STUDYING THE IMPACT OF IMAGING ARTIFACTS ON CNN-BASED BRAIN CANCER DETECTION

by

Parisa Tabassum, BSCSE

A thesis submitted to the Graduate Council of
Texas State University in partial fulfillment
of the requirements for the degree of
Master of Science
with a Major in Computer Science
May 2024

Committee Members:

Mylene Farias, Chair

Jelena Tesic

Damian Valles

DEDICATION

To my parents, whose unwavering love, encouragement, and sacrifices have been my guiding light throughout my life. Your belief in me has been my greatest strength, and I dedicate this thesis to you with heartfelt gratitude.

ACKNOWLEDGEMENTS

In the process of completing this thesis, I would like to express my sincere gratitude to several individuals who have supported and guided me along the way. First and foremost, I am deeply thankful to my thesis advisor, Dr. Mylene Farias, for her invaluable mentorship, expertise, and encouragement throughout this journey. Her insightful feedback and unwavering support have been instrumental in shaping this thesis. I would also like to extend my appreciation to Dr. Jelena Tesic and Dr. Damian Valles for their time and consideration. Additionally, I am grateful to my family and friends for their constant encouragement and understanding during this challenging period. Their unwavering support has provided me with the motivation and strength to persevere. Lastly, I acknowledge the support and resources provided by Texas State University, which have facilitated the completion of this thesis.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

ABSTRACT

Identifying and categorizing brain tumors is crucial for gaining insights into their underlying mechanisms and formulating a treatment plan. Yet, this process often takes a long time and relies heavily on the expertise and experience of radiologists for manual evaluation of Magnetic Resonance Images (MRI). However, Convolutional Neural Networks (CNNs) offer promising tools to aid in brain tumor diagnosis using MRI scans. While MRI is reliable for tumor detection, common artifacts like blurring, noise, contrast, and ringing can compromise the reliability of CNN models. In this study, we investigate the impact of these artifacts on CNN performance by introducing 10 levels of each artifact on MRI scans. We also generate artifacts with similar Structural Similarity Index (SSIM) to assess diagnostic reliability across different image qualities. We evaluate three state-of-the-art CNN models: ResNet50, Inceptionv3, and ResNeXt using the degraded images. The findings from this study provide insights into how each of these artifacts affect CNN models and could help assess the confidence levels of automatic diagnostic results under varying image qualities.

I. INTRODUCTION

A brain tumor is an irregular growth of brain tissue that disrupts normal brain functions. Any unforeseen changes could impact human functioning due to the limited capacity and rigid structure of the human skull, especially if they involve specific areas of the brain. Furthermore, such changes may potentially metastasize to other organs, posing additional risks to overall human function [1]. Brain cancer, categorized according to its characteristics, origin, growth rate, and progression, distinguishes between benign and malignant tumors. Benign tumors typically remain localized and seldom invade nearby healthy tissues, whereas malignant tumors tend to spread to adjacent areas of the brain or spinal cord. Tumors are further classified as primary, originating within the brain, or secondary, arising elsewhere and metastasizing to the brain. Primary tumors are subclassified as glial (comprising glial cells) or non-glial (developing on or within brain structures such as nerves, blood vessels, and glands), and may be benign or malignant. Furthermore, clinicians classify tumors into four grades according to their growth rate and into four stages according to their progression rate [2].

The number of cancer cases and its associated mortality rates are increasing globally [3]. Cancer is one of the primary factors contributing to mortality rates and poses a significant obstacle to extending life expectancy in all countries around the world [4]. The reasons behind the increase in the number of cancer cases and deaths are complex and are attributes to an increase in lifespan or lifestyle choices [5]. Approximately 23.6 million new cancer cases (excluding non-melanoma skin cancer) and nearly 10.0 million cancer-related deaths (excluding non-melanoma skin cancer) were reported in 2019 worldwide [6]. In the same year, 347,992 new cases of brain cancer were recorded and the total number of deaths from brain cancer worldwide was 246,253 [7]. Brain and central nervous system cancer contribute significantly to the worldwide burden of disease, ranking 19th in terms of frequency among all cancers (which represents 1. 9% of all cancers) and 12th among the main causes of cancer-related deaths (which comprises 2. 5% of all cancers) [8]. Brain and central nervous system cancer was placed as the

eighth most impactful cause of Years of Life Lost (YLLs) among all cancers for both genders [9].

Detecting these tumors early is crucial, as it allows for timely intervention and implementation of preventive measures, ultimately decreasing the risk of mortality [10] [11]. Magnetic Resonance Imaging (MRI) is an imaging technique that uses safe, non-ionizing radiation to capture detailed 3D anatomical structures of the body without the need for surgical incisions [12]. Using RF pulses and a strong magnetic field, it generates images [13]. Radiologists use magnetic resonance imaging to detect abnormalities in the brain, assess disease progression, and strategize surgical interventions [14]. However, relying solely on human diagnosis is prone to errors and inconsistencies, as different experts may interpret medical data differently, leading to delays in diagnosing such a sensitive condition. Consequently, the precision of tumor detection through the analysis of brain images from MRI fluctuates depending on the expertise and experience of the healthcare practitioner [15]. Therefore, several recent studies have focused on developing methods to identify and detect brain tumors using MRI images that use machine learning techniques to analyze MRI scans and detect the presence of brain tumors, in hopes of aiding in the early diagnosis and treatment of such conditions [16] [17].

Several types of artifact can arise during MRI scans, due to problems with software, hardware, pulse sequences, or patient-related factors such as tissue variations or movement, and sometimes a combination of these factors can contribute to a single artifact [18]. These artifacts cause the degradation of the image quality of MRI scans, which negatively affects the performance of CNNs when making medical diagnosis [19]. In this context, the primary objective of this thesis is to perform a thorough examination of how the quality of medical images influences the effectiveness of a CNN-based diagnostic system. We develop CNN-based frameworks and assess how the quality of medical images affects the performance of each CNN-based brain tumor detection algorithm, utilizing brain MRI images as input. We use ResNet50, InceptionVv3 and ResNeXt as the CNN models that will perform brain tumor classification. We look at 4 types of commonly occurring artifacts in this study: blurring, noise, contrast, and ringing. We attempt to explain the effects of each artifact on each architecture by

looking at feature maps produced by the ResNet50 architecture. We take the level of image degradation into account by making use of SSIM metric.

The remainder of the report is structured as follows. Chapter 2 provides an overview of the existing literature, including discussions on previous research to diagnosis diseases using image processing and and the effect quality of images have on diagnosis, explaining the importance of this thesis in classification of brain tumor using brain MRI images. Chapter 3 describes The background for this study. Mainly it explains the the CNN architectures used in this study-ResNet50, InceptionV3 and ResNeXt, The performance metrics used to estimate the performance of the CNN models. It also explain confidence intervals which are used to determine the confidence in the results outputted by the architectures and filter outputs that show the effect the input images have on a CNN architecture. Chapter 4 explain the methods to generate the dataset containing images with artifacts. It also describes SSIM, a metric we use in this study to determine the quality of an input image. Chapter 5 outlines the experimental methodology employed in this study, detailing the design, training and testing of the CNN architectures. Chapter 6 presents the empirical findings of the study and discusses the implications of the findings. Finally, Chapter 7 provides a conclusion summarizing the study, and suggesting avenues for future research.

II. RELATED WORKS

In recent years, machine learning has had a significant impact on disease detection using images in various fields, including medical imaging, pathology, and radiology. Hazarika *et al.* [20] proposed a modification to the LeNet model by incorporating MinPooling layers along with MaxPooling layers to improve brain image analysis. Goyal *et al.* [21] used Mask R-CNN for automatic kidney segmentation in coronal T2-weighted Fast Spin Eco MRI slices, augmenting its performance through post-processing morphological operations. Wang *et al.* [22] introduced a modified Inception-v3 CNN architecture to classify breast lesions as benign or malignant, specifically designed for efficient feature extraction from automated breast ultrasound (ABUS) imaging, considering the visualization of ABUS images in both transverse and coronal views. Kanjanasurat *et al.* [20] integrated CNN and recurrent neural network (RNN) models by replacing fully connected CNN layers with a variant of RNN, leveraging CNN feature extraction capabilities and RNN dependency calculation and classification abilities, with CNN models such as VGG19, ResNet152V2, and DenseNet121 combined with long short-term memory (LSTM) and gated recurrent unit (GRU) RNN models. Lin *et al.* [23] proposed deep classifiers, utilizing VGG, ResNet, and DenseNet architectures, to classify SPECT bone images for automated diagnosis of metastasis, employing a pre-processing pipeline involving cropping and geometric transformations to increase original data.

Multiple recent literature have attempted to diagnose brain cancer using Machine Learning based systems. Latif *et al.* [24] introduced a glioma tumor classification method that employs deep learning-based features extracted from MRI scans using a CNN and subsequently classified using a Support Vector Machine classifier by feeding the features to the classifier. Majib *et al.* [25] proposed VGG-SCNet (VGG Stacked Classifier Network), a hybrid model where features are extracted from a top-performing transfer learning model and subsequently utilized as input variables for constructing hybrid models, integrating algorithms such as Stacked Classifier, AdaBoost, CatBoost, and XgBoost. Methil *et al.* [26] introduced a novel approach to the detection

of brain tumors from various brain images, involving preprocessing methods such as histogram equalization and opening, followed by CNN classification, with an experimental evaluation carried out on a data set comprising different tumor shapes, sizes, textures, and locations.

Cinar *et al.* [27] aimed to diagnose brain tumors using MRI images, employing convolutional neural network (CNN) models, specifically utilizing the ResNet50 architecture as the base model with the last 5 layers removed and 8 new layers added for the diagnosis process. Vankdothu *et al.* [28] proposed a fusion of CNN with LSTM units to enhance feature extraction capabilities, resulting in superior image classification performance compared to standard CNN methods. Anaraki *et al.* [29] utilized CNNs and genetic algorithms (GAs) to classify various grades of Glioma brain tumors, wherein the CNN architecture is developed using GA rather than traditional trial and error or predefined structures, and bagging as an ensemble algorithm is used to reduce the variance of the prediction error. ZainEldin *et al.* [30] introduced the CNN brain tumor classification model (BCM-CNN) using an adaptive dynamic sine-cosine fitness gray wolf optimizer (ADSCFGWO) algorithm to optimize CNN hyperparameters, employing a training model built with Inception-ResNetV2 with hyperparameters encompassing both network structure and training, leveraging the strengths of the sine-cosine and grey wolf algorithms within an adaptable framework. Yahyaoui *et al.* [31] presented a novel semantic method for MRI brain tumor classification, integrating 2D and 3D MRI images, which addresses challenges in semantic classification and fusion through preprocessing, classification using two deep learning models and heterogeneous datasets (DenseNet for 2D image classification and 3D-CNN for glioma classification), and fusion using specific domain ontology to merge output classes.

Based on recent research, CNNs have demonstrated impressive accuracy in detecting and segmenting tumors, establishing themselves as the current state-of-the-art solution for certain challenges in medical diagnosis [32]. However, the quality of the input images should be taken into account when assessing the ability of any machine learning model to classify an image. Goodfellow *et al.* [33] explain that as computer vision applications expand, understanding the impact of image quality on computer vision systems becomes crucial, particularly due to the

susceptibility of deep networks to adversarial samples despite their high performance. Hu *et al.* [34] investigated the impact of image quality and lighting consistency on CNN performance in weed mapping, using Faster Region-Based CNN (R-CNN) and Mask R-CNN architectures as examples. Thambawita *et al.* [35] examined how image resolution affects endoscopy image classification by assessing the performance of two CNN models under various quality distortions. Sabottke *et al.* [36] investigated the performance of CNNs, specifically ResNet34 and DenseNet121, in various chest radiograph diagnoses and image resolutions.

A review of the literature shows that the performance of CNNs is greatly affected by the quality of the input images. Multiple factors affect the quality of input images. Several studies have been conducted to study how different factors affect the quality of the input images. Sheikh *et al.* [37] showed that image quality factors, such as resolution, noise, contrast, blur, and compression, affect the visual information contained in the images. Basu *et al.* [38] introduce the n-MNIST dataset, a modified version of the MNIST dataset that incorporates Gaussian noise, motion blur, and reduced contrast, along with a modified deep belief network to improve accuracy in this dataset. Dodge *et al.* [39] evaluated four advanced deep neural network models for image classification across five types of quality distortions: blur, noise, contrast, JPEG, and JPEG2000 compression, highlighting that the etworks are susceptible to these quality distortions. Grm *et al.* [40] explored the impact of image quality on face verification performance in various deep CNN models, finding that high levels of noise, blur, missing pixels, and brightness negatively affect performance, while contrast changes and compression artifacts have a limited impact. Several of the artifacts discussed above appear on MRI images [41]. Therefore, it is paramount that the effects of MRI images degraded with these artifacts on machine learning models.

In our investigation, we did not find a significant number of studies measuring how the quality of (input) MRI scans affects the performance of CNN-based systems. Various degradations, such as ringing effects, noise, and reconstruction artifacts, can affect the quality of MRIs. Although we know that severe image degradation can affect diagnosis, it is difficult to determine how confident we can be in results such as accuracy, precision, recall, and the F1 score.

This is crucial because different MRI machines and factors such as patient movement can create very different images for the same patient. Figuring out the range of input image quality for which we can expect certain accuracy, precision, recall, and F1-score levels is vital for widespread use of machine learning in brain tumor diagnosis.

For the aforementioned reason, the main objective of this study is to carefully examine how the quality of medical images impacts the performance of a diagnostic system for brain tumors from CNN-based magnetic resonance images. We focus on four commonly occurring artifacts in magnetic resonance imaging and used state-of-the-art and widespread CNN architectures in this study. We determine how each artifact affects the input image and the confidence of the results calculated from CNN architectures. We also try to explain why an image degraded with an artifact outputs good or bad classification results while being tested using a CNN architecture. Therefore, we address a big challenge with deep learning methods, which is how much they rely on the quality of the samples in the training dataset. We also try to explain the reason behind this dependence.

III. BACKGROUND ON MACHINE LEARNING

CNN Arhitectures

A Convolutional Neural Network (CNN), also referred to as ConvNet, is a type of neural network designed specifically for handling data with a grid-like structure, such as images. It is a type of deep learning model commonly used for image recognition, classification, and other computer vision tasks. CNNs consists of multiple layers. Convolutional Layers apply convolution operations to input images using learnable filters, known as kernels, to extract features such as edges, textures, and shapes. Pooling Layers downsample the feature maps generated by convolutional layers, reducing their spatial dimensions while retaining important information. Common pooling operations include max pooling and average pooling. Activation functions introduce non-linearity into the network, allowing it to learn complex relationships in the data. Common activation functions include ReLU (Rectified Linear Unit), Sigmoid, and Tanh. Fully Connected Layers are also known as dense layers, which connect every neuron in one layer to every neuron in the next layer, enabling the network to learn high-level representations of the input data. Normalization layers help stabilize and speed up training by normalizing the activations of neurons within each layer. Batch normalization is a commonly used technique in CNNs. The output layer produces the final predictions or classifications based on the features learned by the preceding layers. The number of neurons in this layer depends on the task, with softmax activation often used for classification tasks and linear activation for regression tasks.

