B0 achieves highly accurate classification for all tumor types with minimal misclassification, indicating strong discrimination strength and feature extraction capabilities inherent in this architecture. The glioma tumor type has a large number of correct identifications with minimal error, suggesting intense discrimination against other tumors. There are no instances of tumors that are nearly perfectly segregated, demonstrating the model's strength in categorizing normal tissue. Meningioma tumors exhibit very robust correct classification with no confusion, and pituitary tumors achieve perfect classification with no misclassification. Such performance establishes EfficientNet B0's complex architecture and its success in identifying small and complex patterns in medical images, making it well-suited for tumor classification problems where precision and reliability are crucial.

The performance of various machine learning and deep learning models is evaluated using four key metrics, Table (3).

Table 3: Evaluation Results

	Accuracy	Precision	Recall	F1-Score
CNN	91%	91%	92%	91%
CNN-SVM	91%	91%	90%	90%
CNN-RF	89%	89%	89%	89%
CNN-LR	91%	91%	90%	90%
Efficient Net B0	99.80%	99.90%	99.80%	99.80%

The efficiency of implementing different varieties of machine learning and deep learning was estimated based on the assessment of Accuracy, Precision, recall, actual positive rate, and F1-Score, which collectively provide insight into how effectively the desired application performs in a particular model. From the baseline CNN model, the performance is relatively good, with an accuracy of 91%, a precision of 91%, a recall of 92%, and an F1-score of 91%. Hence, these results suggest a moderately balanced model. However, if Pareto optimization were applied, the IGBM would tend to classify more instances as relevant, as indicated by its somewhat higher recall. The accuracy and precision remained the same at 91%, but a slight decline was observed in the model's recall (90%) and F1-score, which suggests a compromise in the CNN-SVM model's ability to identify all instances.

The CNN-RF model performs poorly, with an accuracy, precision, recall, and F1-score of 89%. This shows that, although the model outperforms the overfitting baseline by a slight margin, it performs significantly poorer than the CNN and the CNN-SVM models. However,

the CNN-LR model performs just as well as the CNN-SVM model with an accuracy of 91%, precision of 91%, recall of 90%, and F1-score of 90%. It means that logistic regression can be applied successfully in conjunction with CNN, providing performance almost as good as that of the SVM model.

In contrast, the EfficientNetB0 model outperforms all other models, achieving near-perfect results with an accuracy of 99.80%, a precision of 99.90%, a recall of 99.80%, and an F1-score of 99.80%. These results demonstrate the excellent capability of the EfficientNetB0 architecture in processing this dataset and providing outstanding classification results. This analysis further emphasizes that when dealing with complex classification problems, one should opt for sophisticated architectures like EfficientNetB0, which ensures at least twice the performance of the baseline CNN and its variations.

Although the performance of EfficientNetB0 is excellent, it is necessary to continue investigating the likelihood of overfitting, particularly given the high precision achieved. The model's performance needs to be tested using other methods, such as cross-validation or other testing approaches, to ensure the robustness of these findings and confirm that the model is not overfitting the data it is provided with. This would give a more reliable measure of its generalizability and applicability in the real world.

Comparison of the results obtained on the models carried out in this study is presented in Table (4), with the ones retrieved in the most important relevant studies of brain tumor classification. The table summarizes the performance of different models based on major parameters, including accuracy, precision, recall, and F1-score. The study emphasizes the excellent performance of the Efficient NetB0 model, achieving nearly perfect results in all metrics. The table highlights the strengths and weaknesses of the models proposed in the related works, underscoring issues such as the limited dataset size, computational cost, and class imbalance. Comparing these results can help gain a better understanding of the effectiveness of the approaches and the reasons why EfficientNetB0 yields better results than other models in this task.

Table 4: Related Works and Comparison

Study / Model	Accuracy	Remarks
Cheng et al. (2015)	91%	The feature extraction methods did not address the challenges of small datasets [7].
Cherukuri et al. (2020)	96%	High accuracy, but computationally expensive, relies on large datasets [8].
Guan et al. (2019)	98%	It relies on image enhancement, which is not always effective with poor-quality data [9].
Badža et al. (2019)	96%	Deep CNN model, but concerns over computational cost and class imbalance [10].
Deepak et al. (2019)	92%, 97%, 98%	It requires large datasets and lacks proper feature selection analysis [11].
Díaz-Pernas et al. (2020)	97.30%	Multiscale CNN, but reduced sensitivity for meningioma [12].
Alhassan et al. (2020)	98%	Uses hard-swish-based ReLU CNN; lacks handling of dataset size and complexity [13].
Ghassemi et al. (2019)	95.60%	DCGAN for feature extraction, but computationally expensive [14].
Noreen et al. (2020)	94.34%	It uses ensemble classifiers but does not address small datasets or class imbalances [15].
Gumaei et al. (2019)	94.23%	It is a hybrid feature extraction method, but it does not address dataset size or complexity [16].
Haq et al. (2020)	91%, 1%	Lower accuracy for specific tumors, with issues related to sensitivity and generalizability [17].
Current Study (EfficientNetB0)	99.80%	Best performance among all models in the current study.