ResNet50 [42] is a neural network model introduced by Microsoft Research consisting of 48 convolutional layers, complemented by 1 MaxPool layer and 1 Average Pool layer. This extensive depth enables ResNet50 to dive into deeper architectures without encountering the problem of vanishing gradients, facilitating more effective training. ResNet50 has been widely used in various computer vision tasks, including image classification, object detection, and image segmentation, achieving state-of-the-art performance on many benchmark datasets. Its structure

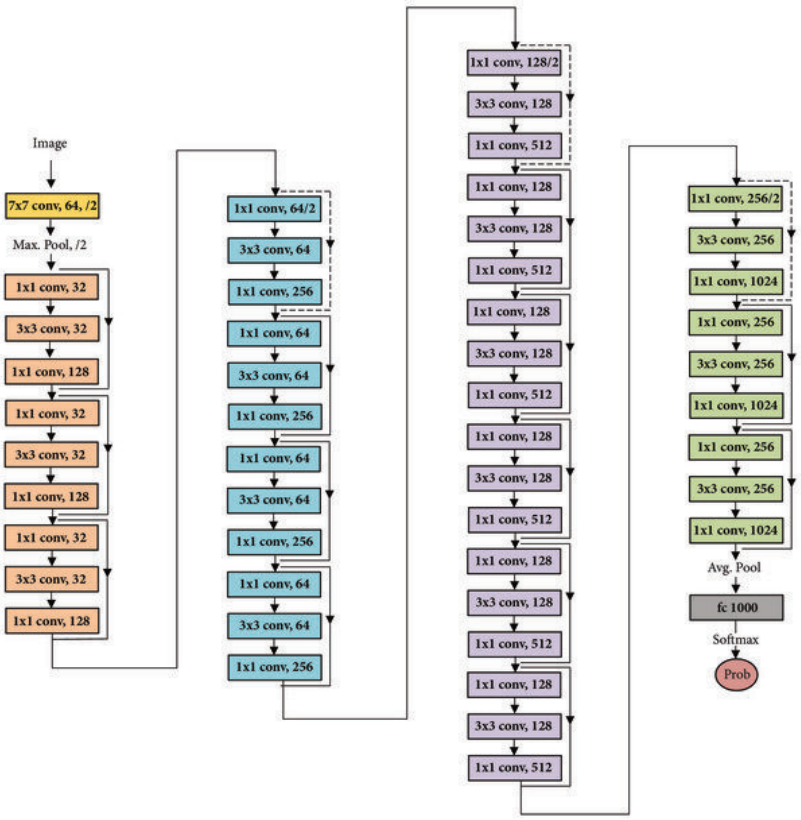


Figure 1: The ResNet50 Architecture

comprises four key components: the convolutional layers, the identity block, the convolutional block, and the fully connected layers. Figure 1 shows the resnet50 architecture.

An Inception Network, developed by Google researchers [43], is a complex neural network characterized by successive blocks, where the output of each block serves as the input of the subsequent one, and each block is referred to as an Inception Block. Inception blocks are made up of multiple parallel convolutional layers with different filter sizes. These modules enable the network to capture features at different scales and resolutions, allowing for more effective feature extraction. Specifically, Inception-v3 represents a convolutional neural network comprising 48 layers, designed to process images with a size of 299 by 299 pixels. Its efficient architecture and excellent performance make it a popular choice for various computer vision applications. Figure 2 shows the Inception-v3 architecture.

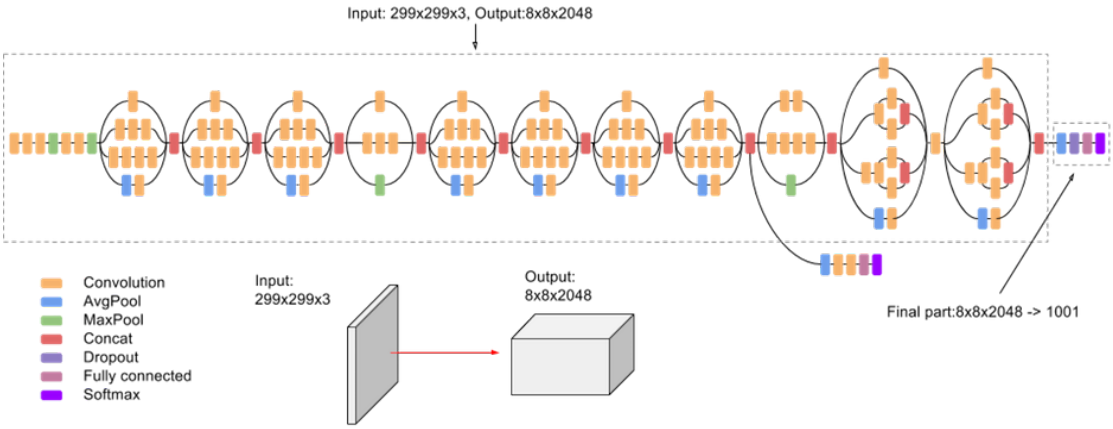


Figure 2: The Inception-v3 Architecture

ResNeXt [44] is a CNN introduced by Facebook AI research. It is a deep convolutional neural network architecture that builds upon the ResNet model by introducing a cardinality parameter, which controls the number of independent paths within each residual block, allowing the network to capture richer representations by aggregating features from multiple paths. This approach enhances the network's ability to learn diverse feature representations while maintaining computational efficiency. A ResNeXt repeats a building block that aggregates a set of transformations with the same topology. We used resnext50_32x4d version of the resNeXt model in this study. This model contains 50 layers. Figure 3 shows the ResNeXt architecture.

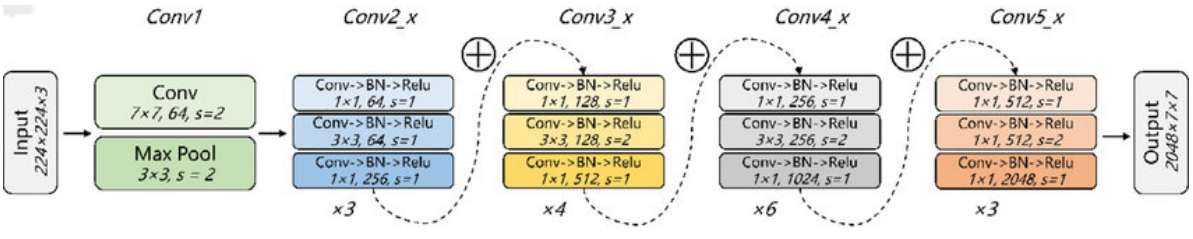


Figure 3: The ResNeXt Architecture

To identify which type of tumor an image contains, we used a CNN-based classification system. More specifically, we used the ResNet50, InceptionetV3 and Resnex architectures to classify the images in the dataset as meningioma, glioma, or pituitary.

We use transfer learning for the purpose of this study, which is a technique where a

pre-trained neural network is used as a starting point for training a new model on a different task or dataset. This approach leverages the pre-trained network's learned features, which are generally useful across different tasks, while allowing the model to adapt and learn task-specific features through fine-tuning. Instead of training the entire network from scratch, all evaluated architectures were pre-trained using weights imported from the Imagenet dataset. The last layer of each network was trained using the dataset from [45]. Training involved the allocation of 68.2% of the images for training, 15.0% for validation, and 16.8% for testing purposes. To avoid the risk that the architecture would be biased by specific characteristics of the patients, such as head shape, the MRI scans of the patients in the training set were different from those of the test set and vice versa. Also, we used cross-entropy as the loss function along with Stochastic Gradient Descent (SGD) as the optimizer. The learning rate was set at 0.0003.

Performance Metrics

Performance metrics are essential in machine learning, as they provide quantitative measures to evaluate, compare, and optimize the effectiveness of models. These metrics help to understand model behavior, make informed decisions about model selection and deployment, and monitor model health in production environments. Using these metrics, practitioners gain insight into the strengths and weaknesses of their models, allowing them to iteratively improve performance and ensure the reliability and effectiveness of machine learning solutions across various tasks and applications. In this study, the performance of the CNN models are evaluated using four metrics: accuracy, precision, recall and F1 score.

In deep learning performance metrics, TP (True Positives) refers to the number of correctly predicted positive instances, TN (True Negatives) represents the number of correctly predicted negative instances, FP (False Positives) indicates the number of negative instances incorrectly classified as positive, and FN (False Negatives) denotes the number of positive instances incorrectly classified as negative. These metrics are used to evaluate the accuracy and effectiveness of classification models in tasks such as classification and object detection.

Accuracy is the ratio of correct predictions (both true positives and true negatives) to the total number of predictions made by the model. It measures the overall correctness of the model's predictions, and it is given by the following equation:

$$c \tag{III.1}$$

Precision is the ratio of true positive predictions to the total number of positive predictions made by the model. It measures the accuracy of positive predictions, and is given by the following equation:

$$\text{Precision} = \frac{TP}{TP + FP}. \tag{III.2}$$

High precision indicates that the model is making fewer false positive predictions. Recall (Sensitivity) is the ratio of true positive predictions to the total number of actual positive instances in the data. It measures the ability of the model to correctly identify positive instances and is given by the following equation:

$$\text{Recall} = \frac{TP}{TP + FN}. \tag{III.3}$$

High recall values indicate that the model is effective in capturing all positive instances. F1 Score is the harmonic mean of precision and recall and is given by the following equation:

$$\text{F1-Score} = \frac{2 * Precision * Recall}{Precision + Recall}. \tag{III.4}$$

The F1-Score provides a balance between precision and recall, taking into account both false positives and false negatives. It reaches its best value at 1 and its worst value at 0.

Confidance Intervals

A confidence interval in statistics is a range of values derived from sample data that is believed to encompass the true value of a population parameter with a certain level of confidence. It provides an estimate of the variability or uncertainty associated with a sample statistic, such as

the mean or proportion, by specifying a range of plausible values around the point estimate. The confidence level indicates the probability that the interval contains the true population parameter, typically expressed as a percentage (e.g., 95% confidence interval).

Confidence intervals are crucial in machine learning classification models because they offer a measure of uncertainty surrounding the model's predictions. By providing a range within which the true value is likely to fall, confidence intervals help practitioners gauge the reliability of classification results. This is particularly valuable in decision-making contexts where understanding the certainty of predictions is essential for risk assessment or resource allocation. For the purpose of this study, we performed each experiment 10 times and calculated the 99% confidence intervals. The 99% confidence interval can be interpreted as there is a 99% probability that the true prediction of a model lies within the range. We used the t.interval() function from the scipy.stats library to get the confidence interval for the CNN models' predictions.

Filter Outputs

For the last part of this study. we attempted to understand why the performance of the networks shows different sensitivity to different artifacts. In machine learning models such as random forests or decision trees, we can understand how they make decisions using a technique called model explainability. Similarly, in CNNs, we can use filters and feature maps to see what the model focuses on in an image. In CNNs, filters are like small grids that slide over the image, extracting features. These filters determine which pixels or parts of the image the model will focus on. Feature maps are the output of a filter passing through the pixel values of an input image. These are what the filters see after scanning the image. This helps us to understand how CNN interprets the input data. We examined the ResNet50 filter outputs separately for the four artifacts to explain the sensitivity of the network to each of the artifacts.

IV. DATASET GENERATION

Dataset With Artifacts

The data used in this research are derived from [45]. Data were collected from Nanfang Hospital in Guangzhou, China, and the General Hospital of Tianjin Medical University in China, covering the period 2005 to 2010. It contains 3,064 T1-weighted contrast-enhanced brain magnetic resonance imaging (MRI). T1-weighted MRI boosts the signal of fatty tissue while suppressing the signal of water [46]. MRI scans were obtained from 233 distinct patients. Each patient is exclusively diagnosed with one of three specific tumor types: meningioma, glioma, or pituitary. The dataset is unbalanced, comprising 708 images in meningioma class, 1426 in glioma class, and 930 in pituitary class. The images are two-dimensional, with pixel values ranging from 0 to 255. They have a resolution of 512x512 pixels, with each pixel covering an area of 0.49x0.49 square millimeters. The slice thickness is 6 mm, and there is a 1 mm gap between slices. Figure 4 illustrates examples of images containing eningioma, glioma, or pituitary tumors [47] from the dataset.

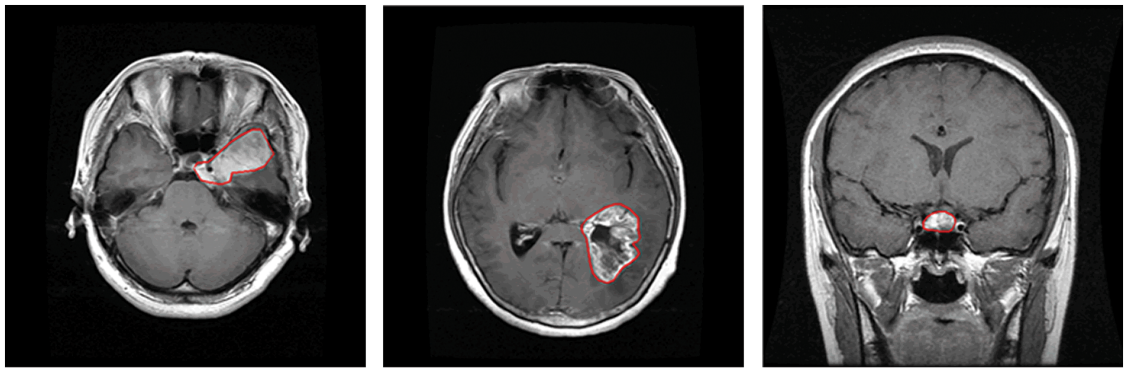


Figure 4: Three types of brain tumors: (a) meningioma (b) glioma and (c) pituitary tumor where red lines indicate the tumor border

The dataset is divided into three subsets according to standard practice: a training set (68.2%), a validation set (15. 0%) and a test set (16.8%). Since there is no direct direct one-to-one

Table 1: Number of images per type of tumor in the Dataset

Tumor type	Number of images
Meningioma	708
Glioma	1426
Pituitary	930

Table 2: Number of images and patients in the training, validation, and test sets in the Dataset

Set	Number of images	Number of patients
Training	2096 (68.2%)	162
Validation	453 (15.0%)	37
Test	515 (16.8%)	34
Total	3064	233

relationship between images and individual patients, as each patient often contributes multiple tumor images to the dataset, it was crucial to preserve the link between patients and their images to avoid unintended information leakage or bias. To ensure this, all images belonging to a specific patient were grouped together within each set, thereby mitigating any potential impact of the model recognizing the patient on tumor classification accuracy. The dataset's image distribution is detailed in Tables 1 and 2.

In real-world scenarios, MRI scans may present a wide array of artifacts or impairments, both visible and invisible, that can negatively impact image quality. These artifacts may arise from hardware malfunctions, software constraints, mishandling of scanning equipment by humans, and issues related to patient movements, whether voluntary or involuntary. To evaluate how these MRI scan degradation affects the performance of a CNN-based diagnostic system, we created modified versions of the dataset images containing typical MRI artifacts, including noise, blur, contrast , and ringing. These artifacts were artificially introduced using Python functions. For each of the four artifact types, we generated a total of 10 degradation levels, ranging from '1' (minimal degradation with scarcely noticeable artifacts) to '10' (high degradation with highly visible and disruptive artifacts). Figures 5, 6, 7, and 8 show examples of images that exhibit 10

levels of blurring, noise, contrast, and ringing, respectively.

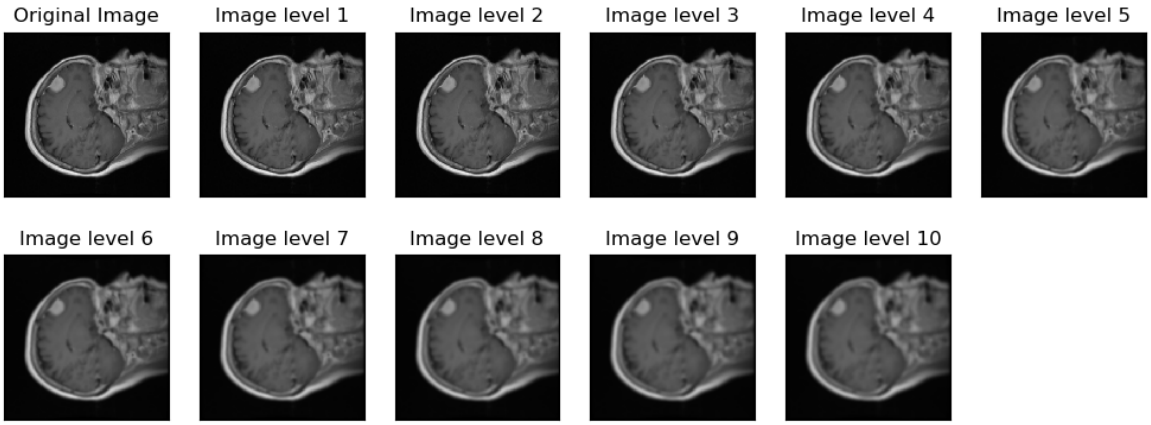


Figure 5: Sample images with different levels of blurring artifacts generated from MRI scans

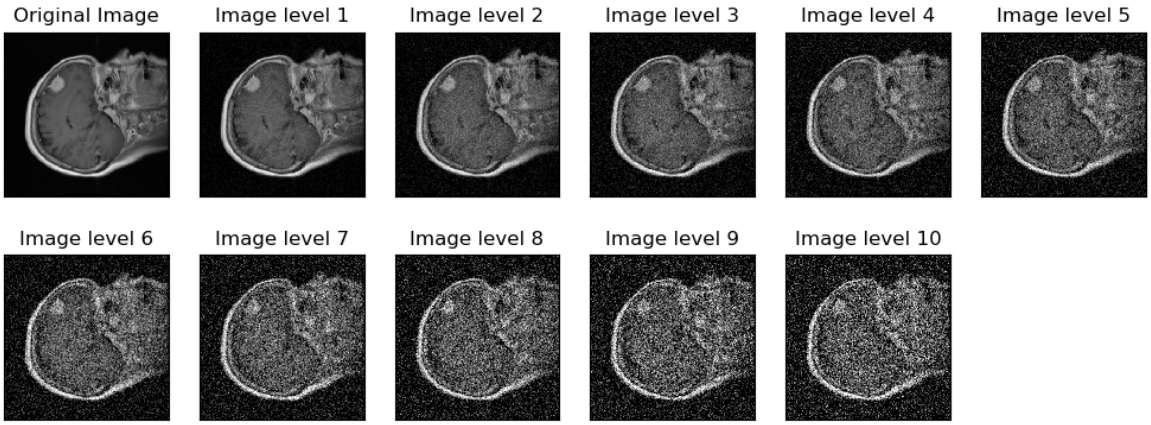


Figure 6: Sample images with different levels of noise artifacts generated from MRI scans

All artifacts were generated following the work of Farias *et al.* [48]. Blurring or blur, a common artifact in image acquisition, is identified by smoothing local intensity fluctuations in pixel values. In magnetic resonance imaging, blurring can result from patient movement during the scan or a limited number of samples. We simulate blur in the images by applying a Gaussian low-pass filter, a technique commonly used to decrease image detail. The level of blurring was adjusted by manipulating the standard deviation of the filter kernel using the GaussianBlur function in Opencv2.1. By modifying the size and standard deviation of the filter kernel, we

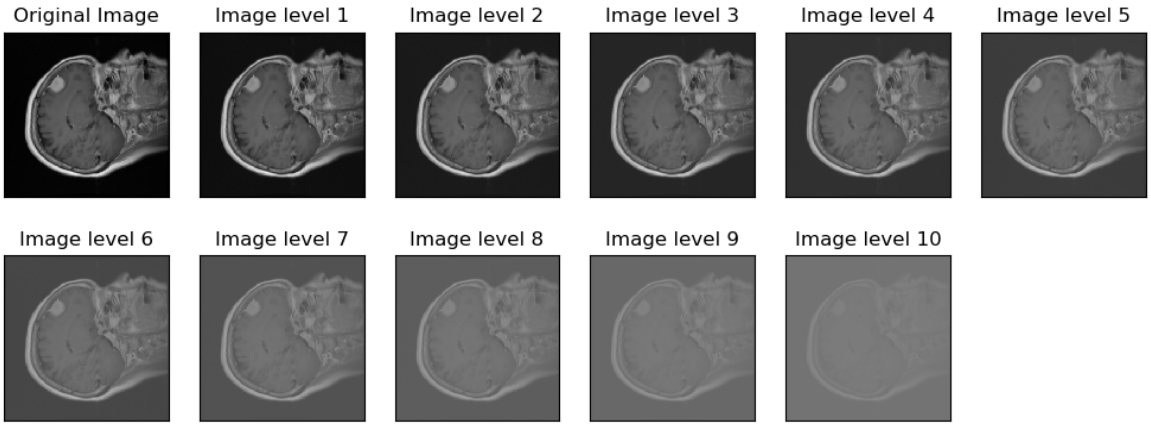


Figure 7: Sample images with different levels of contrast artifacts generated from MRI scans

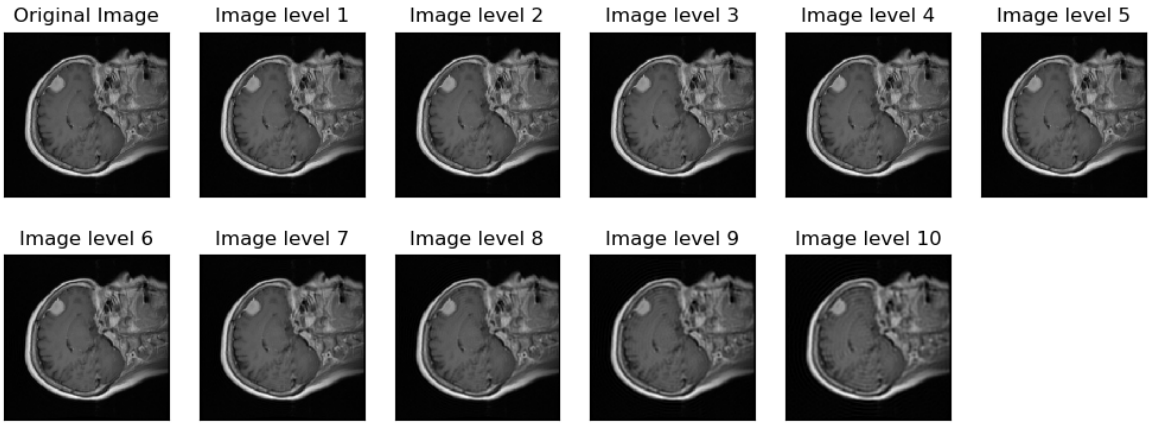


Figure 8: Sample images with different levels of ringing artifacts generated from MRI scans

created various degrees of blur. The blur effect was achieved using the GaussianBlur function in the OpenCV library [49], with parameters that include the size of the kernel K_{size}, the horizontal standard deviation σ_X, and the vertical standard deviation σ_Y. The values of σ_X and σ_Y were adjusted within the range of 0.3 to 12 to produce different levels of blur.

Noise is a type of artifact that is often introduced by the process of acquisition and reconstruction of magnetic resonance images. In addition, noise can be caused by external interference or a low number of samples. In this work, we use a specific type of noise, the additive

Gaussian noise, whose mathematical model is given by the following equation:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \tag{IV.1}$$

where μ is the mean of the noise values and σ^2 is the variance. To introduce Gaussian additive noise into the dataset images, we use the random_noise function from the skimage.util library, which is an open-source library of the Python language [50]. To vary the intensity of the noise (strength of degradation), we vary the values of the horizontal variance σ_X^2 and the vertical variance σ_Y^2 between 0.26 and 16.32, keeping the noise mean zero ($\mu = 0$). We generate noise artifacts by adding Gaussian noise to MRI scans using the Opencv2 random noise function.

A common type of degradation in magnetic resonance images is contrast, or more specifically, limited contrast. In a grayscale image, the dynamic range is defined as the valid range of pixel intensities. The difference between the maximum and minimum values in this range is defined as contrast, while the ratio between these two quantities is defined as the contrast ratio. When the image intensities are not well distributed within the range, the image does not allow a good discernment of its details. To generate images with various levels of contrast, we perform intensity transformation operations, applying the following function to the intensities of the reference image:

$$g(x, y) = \alpha f(x, y) + \beta, \tag{IV.2}$$

where the parameters α and β are the gain and bias parameters, respectively, which control contrast and brightness. The values of α vary between 0.09 and 0.945, while the values of β are defined to keep the histogram centered (values between 115.48 and 6.04).

The Gibbs phenomenon, also known as ringing, is a common deterioration observed in magnetic resonance images [51]. It occurs due to a limited number of high-frequency samples, leading to distinct repetitions of object edges that are transparent and smoothed in the images. This effect is particularly noticeable in areas where there are signal transitions, manifesting as multiple alternating lines of varying brightness near these transition zones. It is crucial to address

this degradation as it alters the image structure, introducing misleading information that can be mistaken for actual image features. To simulate MRI scans with ringing artifacts, we applied frequency domain filtering to the image using ideal low-pass filters with different cutoff frequencies. These filters were designed as circular band-pass regions in the frequency domain, each with its unique characteristics. The brain image's spectrum was then modified by multiplying it with the spectrum of these filters. A larger radius in the filter indicates a higher cutoff frequency, resulting in reduced ringing effects, whereas a smaller radius leads to more pronounced ringing artifacts. Specifically, to generate varying levels of degradation, we adjusted the radius of the passband region between 98 and 14.

Image Quality

Image quality pertains to the faithfulness and sharpness of an image, encompassing elements like resolution, clarity, color precision, contrast, and general aesthetic appeal. This criterion is subjective, being impacted by technical characteristics and the interpretation of observers. Key aspects that contribute to image quality include, among others:

- Sharpness: The sharpness of an image refers to the clarity and precision of its edges and details. A sharp image is characterized by clearly defined edges and minimal blurring.
- Color Accuracy: The accuracy of colors in an image compared to the original scene is crucial. Precise color reproduction is vital for effectively communicating the intended message and atmosphere of the image.
- Contrast: The difference in brightness between the lightest and darkest parts of an image. A good balance of contrast enhances the visual impact and depth of an image.
- Dynamic Range: The range of tones between the darkest and lightest areas of an image. Higher dynamic range allows for better capture of detail in shadows and highlights.
- Presence of Artifacts or distortions: Errors can be added to an image in any stage of the communication pipeline, i.e. during the image capture, processing or editing, compression, and

transmission. When visible, these errors result in visible artifacts or distortions that reduce the perceived image quality. One example of a common artifact is noise, which correspond to random variations in the brightness or color intensities, often visible as graininess or speckles, particularly in low-light conditions or high ISO settings.

The quality of an image is not fixed and can be perceived differently depending on its intended purpose, the viewer's preferences, and the standards of the industry or application where it will be used. For instance, a high-quality image meant for professional printing may have different requirements compared to an image intended for social media sharing or website display. Objective quality metrics are numerical measures utilized to evaluate the quality of digital images, videos, audio, or other multimedia content. These metrics offer a systematic and automated approach to assess various quality aspects, enabling comparisons between different content versions or encoding and processing techniques. In contrast to subjective quality assessment, which depends on human judgment and perception, objective metrics seek to quantify quality through mathematical algorithms and computational analysis.

Among the many available image quality metrics [52] is the SSIM (Structural Similarity Index Measure), which estimates the similarity between two images based on their structural information [53]. SSIM analyzes three primary attributes of an image: luminance, contrast, and structure, and assesses the similarity between two images based on these characteristics. SSIM is calculated using the following formula:

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

where:

x : the reference (undegraded) image

y : the processed (possibly degraded) image

μ_x, μ_y : means of x and y

σ_x^2, σ_y^2 : variances of x and y

σ_{xy} : covariance between x and y

C_1, C_2 : constants to stabilize the division with weak denominator

SSIM provides similarity scores within the range of 0 to 1. A score of 1 suggests high similarity or identical images, whereas a score close to 0 indicates substantial dissimilarity and therefore low quality.

Figures 9, 10, 11 and 12 show MRI scans degraded with different levels of blurring, noise, contrast, and ringing, respectively, and their associated SSIM values. From the images, it can be observed that a certain level of degradation does not yield the same SSIM value for every artifact. So it is important to examine how MRI scans degraded with artifacts that have specific SSIM affect the networks. We used scikit-learn to calculate SSIM between an original image and a degraded version of the same image. For each artifact, we tried to generate degraded MRI scans with SSIM in specific ranges to study how each artifact affects the predictions from CNN models.

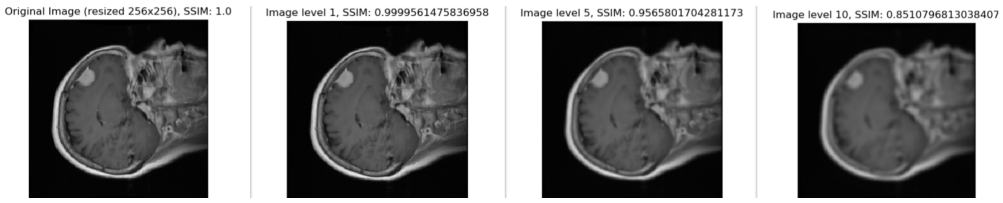


Figure 9: Sample images with different levels of blurring artifacts generated from MRI scans

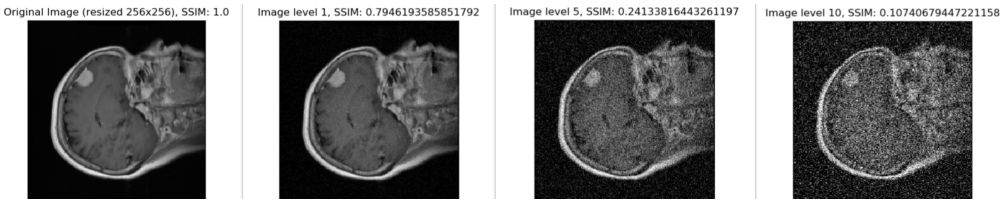


Figure 10: Sample images with different levels of blurring artifacts generated from MRI scans

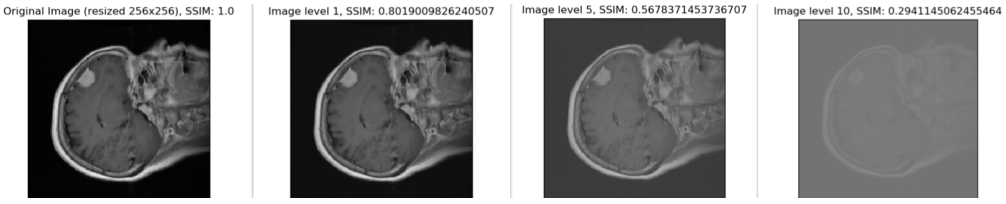


Figure 11: Sample images with different levels of blurring artifacts generated from MRI scans

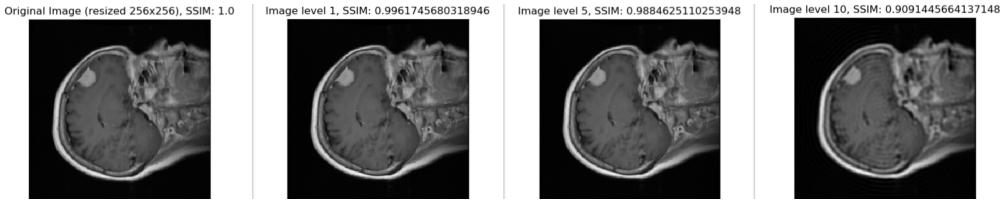


Figure 12: Sample images with different levels of blurring artifacts generated from MRI scans

V. EXPERIMENTAL METHODOLOGY

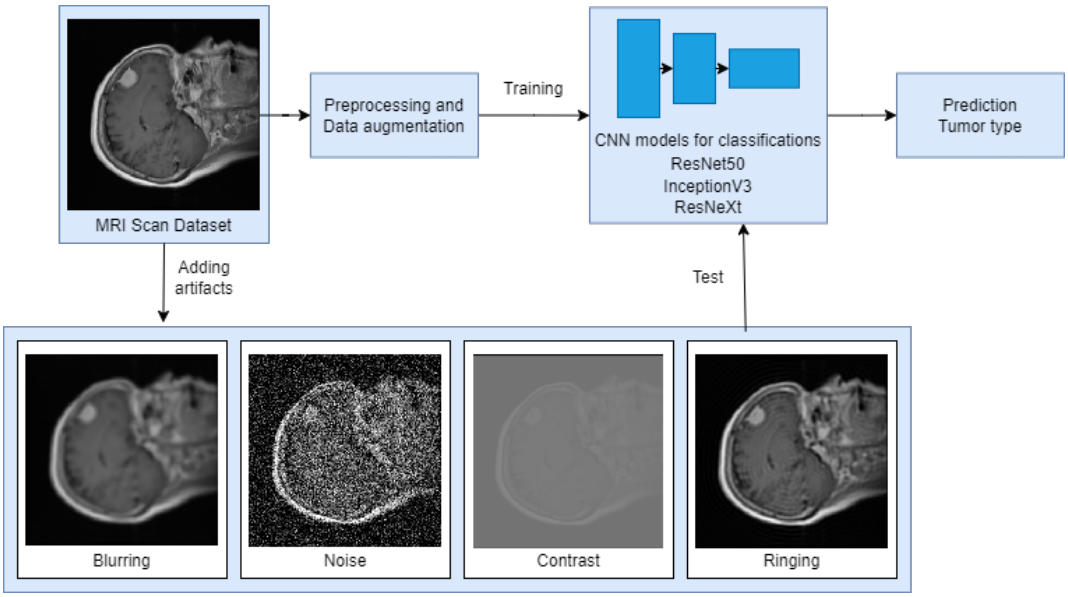


Figure 13: Framework of the experimental methodology used in this work

In this section, we present the systematic approach used for the classification of brain tumors by magnetic resonance imaging using Convolutional Neural Network (CNN) architectures. Figure 13 shows the framework of the experimental methodology used to train and test CNN models and analyze their performance. As described previously, at the beginning of the experiment, we created a custom dataset class that augments each image into 8 different angles: 0, 45, 90, 120, 180, 270, 300, 330 degrees. We fuse this set with Pytorch's DataLoader class so data can be loaded, augmented, and trained in realtime instead of caching all training samples in memory for augmenting. We used 3 CNN architectures in this study, ResNet50, InceptionV3 and ResNeXt, all of which are described in Chapter 3.