5. Conclusions

This study aimed to overcome some challenges of deep learning models by developing new Machine Learning techniques. The research is delivering solutions based on a hybrid approach that utilizes mixed ML and DL techniques, matching the classification performance of Deep Learning models. The study provided a breakdown of performance for each tumor class (Glioma, Meningioma, No Tumor, Pituitary Tumor) for each model. Additionally, it employed statistical validation techniques, including cross-validation and significance testing, to reinforce the findings. The new technique is not affected by the dataset size, processing time, or the required training expertise typically required by deep learning models. This study aims to contribute to the enhancement of brain tumor classification automation, thereby overcoming the limitations typically associated with deep learning techniques. The study compares different machine and deep learning models for brain tumor classification, including CNN, CNN-SVM, CNN-RF, and CNN-LR, with an accuracy of 89% to 91%. For comparison, the model with the maximum rate of 99.80% is EfficientNetB0, showing its high performance. Four types of metrics were used to determine the accuracy of the results and compare the five types of models used in the study. The study highlights that the proposed network, particularly EfficientNetB0, demonstrates the capability to tackle the complexity of medical image classification, providing a solution for accurate diagnosis of brain tumors. It is necessary to reveal the precise strengths of EfficientNetB0, for instance, its easy architecture and scalability.

Furthermore, the study does not reflect on potential shortcomings or overfitting, which should be analyzed more thoroughly to ensure the model's generalizability. Future research may involve strengthening the model by utilizing more heterogeneous datasets, exploring hybrid approaches that combine deep learning and classical machine learning, and investigating techniques for addressing class imbalance. Additionally, research into the application of this model in other medical imaging fields or improving its effectiveness for use in real-time clinical settings could provide valuable insights into establishing the field of medical diagnostics.

References

- [1] S. E. Nassar, I. Yasser, H. M. Amer, and M. A. Mohamed, "A robust MRI-based brain tumor classification via a hybrid deep learning technique," *J. Supercomput.*, vol. 80, no. 2, pp. 2403–2427, 2024, doi: 10.1007/s11227-023-05549-w.
- [2] G. Srinivasa Rao, "Enhanced Brain Tumor Classification: A Hybrid Classifier Approach," *J. Electr. Syst.*, vol. 20, no. 5s, pp. 2898–2908, 2024, doi: 10.52783/jes.3203.
- [3] S. De Campos Vieira Abib et al., "International Society of Paediatric Surgical Oncology (IPSO) Surgical Practice Guidelines," *Ecancermedicalscience*, vol. 16, pp. 1–274, 2022, doi: 10.3332/ECANCER.2022.1356.
- [4] H. Zhang and Y. Qie, "Applying Deep Learning to Medical Imaging: A Review," *Appl. Sci.*, vol. 13, no. 18, 2023, doi: 10.3390/app131810521.
- [5] L. Bote-Curiel, S. Muñoz-Romero, A. Gerrero-Curiseses, and J. L. Rojo-álvarez, "Deep learning and big data in healthcare: A double review for critical beginners," *Appl. Sci.*, vol. 9, no. 11, 2019, doi: 10.3390/app9112331.
- [6] S. Iqbal, A. N. Qureshi, J. Li, and T. Mahmood, *On the Analyses of Medical Images Using Traditional Machine Learning Techniques and Convolutional Neural Networks*, vol. 30, no. 5. Springer Netherlands, 2023.
- [7] J. Cheng et al., "Enhanced Performance of Brain Tumor Classification via Tumor Region Augmentation and Partition," *PLoS One*, vol. 10, no. 10, p. e0140381, Oct. 2015.
- [8] N. S. Shaik and T. K. Cherukuri, "Multi-level attention network: application to brain tumor classification," *Signal, Image Video Process.*, vol. 16, Apr. 2022, doi: 10.1007/s11760-021-02022-0.
- [9] Y. Guan et al., "A framework for efficient brain tumor classification using MRI images," *Math.*