Before starting the training, we redefined the last fully connected layer with sequential convolution layers. The structure begins with a linear transformation from the input size to 2048 units, followed by a Scaled Exponential Linear Unit (SELU) activation function and a dropout layer with a dropout probability of 0.4 to prevent overfitting. This sequence is repeated once again

before the final linear transformation to the number of output classes (3) with a Log-Sigmoid activation function, which is commonly used for classification tasks. Then, the last layer of each network was trained using the dataset from [45].

To train the CNN architectures, we used cross-entropy as the loss function along with Stochastic Gradient Descent (SGD) as the optimizer. The learning rate was set at 0.0003. After saving the trained model, we add artifacts to the test image datsets following the steps explained in Section 4. After the degraded dataset generation, we classify each image as one of 3 types of tumor classes (meningioma, glioma, or pituitary) using the trained models.

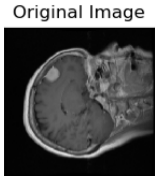
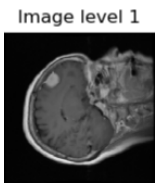
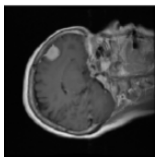
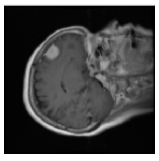
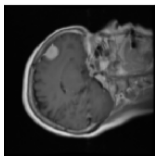
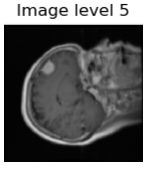
VI. EXPERIMENTAL RESULTS

Classification with artifacts

At the start of this study, we generated the degraded dataset following the description in Chapter 4. Then, we trained the ReNet50, InceptionV3 and ResNeXt architectures using the original dataset from [45]. After training the architectures, we tested them for the classification of MRI brain scans with three tumor types: meningioma, glioma, or pituitary, using both the original and created dataset.

Tables 3, 4, 5, and 6 show the average values for accuracy, precision, recall, and F1 score for images containing blur, noise, contrast, and ringing artifacts, respectively, calculated for the original image and 10 levels of degradation. The tables also show the SSIM value for each image level to show how each level of degradation for each artifact affected the image quality. The first column of the tables shows an example for the images being tested, the second column corresponds to the associated SSIM value, while the rest of the columns present the metrics and the corresponding values.

Table 3: Values for Accuracy, Precision, Recall, and F1 Score for MRI Images Containing Blurring Artifacts using ResNet50, Inception V3 and Resnext models

Image	SSIM	Metrics	ResNet50	InceptionV3	ResNeXt
	1.00	Accuracy	94.130	95.000	93.478
		Precision	94.515	95.464	93.899
		Recall	94.130	95.000	93.478
		F1-Score	94.005	94.894	93.311
	0.99	Accuracy	93.913	95.000	93.478
		Precision	94.292	95.464	93.899
		Recall	9 93.913	95.000	93.478
		F1-Score	93.788	94.894	93.311
	0.99	Accuracy	93.913	95.000	93.478
		Precision	94.292	95.464	93.899
		Recall	93.913	95.000	93.478
		F1-Score	93.788	94.894	93.311
	0.99	Accuracy	93.913	95.217	93.478
		Precision	94.391	95.665	93.764
		Recall	93.913	95.217	93.478
		F1-Score	93.761	95.137	93.310
	0.97	Accuracy	93.695	94.130	94.130
		Precision	94.264	94.361	94.438
		Recall	93.695	94.130	94.130
		F1-Score	93.430	94.070	93.958
	0.95	Accuracy	94.130	95.000	93.478
		Precision	94.515	95.464	93.899
		Recall	94.130	95.000	93.478

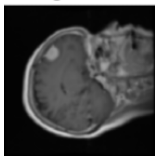
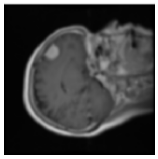
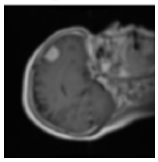
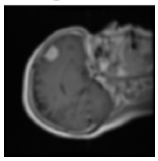
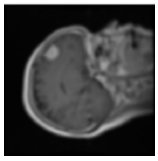
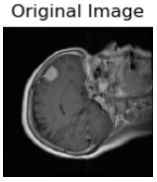
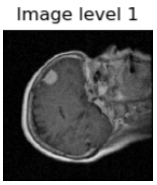
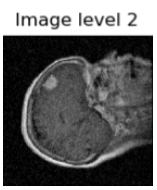
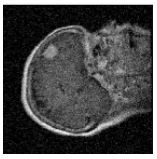
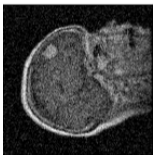
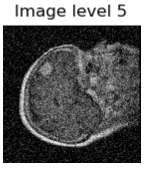
		F1-Score	94.005	94.894	93.311
Image level 6 	0.93	Accuracy	90.869	84.347	92.608
		Precision	91.829	86.920	92.640
		Recall	90.869	84.347	92.608
		F1-Score	90.457	84.703	92.434
Image level 7 	0.90	Accuracy	88.478	78.913	91.956
		Precision	89.369	83.840	91.940
		Recall	88.478	78.913	91.956
		F1-Score	87.797	79.294	91.797
Image level 8 	0.88	Accuracy	88.043	74.130	87.608
		Precision	89.121	80.884	87.462
		Recall	88.043	74.130	87.608
		F1-Score	87.319	74.484	87.315
Image level 9 	0.86	Accuracy	86.739	72.173	85.434
		Precision	88.073	79.989	85.669
		Recall	86.739	72.173	85.434
		F1-Score	85.765	72.363	84.910
Image level 10 	0.85	Accuracy	86.956	71.739	83.695
		Precision	88.227	79.765	84.086
		Recall	86.956	71.739	83.695
		F1-Score	86.087	71.785	82.949

Table 4: Values for Accuracy, Precision, Recall, and F1 Score for MRI Images Containing Noise Artifacts using ResNet50, Inception V3 and Resnext models

Image	Metrics	ResNet50	InceptionV3	ResNeXt	
	1.00	Accuracy	94.130	95.000	93.478
		Precision	94.515	95.464	93.899
		Recall	94.130	95.000	93.478
		F1-Score	94.005	94.894	93.311
	0.79	Accuracy	88.695	93.043	83.695
		Precision	90.916	93.584	88.122
		Recall	9 88.695	93.043	83.695
		F1-Score	88.354	92.863	81.997
	0.55	Accuracy	67.826	92.608	69.565
		Precision	79.630	92.840	82.642
		Recall	67.826	92.608	69.565
		F1-Score	65.346	92.472	67.486
	0.40	Accuracy	66.086	78.695	64.347
		Precision	77.664	81.246	78.536
		Recall	66.086	78.695	64.347
		F1-Score	61.976	78.667	60.316
	0.30	Accuracy	63.478	59.347	67.391
		Precision	74.408	70.362	77.391
		Recall	63.478	59.347	67.391
		F1-Score	57.871	56.289	62.670
	0.24	Accuracy	58.043	44.782	70.217
		Precision	46.745	59.929	76.625
		Recall	58.043	44.782	70.217

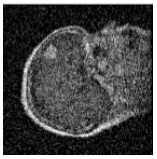
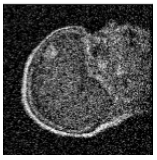
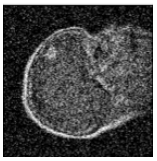
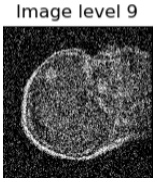
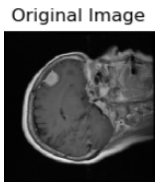
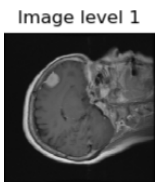
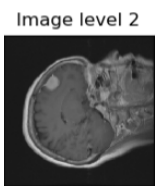
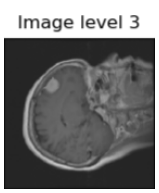
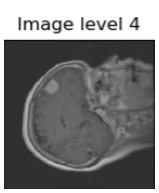
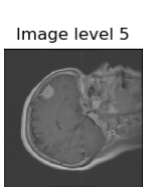
		F1-Score	50.378	35.910	63.919
Image level 6 	0.19	Accuracy	52.391	37.826	58.695
		Precision	44.181	66.480	68.569
		Recall	52.391	37.826	58.695
		F1-Score	45.145	26.868	50.032
Image level 7 	0.16	Accuracy	43.913	35.434	52.391
		Precision	39.647	19.375	63.896
		Recall	43.913	35.434	52.391
		F1-Score	35.958	24.918	41.487
Image level 8 	0.14	Accuracy	38.260	35.652	47.391
		Precision	37.727	20.559	49.136
		Recall	38.260	35.652	47.391
		F1-Score	28.510	24.819	32.484
Image level 9 	0.12	Accuracy	35.652	34.782	47.391
		Precision	35.411	23.594	55.235
		Recall	35.652	34.782	47.391
		F1-Score	24.366	23.118	31.971
Image level 10 	0.10	Accuracy	33.695	31.521	47.173
		Precision	34.845	23.641	44.516
		Recall	33.695	31.521	47.173
		F1-Score	20.535	17.458	31.372

Table 5: Values for Accuracy, Precision, Recall, and F1 Score for MRI Images Containing Contrast Artifacts using ResNet50, Inception V3 and Resnext models

Image	SSIM	Metrics	ResNet50	InceptionV3	ResNeXt
	1.00	Accuracy	94.130	93.695	93.478
		Precision	94.515	95.464	93.899
		Recall	94.130	95.000	93.478
		F1-Score	94.005	94.894	93.311
	0.80	Accuracy	94.347	94.130	93.695
		Precision	94.848	94.720	94.215
		Recall	94.347	94.130	93.695
		F1-Score	94.193	93.988	93.491
	0.70	Accuracy	94.347	93.695	91.521
		Precision	95.005	94.316	92.876
		Recall	94.347	93.695	91.521
		F1-Score	94.128	93.540	91.003
	0.65	Accuracy	90.217	93.695	89.347
		Precision	91.879	94.392	91.517
		Recall	90.217	93.695	89.347
		F1-Score	89.359	93.525	88.446
	0.61	Accuracy	88.043	93.478	85.217
		Precision	90.384	94.111	89.279
		Recall	88.043	93.478	85.217
		F1-Score	86.710	93.310	83.537
	0.56	Accuracy	83.478	92.826	82.608
		Precision	86.859	93.473	88.076
		Recall	83.478	92.826	82.608

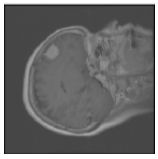
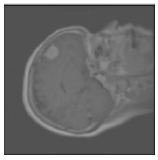
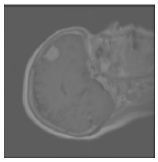
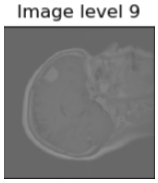
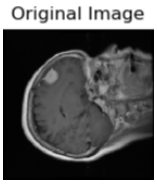
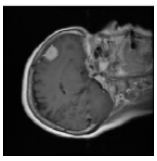
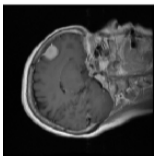
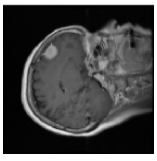
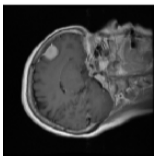
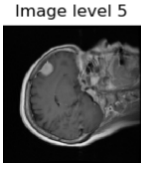
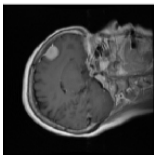
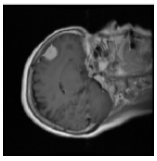
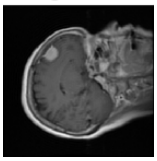
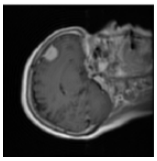
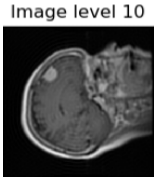
		F1-Score	80.767	92.669	79.965
Image level 6 	0.52	Accuracy	75.869	90.652	82.173
		Precision	80.730	91.789	88.019
		Recall	75.869	90.652	82.173
		F1-Score	70.250	90.527	79.032
Image level 7 	0.47	Accuracy	64.782	88.913	81.086
		Precision	73.056	90.227	85.953
		Recall	64.782	88.913	81.086
		F1-Score	58.249	88.813	77.172
Image level 8 	0.41	Accuracy	52.173	82.608	77.608
		Precision	65.428	84.816	82.059
		Recall	52.173	82.608	77.608
		F1-Score	41.189	82.294	72.682
Image level 9 	0.35	Accuracy	47.608	62.173	69.565
		Precision	54.997	64.294	77.922
		Recall	47.608	62.173	69.565
		F1-Score	32.368	61.009	65.068
Image level 10 	0.29	Accuracy	46.521	27.391	53.695
		Precision	38.607	26.318	73.101
		Recall	46.521	27.391	53.695
		F1-Score	30.371	18.995	44.179

Table 6: Values for Accuracy, Precision, Recall, and F1 Score for MRI Images Containing Ringing Artifacts using ResNet50, Inception V3 and Resnext models

Image	SSIM	Metrics	ResNet50	InceptionV3	ResNeXt
	1.00	Accuracy	94.130	93.695	93.478
		Precision	94.515	95.464	93.899
		Recall	94.130	95.000	93.478
		F1-Score	94.005	94.894	93.311
	0.99	Accuracy	94.565	95.217	93.913
		Precision	94.965	95.683	94.362
		Recall	94.565	95.217	93.913
		F1-Score	94.438	95.135	93.717
	0.99	Accuracy	94.347	95.217	93.695
		Precision	94.739	95.683	94.141
		Recall	94.347	95.217	93.6956
		F1-Score	94.222	95.135	93.502
	0.99	Accuracy	94.1304	95.000	93.260869
		Precision	94.448	95.513	93.705
		Recall	94.130	95.000	93.260
		F1-Score	94.003	94.908	93.071
	0.99	Accuracy	93.913	94.782	93.260
		Precision	94.223	95.272	93.705
		Recall	93.913	94.782	93.260
		F1-Score	93.785	94.691	93.071
	0.98	Accuracy	93.913	94.565	93.043
		Precision	94.223	95.126	93.513
		Recall	93.913	94.565	93.043

		F1-Score	93.785	94.436	92.830
Image level 6 	0.98	Accuracy	93.913	94.130	93.260
		Precision	94.339	94.592	93.705
		Recall	93.913	94.130	93.2608
		F1-Score	93.743	94.006	93.071
Image level 7 	0.97	Accuracy	93.913	93.695	92.391
		Precision	94.197	94.427	92.951
		Recall	93.913	93.695	92.391
		F1-Score	93.739	93.652	92.217
Image level 8 	0.96	Accuracy	93.260	90.652	90.652
		Precision	93.818	91.472	91.130
		Recall	93.260	90.652	90.652
		F1-Score	92.961	90.671	90.333
Image level 9 	0.93	Accuracy	91.521	81.304	88.2608
		Precision	91.867	84.771	88.412
		Recall	91.521	81.304	88.260
		F1-Score	91.244	81.454	87.908
Image level 10 	0.90	Accuracy	90.434	71.739	84.347
		Precision	90.790	81.032	84.372
		Recall	90.434	71.739	84.347
		F1-Score	90.1611	71.200	83.809

The image quality and CNN architecture influence performance metrics across all four types of artifacts. To take a closer look at this effect, we plotted the results from all three CNN architectures for each of the performance metrics and each of the four artifacts. Figure 14(a)-(d) shows the plots of accuracy, precision, recall, and F1 score, respectively, achieved from ResNet50, Inceptionv3 and ResNeXt models for images containing blur artifacts. As the blur increases, the

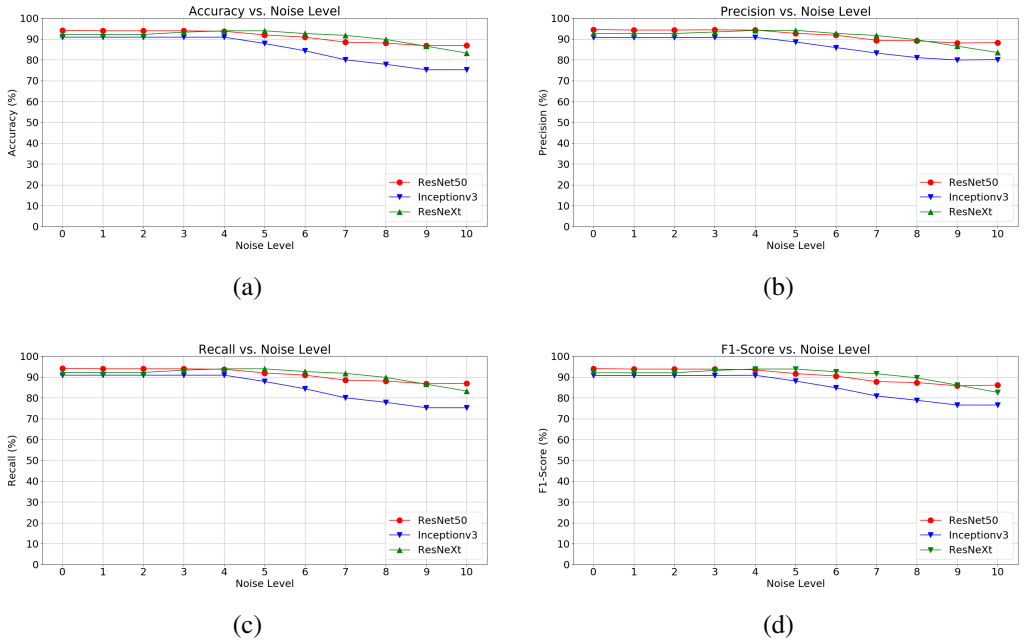


Figure 14: Graphs of values for (a) accuracy, (b) precision, (c) recall, and (d) F1 score achieved from ResNet50, Inceptionv3 and ResNeXt models for images containing blur artifacts

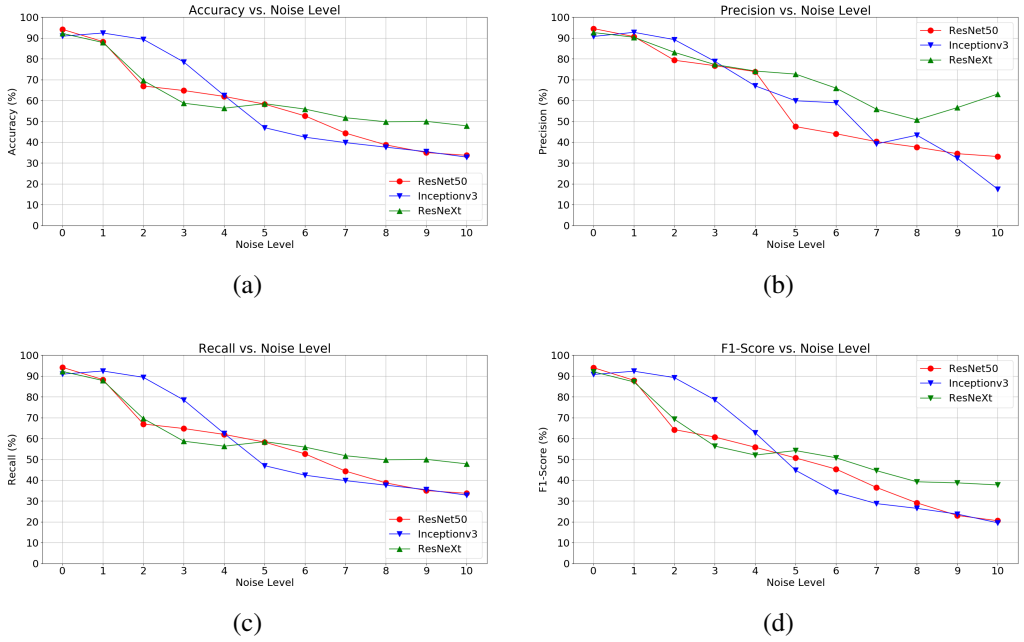


Figure 15: Graphs of values for (a) accuracy, (b) precision, (c) recall, and (d) F1 score achieved from ResNet50, Inceptionv3 and ResNeXt models for images containing noise artifacts

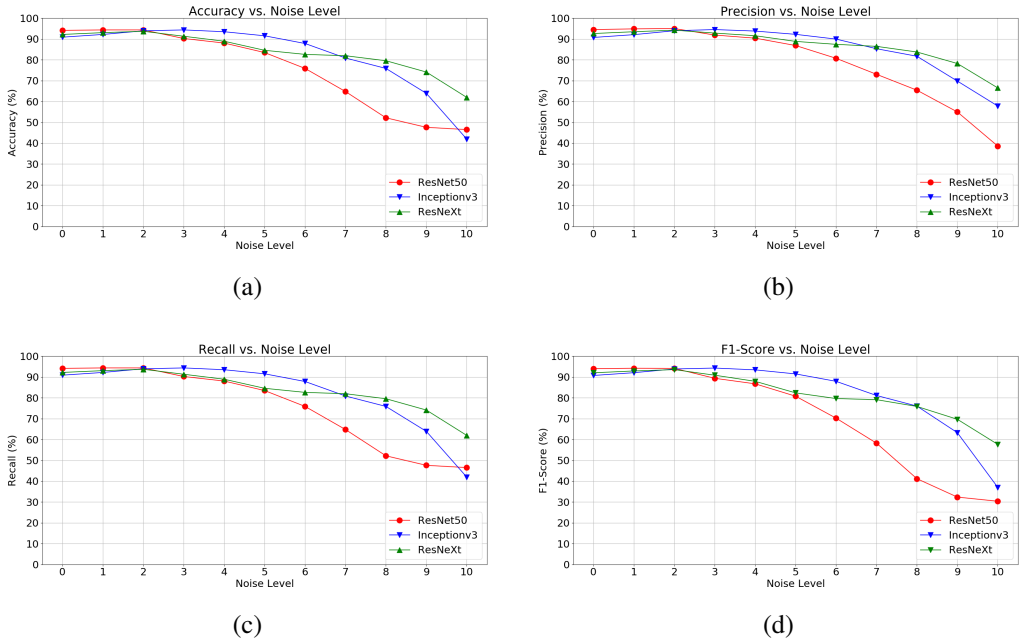


Figure 16: Graphs of values for (a) accuracy, (b) precision, (c) recall, and (d) F1 score achieved from ResNet50, Inceptionv3 and ResNeXt models for images containing contrast artifacts

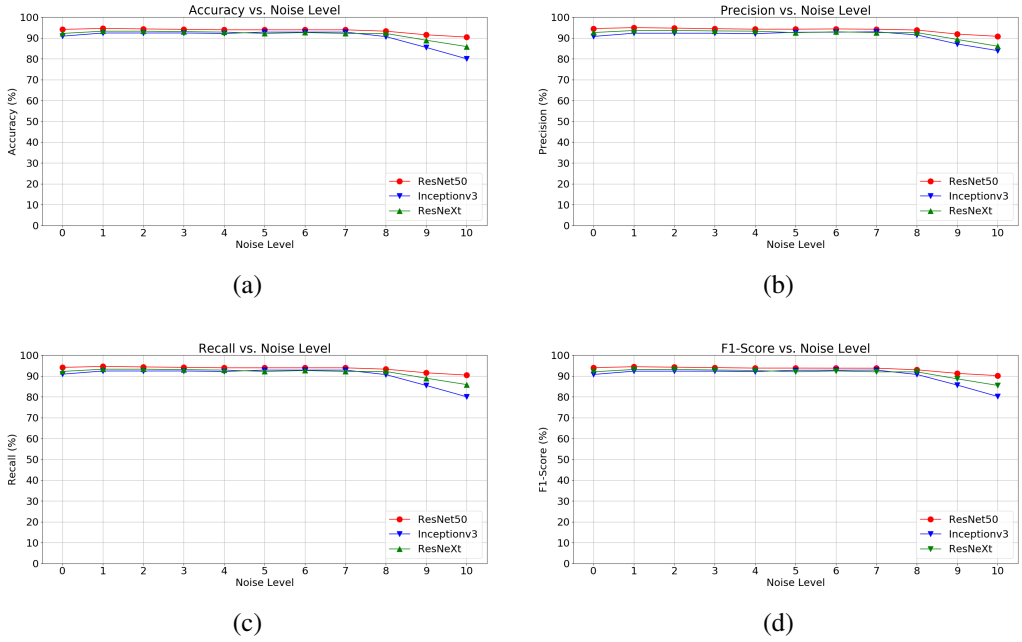


Figure 17: Graphs of values for (a) accuracy, (b) precision, (c) recall, and (d) F1 score achieved from ResNet50, Inceptionv3 and ResNeXt models for images containing ringing artifacts

performance decreases, although it does not decrease a lot. Figure 15(a)-(d) shows the plots of accuracy, precision, recall, and F1 score, respectively, achieved from ResNet50, Inceptionv3 and ResNeXt models for images containing noise artifacts. As the noise increases, the performance decreases significantly. Figure 16(a)-(d) shows the plots of accuracy, precision, recall, and F1 score, respectively, achieved from ResNet50, Inceptionv3 and ResNeXt models for images containing contrast artifacts. As the contrast increases, the performance decreases significantly. Figure 17(a)-(d) shows the plots of accuracy, precision, recall, and F1 score, respectively, achieved from ResNet50, Inceptionv3 and ResNeXt models for images containing ringing artifacts. As the ringing increases, the performance decreases, although it does not decrease much.

From figures 14, 15, 16, and 17, it can be observed that noise and contrast affect the performance values more significantly compared to blurring and ringing. To explain this difference, we can take a closer look at the SSIM values associated with each level of images shown in Tables 3, 4, 5, and 6. The lowest value of SSIM for blur is 0.85 and ringing is 0.92. The lowest value of SSIM for noise is 0.10 and contrast is 0.29. From these values, we can see that for noise and contrast, images get significantly more degraded compared to blur and ringing, which can explain the difference in performance values.

Confidence Intervals

In this section, we calculated the confidence intervals for accuracy of all four artifacts (blur, noise, contrast, and ringing) tested using the three CNN architectures. We performed each experiment 10 times and calculated the 99% confidence intervals. Tables 7, 8, 9, 10 show confidence intervals for accuracy of blur, noise, contrast, and ringing, respectively. The first column of the tables corresponds to the degradation lavels, the second, third, and fourth columns reports the confidence intervals for accuracy while testing with ResNet50, InceptionV3 and ResNeXt, respectively. From the values in the tables, we can notice that the difference between the upper limit and the lower limit of the confidence intervals is very small. Therefore, the results will remain highly consistent throughout multiple test runs.

Table 7: Confidence Intervals for accuracy of prediction with MRI scans with blurring artifacts

Level of Degradation	ResNet50	Inceptionv3	Resnext
Level 1	93.91304347826087	95.0	93.47826086956522
Level 2	93.04347826086953, 93.04347826086956	95.0	93.47826086956522
Level 3	93.04347826086953, 93.04347826086956	95.21739130434783	93.47826086956522
Level 4	93.04347826086953, 93.04347826086956	95.21739130434783, 95.21739130434786	94.1304347826087
Level 5	91.95652173913044	94.1304347826087, 94.13043478260873	94.1304347826087, 94.13043478260873
Level 6	91.30434782608695, 91.30434782608698	84.34782608695652	93.69565217391302, 93.69565217391305
Level 7	89.1304347826087, 89.13043478260873	78.91304347826087	91.95652173913044
Level 8	88.04347826086956	74.1304347826087	91.95652173913044, 91.95652173913047
Level 9	86.73913043478261	74.1304347826087, 74.13043478260873	85.43478260869566
Level 10	83.69565217391302, 83.69565217391305	71.73913043478261	83.69565217391305

Table 8: Confidence Intervals for accuracy of prediction with MRI scans with noise artifacts

Level of Degradation	ResNet50	Inceptionv3	Resnext
Level 1	88.26086956521739	93.69565217391305	83.26086956521739
Level 2	81.57679693050096, 82.33624654775991	93.1370433268834, 94.03686971659485	83.28572316119093, 84.322972490983
Level 3	55.02850796922677, 56.841057248164525	91.84406265478763, 92.93854604086455	69.32320654478715, 70.32896736825629
Level 4	47.393639887271135, 49.08462098229408	78.70395214175845, 80.51343916258936	63.798500148002105, 65.33193463460658
Level 5	44.21176138927963, 45.223021219416026	58.78011290313251, 60.132930575128334	66.20749265069236, 68.66207256669895
Level 6	40.71181329229986, 42.46209975117841	44.55772435971331, 46.44227564028669	67.92611298052391, 69.50866962817175
Level 7	36.125032176170905, 37.39670695426388	37.96519306189826, 39.86089389462349	59.08607061436849, 61.78349460302284
Level 8	33.056179679360426, 33.85686379890045	34.70576792829051, 37.511623376057315	51.26496500262794, 52.60460021476336
Level 9	31.687580141163487, 32.3124198588365	33.780406025048364, 37.30655049669076	47.980261909883694, 48.93278156837718
Level 10	30.867249693201952, 31.219706828537188	32.486630489619486, 35.209021684293546	46.926334578249225, 48.20410020435946

Table 9: Confidence Intervals for accuracy of prediction with MRI scans with contrast artifacts

Level of Degradation	ResNet50	Inceptionv3	Resnext
Level 1	94.34782608695652	94.1304347826087	93.69565217391305
Level 2	94.34782608695652	94.1304347826087, 94.13043478260873	93.69565217391302, 93.69565217391305
Level 3	90.21739130434783	93.69565217391302, 93.69565217391305	89.34782608695652
Level 4	85.21739130434783, 85.21739130434786	93.69565217391302, 93.69565217391305	85.21739130434783
Level 5	83.04347826086953, 83.04347826086956	92.82608695652173	85.21739130434783, 85.21739130434786
Level 6	75.8695652173913	90.65217391304348	82.17391304347827
Level 7	64.78260869565217	88.91304347826087	81.08695652173913
Level 8	52.17391304347826	82.6086956521739	77.6086956521739
Level 9	47.608695652173914	62.17391304347826	69.56521739130434
Level 10	46.52173913043478	62.17391304347826, 62.17391304347827	53.69565217391305

Table 10: Confidence Intervals for accuracy of prediction with MRI scans with ringing artifacts

Level of Degradation	ResNet50	Inceptionv3	Resnext
Level 1	94.1304347826087	95.21739130434783	93.91304347826087
Level 2	94.56521739130434	[95.21739130434783, 95.21739130434786]	93.69565217391305
Level 3	94.34782608695652	[95.21739130434783, 95.21739130434786]	[93.69565217391302, 93.69565217391305]
Level 4	[94.1304347826087, 94.13043478260873]	94.56521739130434	93.26086956521739
Level 5	[94.1304347826087, 94.13043478260873]	[94.78260869565214, 94.78260869565217]	93.04347826086956
Level 6	93.91304347826087	94.1304347826087	[93.04347826086953, 93.04347826086956]
Level 7	[93.69565217391302 , 93.69565217391305]	[94.1304347826087, 94.13043478260873]	92.3913043478261
Level 8	[93.04347826086953 , 93.04347826086956]	93.69565217391302, 93.69565217391305	90.65217391304348
Level 9	91.52173913043478	81.30434782608695	88.26086956521739
Level 10	[90.21739130434783 , 90.21739130434786]	81.30434782608695, 81.30434782608698	84.34782608695652

Feature Maps

In this section, we examine the response of the neural network architectures for the MRI scan image shown in Figure 18. Figures 19, 20, 21 and 22 illustrate feature maps from the ResNet50 network on the first and last convolutional layers for the original image without artifacts and the same image degraded with blur, noise, contrast, and ringing artifacts, respectively. For the image with artifact, we used 'level 5' degradation for all four artifacts.