- Biosci. Eng.*, vol. 18, no. 5, pp. 5790–5815, 2021, doi: 10.3934/MBE.2021292.
- [10] M. M. Badža and M. C. Barjaktarović, “Classification of brain tumors from mri images using a convolutional neural network,” *Appl. Sci.*, vol. 10, no. 6, 2020, doi: 10.3390/app10061999.
 - [11] S. Deepak and P. M. Ameer, “Brain tumor classification using deep CNN features via transfer learning,” *Comput. Biol. Med.*, vol. 111, p. 103345, 2019, doi: <https://doi.org/10.1016/j.compbiomed.2019.103345>.
 - [12] F. J. Díaz-Pernas, M. Martínez-Zarzuela, D. González-Ortega, and M. Antón-Rodríguez, “A deep learning approach for brain tumor classification and segmentation using a multiscale convolutional neural network,” *Healthc.*, vol. 9, no. 2, 2021, doi: 10.3390/healthcare9020153.
 - [13] A. M. Alhassan and W. M. N. W. Zainon, “Brain tumor classification in magnetic resonance image using hard swish-based RELU activation function-convolutional neural network,” *Neural Comput. Appl.*, vol. 33, no. 15, pp. 9075–9087, 2021, doi: 10.1007/s00521-020-05671-3.
 - [14] N. Ghassemi, A. Shoeibi, and M. Rouhani, “Deep neural network with generative adversarial networks pre-training for brain tumor classification based on MR images,” *Biomed. Signal Process. Control*, vol. 57, pp. 101678–101688, Nov. 2019, doi: 10.1016/j.bspc.2019.101678.
 - [15] N. Noreen, S. Palaniappan, A. Qayyum, I. Ahmad, and M. O. Alassafi, “Brain Tumor Classification Based on Fine-Tuned Models and the Ensemble Method,” *Comput. Mater. Contin.*, vol. 67, no. 3, pp. 3967–3982, 2021, doi: 10.32604/cmc.2021.014158.
 - [16] A. Gumaei, M. M. Hassan, M. R. Hassan, A. Alelaiwi, and G. Fortino, “A Hybrid Feature Extraction Method with Regularized Extreme Learning Machine for Brain Tumor Classification,” *IEEE Access*, vol. 7, no. c, pp. 36266–36273, 2019, doi: 10.1109/ACCESS.2019.2904145.
 - [17] E. U. Haq, H. Jianjun, K. Li, H. U. Haq, and T. Zhang, “An MRI-based deep learning approach for efficient classification of brain tumors,” *J. Ambient Intell. Humaniz. Comput.*, vol. 14, no. 6, pp. 6697–6718, 2023, doi: 10.1007/s12652-021-03535-9.
 - [18] A. R. Al-Qubaa, A. Al-Shiha, and G. Y. Tian, “Gun detection and classification based on feature extraction from a new sensor array imaging system,” *2013 Int. Conf. Electr. Commun. Comput. Power, Control Eng. ICECCPCE 2013*, pp. 88–94, 2014, doi: 10.1109/ICECCPCE.2013.6998740.
 - [19] M. W. Al-Neama, A. A. Mohamad Alshiha, and M. G. Saeed, “A parallel algorithm of multiple face detection on multi-core system,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 29, no. 2, pp. 1166–1173, 2023, doi: 10.11591/ijeecs.v29.i2.pp1166-1173.
 - [20] H. M. Al-Dabbas and M. Salih, “Classification of Brain Tumor Diseases Using Data Augmentation and Transfer Learning,” *Iraqi J. Sci.*, pp. 2275–2286, 2024.
 - [21] H. K. Dhahir and N. H. Salman, “A Review on Face Detection Based on Convolution Neural Network Techniques,” *Iraqi J. Sci.*, pp. 1823–1835, 2022.
 - [22] Z. Y. Younes and E. M. T. A. Alsaadi, “A Review on Plant Leaf Disease Classification Using Deep Learning,” *Iraqi J. Sci.*, pp. 6738–6752, 2024.
 - [23] Ž. Vujović, “Classification Model Evaluation Metrics,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 6, pp. 599–606, 2021, doi: 10.14569/IJACSA.2021.0120670.
 - [24] I. C. Dipto, M. A. Rahman, T. Islam, and H. M. M. Rahman, “Prediction of Accident Severity Using Artificial Neural Network: A Comparison of Analytical Capabilities between Python and R,” *J. Data Anal. Inf. Process.*, vol. 08, no. 03, pp. 134–157, 2020, doi: 10.4236/jdaip.2020.83008.
 - [25] I. Markoulidakis, I. Rallis, I. Georgoulas, G. Kopsiaftis, A. Doulamis, and N. Doulamis, “Multiclass Confusion Matrix Reduction Method and Its Application on Net Promoter Score Classification Problem,” *Technologies*, vol. 9, no. 4, 2021, doi: 10.3390/technologies9040081.
 - [26] T. P. Pagano et al., “Bias and Unfairness in Machine Learning Models: A Systematic Review on Datasets, Tools, Fairness Metrics, and Identification and Mitigation Methods,” *Big Data Cogn. Comput.*, vol. 7, no. 1, pp. 1–31, 2023, doi: 10.3390/bdcc7010015.