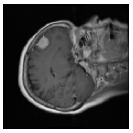


Figure 18: The reference image used to extract the feature maps.

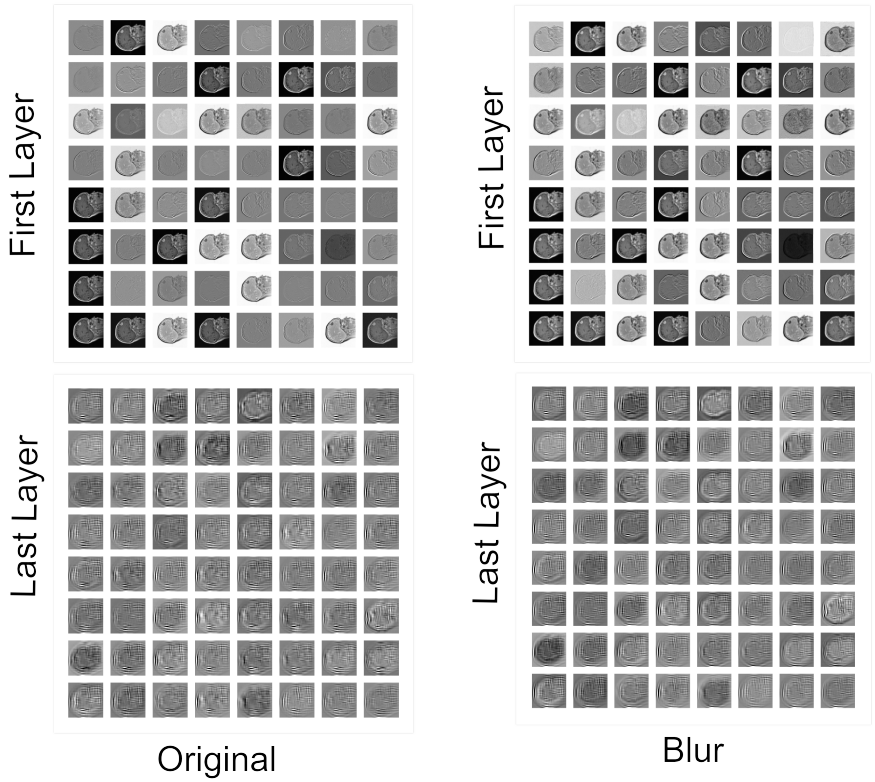


Figure 19: Feature maps from the ResNet50 network on the first and last convolutional layers for original image without artifacts and the same image degraded with blur artifact

In Figure 19, it is evident that the blurring process induces slight modifications in the filter

responses within the initial and final layers of the convolutional neural network (CNN). Despite these alterations, they are minimal. Specifically, for a degradation level of 5, the Structural Similarity Index (SSIM) of the blurred image amounts to 0.95.

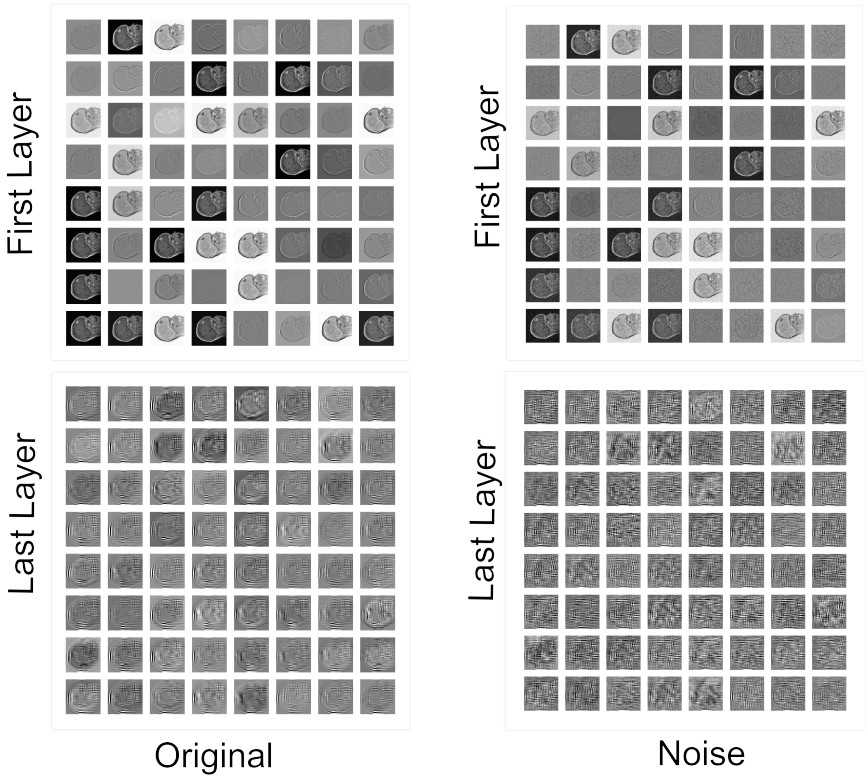


Figure 20: Feature maps from the ResNet50 network on the first and last convolutional layers for original image without artifacts and the same image degraded with noise artifact

In Figure 20, it is evident that noise induces modifications in the filter response within the initial layer of the CNN. The impact is notably pronounced in the final layer, suggesting that minor adjustments in the first layer's response result in more substantial effects in the following layers. Specifically, for a degradation level of 5, the SSIM value for the noisy image is measured at 0.24.

In Figure 21, it is evident that contrast induces notable alterations in the filter response within the initial layer of the CNN. Likewise, substantial changes are observed in the final layer. Contrast triggers a multitude of activations in the initial layer, which consequently impact the responses in the ultimate layer. The Structural Similarity Index (SSIM) for the contrast image is

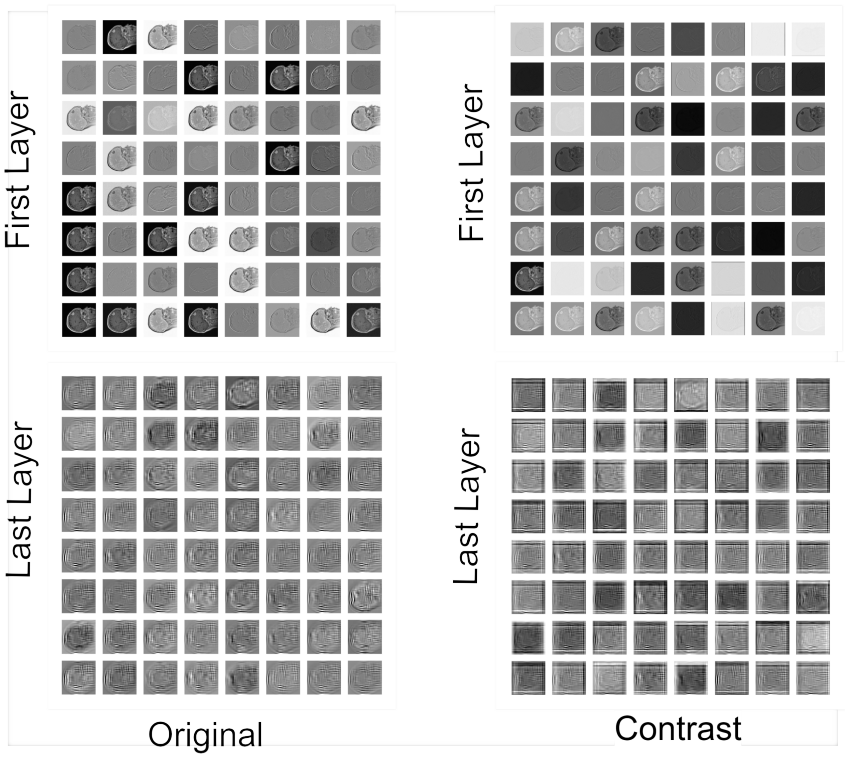


Figure 21: Feature maps from the ResNet50 network on the first and last convolutional layers for original image without artifacts and the same image degraded with contrast artifact

0.56 under level 5 degradation

In Figure 22, it is evident that the presence of ringing induces slight modifications in the filter responses at the initial and final layers of the CNN. Specifically, the alterations are minimal. When considering a degradation level of 5, the SSIM value for the ringing image is measured at 0.98. These observations suggest that the differences in filter outputs between the original image and the image with ringing or blur artifacts are marginal, as indicated by the high SSIM values of the degraded images. Conversely, for noise and contrast, substantial disparities are observed in the filter outputs between the original image and the images with artifacts, which aligns with the low SSIM values of the degraded images.

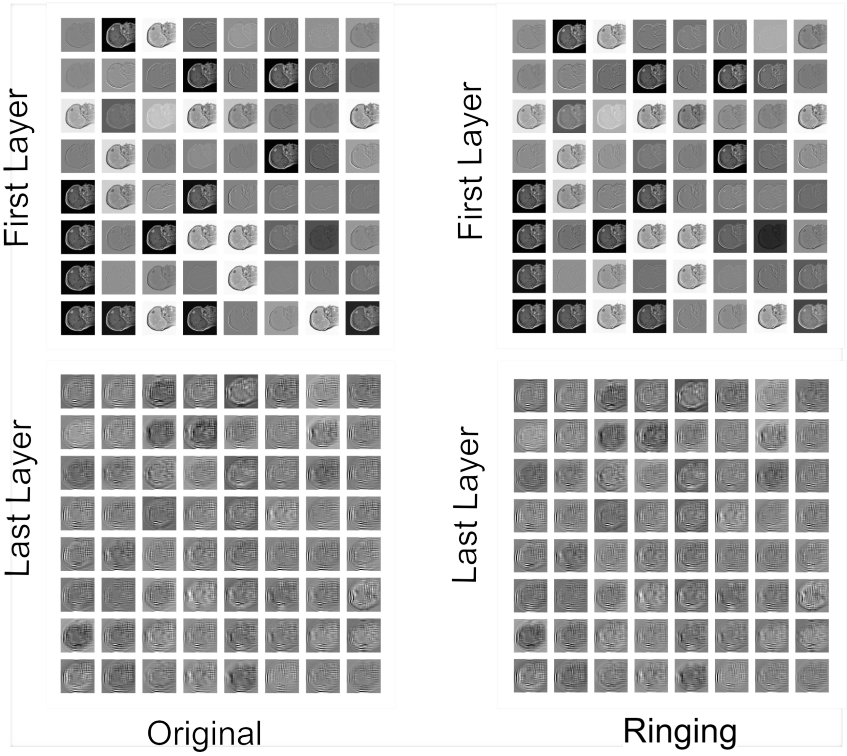


Figure 22: Feature maps from the ResNet50 network on the first and last convolutional layers for original image without artifacts and the same image degraded with ringing artifact.

SSIM Analysis

Up to this point, we examined an image dataset that contained various levels of artifacts. In this section, we constructed a dataset where each of the four artifacts exhibits comparable SSIM values. The SSIM values were grouped into the following intervals: 0.10-0.19, 0.20-0.29, 0.30-0.39, 0.40-0.49, 0.50-0.59, 0.60-0.69, 0.70-0.79, 0.80-0.89, and 0.90-0.99. Next, we produced images with SSIM values that fell within the specified ranges for all four artifacts. While we were not successful in generating images for all SSIM ranges across all artifacts, we were able to do so for the majority of cases. This new dataset underwent testing for the classification of brain MRI scans with meningioma, glioma, or pituitary tumors. This classification task was carried out using ReNet50, InceptionV3, and ResNeXt architectures, which were trained using the original dataset as detailed in [45].

Tables 11, 12, 13, and 14 display the mean results for accuracy, precision, recall, and F1 score concerning MRI images that include blur, noise, contrast, and ringing artifacts. These results are computed for both the initial image and images with different SSIM values, utilizing ResNet50, Inception V3, and Resnext models. The initial column of each table provides an illustration of the images under examination, while the subsequent columns present the metrics alongside their respective values.

Table 11: Values for Accuracy, Precision, Recall, and F1 Score for MRI Images Containing Blurring Artifacts with varying SSIM values using ResNet50, Inception V3 and Resnext models

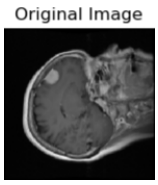
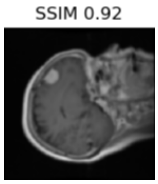
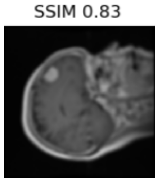
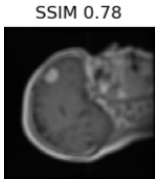
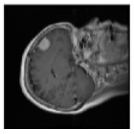
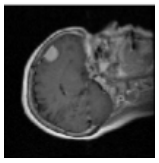
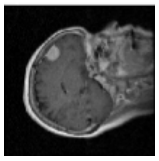
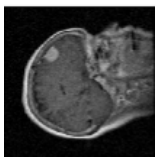
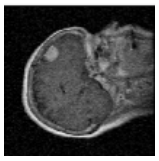
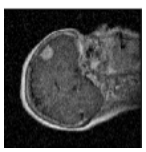
Image	Metrics	ResNet50	InceptionV3	ResNeXt
	Accuracy	94.130	93.695	93.478
	Precision	94.515	95.464	93.899
	Recall	94.130	95.000	93.478
	F1-Score	94.005	94.894	93.311
	Accuracy	89.782	83.043	92.608
	Precision	90.596	85.049	92.664
	Recall	89.782	83.043	92.608
	F1-Score	89.268	83.637	92.485
	Accuracy	85.652	75.869	81.304
	Precision	86.962	80.027	81.928
	Recall	85.652	75.869	81.304
	F1-Score	84.654	77.036	80.357
	Accuracy	81.956	77.173	75.869
	Precision	83.900	79.924	78.369
	Recall	81.956	77.173	75.869
	F1-Score	80.339	78.034	73.997

Table 12: Values for Accuracy, Precision, Recall, and F1 Score for MRI Images Containing Noise Artifacts with varying SSIM values using ResNet50, Inception V3 and Resnext models

Image	Metrics	ResNet50	InceptionV3	ResNeXt
Original Image 	Accuracy	94.130	93.695	93.478
	Precision	94.515	95.464	93.899
	Recall	94.130	95.000	93.478
	F1-Score	94.005	94.894	93.311
SSIM 0.94 	Accuracy	94.347	94.130	93.913
	Precision	94.872	94.114	94.465
	Recall	94.347	94.130	93.913
	F1-Score	94.166	94.069	93.676
SSIM 0.81 	Accuracy	93.478	93.695	90.434
	Precision	94.223	93.8541	91.807
	Recall	93.478	93.695	90.434
	F1-Score	93.140	93.617	89.881
SSIM 0.74 	Accuracy	83.695	91.739	84.565
	Precision	87.720	92.003	88.584
	Recall	83.695	91.739	84.565
	F1-Score	83.196	91.567	83.712
SSIM 0.61 	Accuracy	71.739	90.217	77.391
	Precision	81.481	90.274	85.453
	Recall	71.739	90.217	77.391
	F1-Score	69.920	90.117	76.928
SSIM 0.51 	Accuracy	67.173	88.260	67.826
	Precision	78.654	88.105	81.869
	Recall	67.173	88.260	67.826

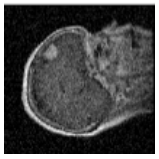
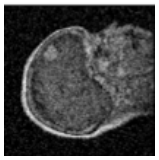
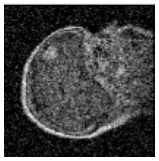
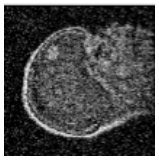
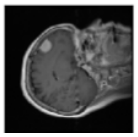
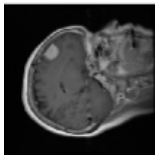
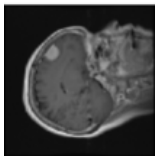
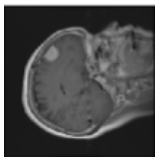
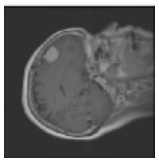
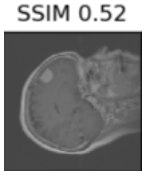
	F1-Score	64.488	88.135	67.032
SSIM 0.43 	Accuracy	65.652	83.043	62.173
	Precision	77.003	83.129	79.510
	Recall	65.652	83.043	62.173
	F1-Score	61.418	83.082	60.783
SSIM 0.33 	Accuracy	61.956	71.304	56.086
	Precision	74.459	73.255	75.610
	Recall	61.956	71.304	56.086
	F1-Score	56.888	71.821	52.672
SSIM 0.21 	Accuracy	54.130	45.217	56.739
	Precision	44.443	60.816	68.925
	Recall	54.130	45.217	56.739
	F1-Score	46.691	38.992	52.089
SSIM 0.17 	Accuracy	44.782	40.869	50.652
	Precision	39.718	50.193	56.618
	Recall	44.782	40.869	50.652
	F1-Score	37.027	30.748	43.995

Table 13: Values for Accuracy, Precision, Recall, and F1 Score for MRI Images Containing Contrast Artifacts with varying SSIM values using ResNet50, Inception V3 and Resnext models

Image	Metrics	ResNet50	InceptionV3	ResNeXt
Original Image 	Accuracy	94.130	93.695	93.478
	Precision	94.515	95.464	93.899
	Recall	94.130	95.000	93.478
	F1-Score	94.005	94.894	93.311
SSIM 0.98 	Accuracy	93.913	90.652	91.956
	Precision	94.292	90.522	92.403
	Recall	93.913	90.652	91.956
	F1-Score	93.788	90.502	91.789
SSIM 0.8 	Accuracy	94.347	92.173	93.043
	Precision	94.848	92.075	93.478
	Recall	94.347	92.173	93.043
	F1-Score	94.193	92.058	92.884
SSIM 0.7 	Accuracy	94.347	93.913	93.695
	Precision	95.005	93.953	94.221
	Recall	994.347	93.913	93.695
	F1-Score	94.128	93.835	93.529
SSIM 0.61 	Accuracy	88.043	93.478	88.913
	Precision	90.384	93.793	91.558
	Recall	88.043	93.478	88.913
	F1-Score	86.710	93.411	87.913
SSIM 0.52 	Accuracy	75.869	87.826	82.608
	Precision	80.730	89.939	87.426
	Recall	75.869	87.826	82.608

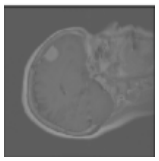
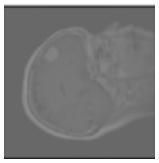
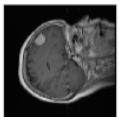
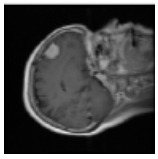
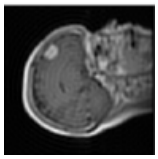
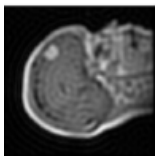
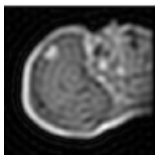
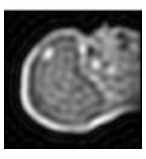
	F1-Score	70.250	87.909	79.653
SSIM 0.41 	Accuracy	52.173	75.869	79.565
	Precision	65.428	81.710	83.752
	Recall	52.173	75.869	79.565
	F1-Score	41.189	76.026	75.860
SSIM 0.35 	Accuracy	47.608	63.913	74.130
	Precision	54.997	69.806	78.229
	Recall	47.608	63.913	74.130
	F1-Score	32.368	63.278	69.698
SSIM 0.23 	Accuracy	23.043	30.869	23.043
	Precision	5.310	9.529	5.310
	Recall	23.043	30.869	23.043
	F1-Score	8.631	14.563	8.631
SSIM 0.19 	Accuracy	51.739	58.695	53.043
	Precision	73.032	58.498	43.096
	Recall	51.739	58.695	53.043
	F1-Score	39.910	56.799	41.529

Table 14: Values for Accuracy, Precision, Recall, and F1 Score for MRI Images Containing Ringing Artifacts with varying SSIM values using ResNet50, Inception V3 and Resnext models

Image	Metrics	ResNet50	InceptionV3	ResNeXt
Original Image 	Accuracy	94.130	93.695	93.478
	Precision	94.515	95.464	93.899
	Recall	94.130	95.000	93.478
	F1-Score	94.005	94.894	93.311
SSIM 0.96 	93.478	90.869	92.173	
	Precision	93.903	91.553	92.563
	Recall	94.565	93.478	92.173
	F1-Score	93.236	90.916	91.995
SSIM 0.82 	Accuracy	85.217	67.826	76.086
	Precision	85.278	77.948	76.866
	Recall	94.347	85.217	76.086
	F1-Score	84.950	67.898	74.370
SSIM 0.71 	Accuracy	70.217	57.173	67.173
	Precision	69.664	76.095	70.173
	Recall	93.913	70.217	67.173
	F1-Score	69.515	55.9459	64.502
SSIM 0.62 	Accuracy	53.913	46.956	56.956
	Precision	55.715	69.753	63.528
	Recall	93.913	53.913	56.956
	F1-Score	52.429	44.331	56.863
SSIM 0.54 	Accuracy	45.652	45.869	45.434
	Precision	52.346	63.484	55.812
	Recall	93.913	45.652	45.434

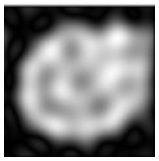
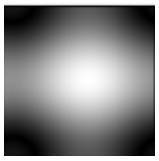
	F1-Score	42.624	38.858	47.057
SSIM 0.41 	Accuracy	34.565	38.478	35.217
	Precision	50.905	21.706	945.514
	Recall	93.260	34.565	35.217
	F1-Score	22.440	27.667	32.809
SSIM 0.30 	Accuracy	23.695	34.130	21.956
	Precision	20.648	18.909	7.276
	Recall	91.521	23.695	21.956
	F1-Score	11.882	24.280	8.879
SSIM 0.26 	Accuracy	23.043	31.086	22.826
	Precision	5.310	32.593	12.3331
	Recall	91.521	23.043	22.826
	F1-Score	8.631	15.018	11.5034

Figure 23(a)-(d) shows the plots of accuracy, precision, recall, and F1 score, respectively, achieved from ResNet50, Inceptionv3, and ResNeXt models for images containing blur artifacts with varying SSIM values. As blurring increases, performance decreases. We do not see a significant decrease in performance as we could only generate 3 blur images. Figure 24(a)-(d) shows the plots of accuracy, precision, recall, and F1 score, respectively, achieved from ResNet50, Inceptionv3 and ResNeXt models for images containing noise artifacts with varying SSIM values. As the noise increases, the performance.

Figure 25(a)-(d) shows the plots of accuracy, precision, recall, and F1 score, respectively, achieved from ResNet50, Inceptionv3 and ResNeXt models for images containing contrast artifacts with varying SSIM values. As the contrast increases, the performance decreases. However, for contrast with SSIM 0.23, the performance values are always significantly lower. To explain this, take a look at the example of a contrast image with SSIM 0.23 in Table 13. The contrast artifact degraded the image so that it is almost impossible to see the content of the image,

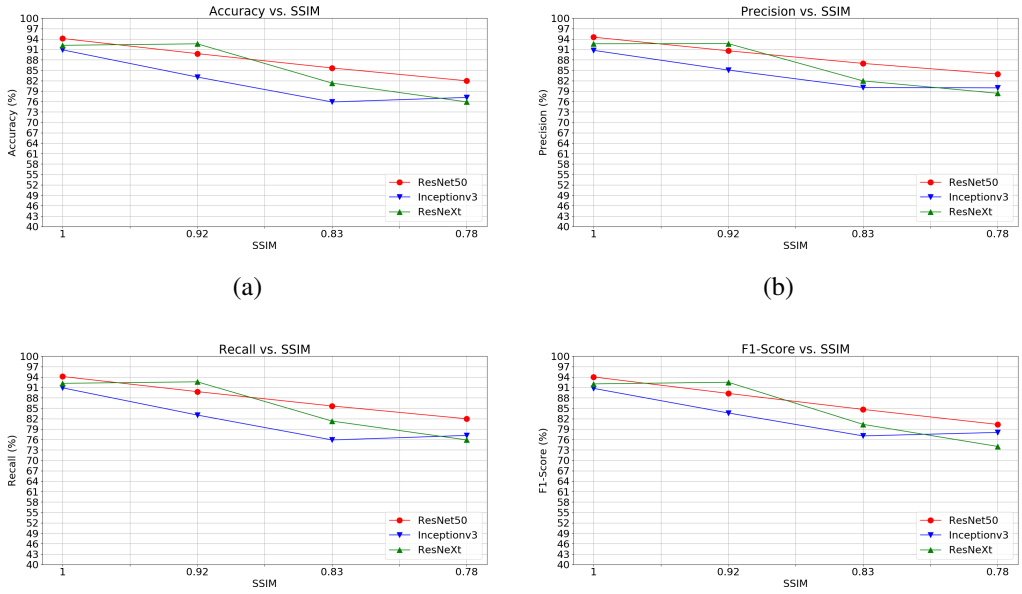


Figure 23: Graphs of values for (a) accuracy, (b) precision, (c) recall, and (d) F1 score achieved from ResNet50, Inceptionv3 and ResNeXt models for images containing blur artifacts with varying SSIM values

which is the reason behind its performance. Figure 26(a)-(d) shows the plots of accuracy, precision, recall, and F1 score, respectively, achieved from ResNet50, Inceptionv3 and ResNeXt models for images containing ringing artifacts with varying SSIM values. As the ringing increases, the performance decreases significantly. Therefore, as SSIM increases, performance almost always decreases.

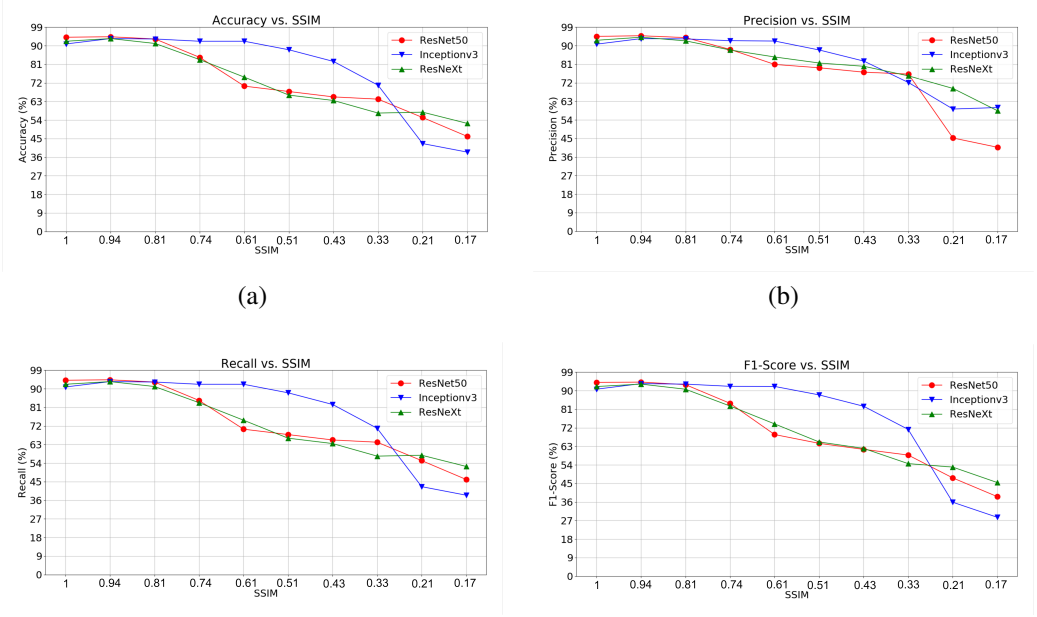


Figure 24: Graphs of values for (a) accuracy, (b) precision, (c) recall, and (d) F1 score achieved from ResNet50, Inceptionv3 and ResNeXt models for images containing noise artifacts with varying SSIM values

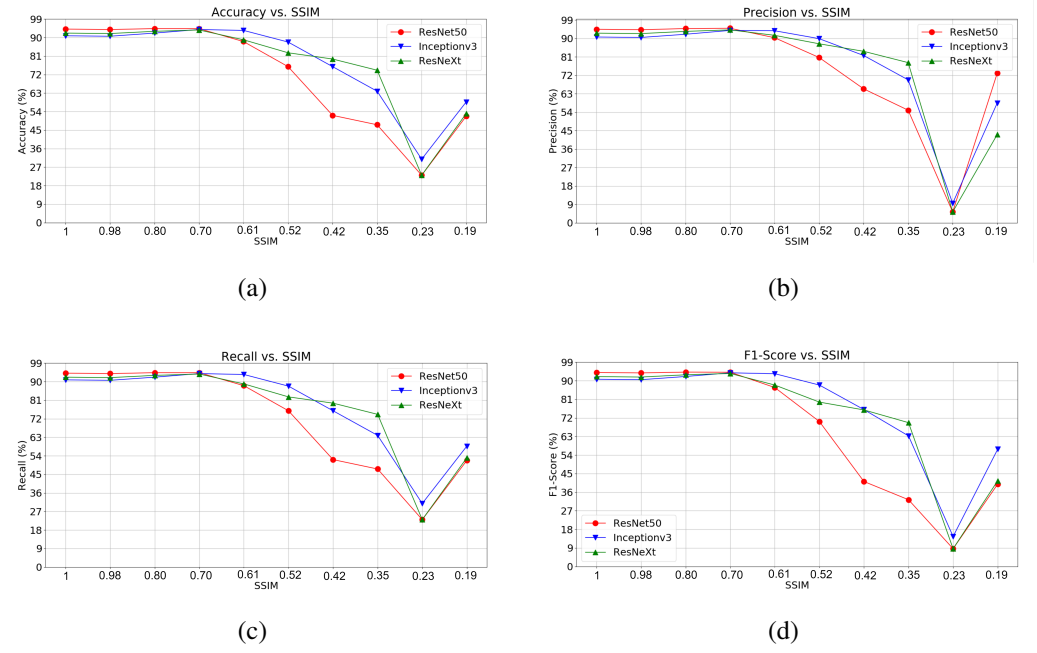


Figure 25: Graphs of values for (a) accuracy, (b) precision, (c) recall, and (d) F1 score achieved from ResNet50, Inceptionv3 and ResNeXt models for images containing contrast artifacts with varying SSIM values

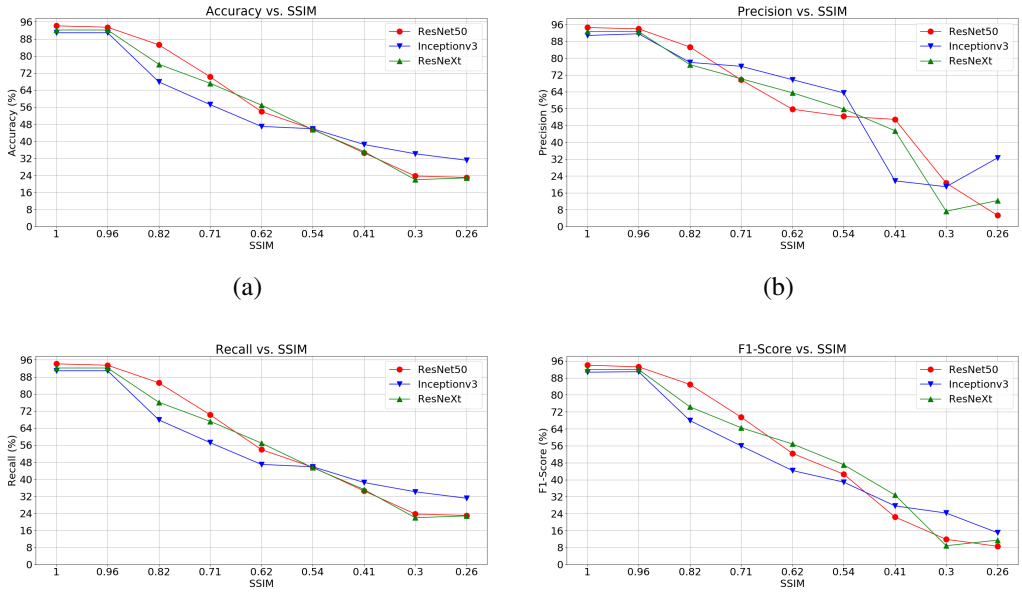


Figure 26: Graphs of values for (a) accuracy, (b) precision, (c) recall, and (d) F1 score achieved from ResNet50, Inceptionv3 and ResNeXt models for images containing ringing artifacts with varying SSIM values

VII. CONCLUSION AND FUTURE WORK

Conclusion

In this study, we examined how the quality of medical image inputs influences the efficacy of three CNN-based systems designed for brain tumor classification. Our objective was to assess the impact of introducing artifacts like blurring, noise, ringing, and contrast into MRI scans on the detection of brain tumors. Additionally, we explored the influence of these four artifacts when the Structural Similarity Index (SSIM) values were within a similar range. We computed performance metrics (accuracy, precision, recall, and F1 score) for 10 intensity levels of each of the five common MRI artifacts. Furthermore, we assessed these metrics under conditions where the image quality for these artifacts resulted in similar SSIM values, aiming to understand the individual effects of each artifact on cancer detection. The results revealed a substantial impact of these artifacts on performance metrics when images were impaired by artifacts.

Future Work

Our future work aims to investigate the effects of multiple artifacts occurring simultaneously in each image. we also plan to evaluate more CNN architectures to gain more insight on how the degradation caused by artifacts affect neural networks. We plan to extract feature maps for all the evaluated models to determine now each artifact affect each network. Furthermore, we want to look at possible solutions toh elp CNN make better diagnosis when the input image contains artifacts.

REFERENCES

[1] L. M. DeAngelis, "Brain tumors," *New England journal of medicine*, vol. 344, no. 2, pp. 114–123, 2001.

[2] M. Bernstein, M. Berger, A. A. of Neurological Surgeons. Joint Tumor Section, and C. of Neurological Surgeons, *Neuro-oncology: The Essentials*, ser. Thieme Publishers Series. Thieme Medical Publishers, 2000. [Online]. Available: https://books.google.com/books?id=gbwSC43Eia4C

[3] P. Kanavos, "The rising burden of cancer in the developing world," *Annals of Oncology*, vol. 17, pp. viii15–viii23, 2006, cancer Initiatives in Developing Countries., 30 October 2005: Paris, France. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0923753419414233

[4] F. Bray, M. Laversanne, E. Weiderpass, and I. Soerjomataram, "The ever-increasing importance of cancer as a leading cause of premature death worldwide," *Cancer*, vol. 127, no. 16, pp. 3029–3030, 2021.

[5] P. Brennan and G. Davey-Smith, "Identifying Novel Causes of Cancers to Enhance Cancer Prevention: New Strategies Are Needed," *JNCI: Journal of the National Cancer Institute*, vol. 114, no. 3, pp. 353–360, 11 2021. [Online]. Available: https://doi.org/10.1093/jnci/djab204

[6] G. B. of Disease 2019 Cancer Collaboration, "Cancer Incidence, Mortality, Years of Life Lost, Years Lived With Disability, and Disability-Adjusted Life Years for 29 Cancer Groups From 2010 to 2019: A Systematic Analysis for the Global Burden of Disease Study 2019," *JAMA Oncology*, vol. 8, no. 3, pp. 420–444, 03 2022. [Online]. Available: https://doi.org/10.1001/jamaoncol.2021.6987

[7] I. Ilic and M. Ilic, "International patterns and trends in the brain cancer incidence and mortality: An observational study based on the global burden of disease," *Heliyon*, vol. 9, p. e18222, Jul 2023, the authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. [Online]. Available: https://doi.org/10.1016/j.heliyon.2023.e18222

[8] J. Ferlay, M. Colombet, and F. Bray, "Cancer incidence in five continents, ci5plus: Iarc cancerbase no. 9," *Lyon, France: International Agency for Research on Cancer*, 2018.

[9] C. Fitzmaurice, C. Allen, R. M. Barber, L. Barregard, Z. A. Bhutta, H. Brenner, D. J. Dicker, O. Chimed-Orchir, R. Dandona, L. Dandona *et al.*, "Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life-years for 32 cancer groups, 1990 to 2015: a systematic analysis for the global burden of disease study," *JAMA oncology*, vol. 3, no. 4, pp. 524–548, 2017.

[10] D. N. Louis, A. Perry, P. Wesseling, D. J. Brat, I. A. Cree, D. Figarella-Branger, C. Hawkins, H. Ng, S. M. Pfister, G. Reifenberger *et al.*, "The 2021 who classification of tumors of the central nervous system: a summary," *Neuro-oncology*, vol. 23, no. 8, pp. 1231–1251, 2021.

[11] U. Raghavendra, A. Gudigar, A. Paul, T. Goutham, M. A. Inamdar, A. Hegde, A. Devi, C. P. Ooi, R. C. Deo, P. D. Barua, F. Molinari, E. J. Ciaccio, and U. R. Acharya, "Brain tumor detection and screening using artificial intelligence techniques: Current trends and future perspectives," *Computers in Biology and Medicine*, vol. 163, p. 107063, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010482523005280

[12] S. Ammari, S. Pitre-Champagnat, L. Dercle, E. Chouzenoux, S. Moalla, S. Reuze, H. Talbot, T. Mokoyoko, J. Hadchiti, S. Diffetocq *et al.*, "Influence of magnetic field strength on magnetic resonance imaging radiomics features in brain imaging, an in vitro and in vivo study," *Frontiers in Oncology*, vol. 10, p. 541663, 2021.

[13] G. Pradhan, S. Morris, and N. Nayak, *Advances in Electrical Control and Signal Systems: Select Proceedings of AECSS 2019*. Springer, 2020.

[14] M. S. Mahaley, C. Mettlin, N. Natarajan, E. R. Laws, and B. B. Peace, "National survey of patterns of care for brain-tumor patients," *Journal of neurosurgery*, vol. 71, no. 6, pp. 826–836, 1989.

[15] R. M. Hayward, N. Patronas, E. H. Baker, G. Vézina, P. S. Albert, and K. E. Warren, "Inter-observer variability in the measurement of diffuse intrinsic pontine gliomas," *Journal of neuro-oncology*, vol. 90, pp. 57–61, 2008.

[16] N. V. Dhole and V. V. Dixit, "Review of brain tumor detection from mri images with hybrid approaches," *Multimedia tools and applications*, vol. 81, no. 7, pp. 10 189–10 220, 2022.

[17] R. Kaifi, "A review of recent advances in brain tumor diagnosis based on ai-based classification," *Diagnostics*, vol. 13, no. 18, p. 3007, 2023.

[18] C. Noda, B. Ambale Venkatesh, J. D. Wagner, Y. Kato, J. M. Ortman, and J. A. Lima, "Primer on commonly occurring mri artifacts and how to overcome them," *Radiographics*, vol. 42, no. 3, pp. E102–E103, 2022.

[19] H. Liu, H. Li, X. Wang, H. Li, M. Ou, L. Hao, Y. Hu, and J. Liu, "Understanding how fundus image quality degradation affects cnn-based diagnosis," in *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2022, pp. 438–442.

[20] R. A. Hazarika, A. Abraham, D. Kandar, and A. K. Maji, "An improved lenet-deep neural network model for alzheimerâs disease classification using brain magnetic resonance images," *IEEE Access*, vol. 9, pp. 161 194–161 207, 2021.

[21] M. Goyal, J. Guo, L. Hinojosa, K. Hulsey, and I. Pedrosa, "Automated kidney segmentation by mask r-cnn in t2-weighted magnetic resonance imaging," in *Medical Imaging 2022: Computer-Aided Diagnosis*, vol. 12033. SPIE, 2022, pp. 789–794.

[22] Y. Wang, E. J. Choi, Y. Choi, H. Zhang, G. Y. Jin, and S.-B. Ko, "Breast cancer classification in automated breast ultrasound using multiview convolutional neural network with transfer learning," *Ultrasound in medicine & biology*, vol. 46, no. 5, pp. 1119–1132, 2020.

[23] Q. Lin, T. Li, C. Cao, Y. Cao, Z. Man, and H. Wang, "Deep learning based automated diagnosis of bone metastases with spect thoracic bone images," *Scientific Reports*, vol. 11, no. 1, p. 4223, 2021.

[24] G. Latif, G. Ben Brahim, D. A. Iskandar, A. Bashar, and J. Alghazo, "Glioma tumorsâ classification using deep-neural-network-based features with svm classifier," *Diagnostics*, vol. 12, no. 4, p. 1018, 2022.

[25] M. S. Majib, M. M. Rahman, T. S. Sazzad, N. I. Khan, and S. K. Dey, "Vgg-scnet: A vgg net-based deep learning framework for brain tumor detection on mri images," *IEEE Access*, vol. 9, pp. 116 942–116 952, 2021.

[26] A. S. Methil, "Brain tumor detection using deep learning and image processing," in *2021 international conference on artificial intelligence and smart systems (ICAIS)*. IEEE, 2021, pp. 100–108.

[27] A. Çinar and M. Yildirim, "Detection of tumors on brain mri images using the hybrid convolutional neural network architecture," *Medical hypotheses*, vol. 139, p. 109684, 2020.

[28] R. Vankdothu, M. A. Hameed, and H. Fatima, "A brain tumor identification and classification using deep learning based on cnn-lstm method," *Computers and Electrical Engineering*, vol. 101, p. 107960, 2022.

[29] A. K. Anaraki, M. Ayati, and F. Kazemi, "Magnetic resonance imaging-based brain tumor grades classification and grading via convolutional neural networks and genetic algorithms," *biocybernetics and biomedical engineering*, vol. 39, no. 1, pp. 63–74, 2019.

[30] H. ZainEldin, S. A. Gamel, E.-S. M. El-Kenawy, A. H. Alharbi, D. S. Khafaga, A. Ibrahim, and F. M. Talaat, "Brain tumor detection and classification using deep learning and sine-cosine fitness grey wolf optimization," *Bioengineering*, vol. 10, no. 1, p. 18, 2022.

[31] H. Yahyaoui, F. Ghazouani, and I. R. Farah, "Deep learning guided by an ontology for medical images classification using a multimodal fusion," in *2021 International Congress of Advanced Technology and Engineering (ICOTEN)*. IEEE, 2021, pp. 1–6.

[32] M. S. I. Khan, A. Rahman, T. Debnath, M. R. Karim, M. K. Nasir, S. S. Band, A. Mosavi, and I. Dehzangi, "Accurate brain tumor detection using deep convolutional neural network," *Computational and Structural Biotechnology Journal*, vol. 20, pp. 4733–4745, 2022.

[33] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," *arXiv preprint arXiv:1412.6572*, 2014.

[34] C. Hu, B. B. Sapkota, J. A. Thomasson, and M. V. Bagavathiannan, "Influence of image quality and light consistency on the performance of convolutional neural networks for weed mapping," *Remote Sensing*, vol. 13, no. 11, p. 2140, 2021.

[35] V. Thambawita, I. Strümke, S. A. Hicks, P. Halvorsen, S. Parasa, and M. A. Riegler, "Impact of image resolution on deep learning performance in endoscopy image classification: an experimental study using a large dataset of endoscopic images," *Diagnostics*, vol. 11, no. 12, p. 2183, 2021.

[36] C. F. Sabottke and B. M. Spieler, "The effect of image resolution on deep learning in radiography," *Radiology: Artificial Intelligence*, vol. 2, no. 1, p. e190015, 2020.

[37] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Transactions on image processing*, vol. 15, no. 2, pp. 430–444, 2006.

[38] S. Basu, M. Karki, S. Ganguly, R. DiBiano, S. Mukhopadhyay, S. Gayaka, R. Kannan, and R. Nemani, "Learning sparse feature representations using probabilistic quadtrees and deep belief nets," *Neural Processing Letters*, vol. 45, pp. 855–867, 2017.

[39] S. Dodge and L. Karam, "Understanding how image quality affects deep neural networks," in *2016 eighth international conference on quality of multimedia experience (QoMEX)*. IEEE, 2016, pp. 1–6.

[40] K. Grm, V. Štruc, A. Artiges, M. Caron, and H. K. Ekenel, "Strengths and weaknesses of deep learning models for face recognition against image degradations," *Iet Biometrics*, vol. 7, no. 1, pp. 81–89, 2018.

[41] K. Krupa and M. Bekiesińska-Figatowska, "Artifacts in magnetic resonance imaging," *Polish journal of radiology*, vol. 80, p. 93, 2015.

[42] B. Koonce, *ResNet 50*. Berkeley, CA: Apress, 2021, pp. 63–72. [Online]. Available: https://doi.org/10.1007/978-1-4842-6168-2_6

[43] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818–2826.

[44] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1492–1500.

[45] J. Cheng, W. Yang, M. Huang, W. Huang, J. Jiang, Y. Zhou, R. Yang, J. Zhao, Y. Feng, Q. Feng *et al.*, "Retrieval of brain tumors by adaptive spatial pooling and fisher vector representation," *PloS one*, vol. 11, no. 6, p. e0157112, 2016.

[46] D. Kawahara and Y. Nagata, "T1-weighted and t2-weighted mri image synthesis with convolutional generative adversarial networks," *reports of practical Oncology and radiotherapy*, vol. 26, no. 1, pp. 35–42, 2021.

[47] J. Cheng, W. Huang, S. Cao, R. Yang, W. Yang, Z. Yun, Z. Wang, and Q. Feng, "Enhanced performance of brain tumor classification via tumor region augmentation and partition," *PloS one*, vol. 10, no. 10, p. e0140381, 2015.

[48] M. Farias, P. de Castro Oliveira, G. dos Santos Lopes, C. Miosso, and J. Lima, "The influence of magnetic resonance imaging artifacts on cnn-based brain cancer detection algorithms," *Computational Mathematics and Modeling*, vol. 33, no. 2, pp. 211–229, 2022.

[49] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2000.

[50] S. V. der Walt, J. L. SchÃ¶nberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, N. Yager, E. Gouillart, and T. Yu, "scikit-image: image processing in python," *PeerJ*, vol. 2, p. e453, jun 2014.

[51] E. Kellner, B. Dhital, V. G. Kiselev, and M. Reisert, "Gibbs-ringing artifact removal based on local subvoxel-shifts," *Magnetic resonance in medicine*, vol. 76, no. 5, pp. 1574–1581, 2016.

[52] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[53] Z. Wang and A. C. Bovik, "Mean squared error: Love it or leave it? a new look at signal fidelity measures," *IEEE signal processing magazine*, vol. 26, no. 1, pp. 98–117, 2009.