Independent University **Bangladesh (IUB)**

IUB Academic Repository

Computer Science and Engineering | Undergraduate Thesis

2024-12

Comparative Analysis of Deep Learning Models for Intracerebral Hemorrhage Detection from Computed Tomography

Ahad, Taiseer Rakiin

IUB

http://ar.iub.edu.bd/handle/11348/989

Downloaded from IUB Academic Repository



Comparative Analysis of Deep Learning Models for Intracerebral Hemorrhage Detection from Computed Tomography

Taiseer Rakiin Ahad (1911242)

Farhan Zaman Ha-mim (2022157)

Md. Ashaduzzaman Nur (1910397)

Dissertation submitted in partial fulfillment for the Degree of
Bachelor of Science in Computer Science and Engineering
Department of Computer Science & Engineering
Independent University, Bangladesh

December 2024

Attestation

I understand the nature of plagiarism and am aware of the university's strict policy on the matter. I certify that this is my original work. However, others' written work or software used in any part of this project is properly cited following internationally accepted academic guidelines.

Signature:

I. Taiseer Rakiin Ahad: _____

II. Farhan Zaman Ha-mim: _____

III. Md. Ashaduzzaman Nur: _____

Evaluation Committee

Supervision Panel

.....

Supervisor Co-Supervisor

External Examiners

.....

External Examiner 1

Internal Examiners

.....

Internal Examiner 1 Internal Examiner 2

Office Use

.....

Program Coordinator Head of the Department

.....

Director Graduate Studies, Research and Industrial Relations

.....

Library

Acknowledgment

We are deeply thankful to our supervisor, Dr. Saadia Binte Alam, Head of the Department of Computer Science & Engineering at Independent University Bangladesh, for her immense guidance and support. We also express our deepest gratitude to our advisor, Dr. MD Rashedur Rahman, a Doctoral Fellow at the University of Hyogo, Japan, and Adjunct Assistant Professor at Independent University, Bangladesh. With his expert guidance, insightful feedback, and unwavering support, we completed this dissertation.

Additionally, we sincerely thank Mr. Siam Tahsin Bhuiyan and Mr. Sefatul Wasi from Independent University Bangladesh for their valuable suggestions and contributions throughout the experiments. Their assistance and encouragement were invaluable throughout this research.

List of Publications:

T. R. Ahad, S. B. Alam, F. Hamim, A. Nur, S. Wasi, and M. R. Rahman, "Comparative Analysis of Deep Learning Models for Intracerebral Hemorrhage Detection from Computed Tomography," in *Proc. 2024 27th International Conference on Computer and Information Technology (ICCIT)*, Cox's Bazar, Bangladesh, Dec. 20-22, 2024. (Accepted)

Abstract

Intracerebral hemorrhage (ICH) is a serious medical condition caused by bleeding within the brain tissue, which can lead to life-threatening complications if not diagnosed quickly. This study compares the performance of three advanced deep learning models, InceptionV3, DenseNet201, and EfficientNetB4, for detecting ICH from computed tomography (CT) scans. This work's novelty lies in preprocessing techniques such as windowing with Hounsfield Units and Skull segmentation to provide cleaner input data and Grad-CAM visualizations to ensure the model's predictions are interpretable. Using comprehensive evaluations, including metrics like AUROC, precision, and F1 score, EfficientNetB4 achieved the best results with an accuracy of 0.85, precision of 0.87, and F1-score of 0.86, demonstrating its potential to assist medical professionals in making faster and more reliable diagnoses.

Table of Contents

Chapter-1: Introduction 1
1.1. Background and Context 1
1.2. Objectives 5
1.3. Scope and Limitations 6
1.4. Outline of Thesis 7
Chapter-2: Literature Review 8
Chapter-3: Dataset 12
3.1. Defining the dataset 12
3.2. Types of Intracerebral Hemorrhages 13
3.3 Handling Class Imbalance 16
3.4 Comparison with Other Medical Imaging Datasets 18
Chapter-4: Methods 19
4.1 Pre-Processing and Data Augmentation 19
4.1.1 Data Augmentation 19
4.1.2 Windowing Techniques 20
4.1.3 Resizing and Batch Processing 22
4.1.4 Segmentation and Skull Extraction 22
4.2 Model 24
4.2.1 Model Layers 24
4.2.2 Training Procedure 25
4.3Evaluation Metrics 25
4.3.1 ROC Curve and AUC 27
4.3.2 Grad-CAM Heatmaps 28
Chapter-5: Results and Discussion 29
5.1 Architectural Significance of EfficientNetB4 30
5.2 Comparison with DenseNet201 31
5.3 Comparison with InceptionV3 31
5.4 Performance Analysis 34
Chapter-6: Conclusion and Future Work 39
6.1 Summary 39
6.2 Challenges and Limitations 39
6.3 Future Directions 40
Chapter-7: References 41

List of Figures

FIGURE 3.1: Types of Hemorrhages in the RSNA dataset.....13

FIGURE 3.2: Illustration of how the dataset was split.....15

FIGURE 3.3: Samples of (a) Normal, (b) Multiple, (c) Single training class images.....17

FIGURE 4.1: Combination of window channels to form RGB CT slice.....21

FIGURE 4.2: Segmentation of skull region.....23

FIGURE 4.3: Samples of CT slices used for training.....23

FIGURE 4.4: Samples of GRAD-Cam image.....28

FIGURE 5.1: EfficientNetB4 plot history.....32

FIGURE 5.2: DenseNet201 plot history.....33

FIGURE 5.3: InceptionV3 plot history.....33

FIGURE 5.4: Confusion Matrices of (a) EfficientNetB4, (b) DenseNet201, (C) InceptionV3.....35

FIGURE 5.5: ROC curves of all models.....36

FIGURE 5.6: EfficientNetB4 Grad-CAM Heatmap.....37

FIGURE 5.7: DenseNet201 Grad-CAM heatmap.....38

FIGURE 5.8: InceptionV3 Grad-CAM heatmap.....38

List of Tables

TABLE 3.1: Sample list of IDs and their corresponding labels in the dataset……..……..……..14

TABLE 5.1: Initial training Metrics.....29

TABLE 5.2: Fine-Tuned Metrics.....29

Chapter 1: Introduction

1.1. Background and Context

Intracerebral hemorrhage (ICH) is a serious medical condition caused by bleeding within the brain tissue [1], which increases pressure in the brain, obstructs blood flow, and reduces oxygen supply. This can cause damage to brain cells, leading to severe outcomes such as loss of consciousness, disability, or even death [2]. Computed Tomography (CT) scans are a key tool for diagnosing ICH, as they provide rapid brain imaging, allowing for early detection of hemorrhage and guidance in treatment decisions. Magnetic Resonance Imaging (MRI) is also used in some instances for detailed brain imaging, though it is less commonly used in emergency settings because of its longer processing time.

ICH manifests in several types, each with distinct characteristics, causes, and CT scan appearances:

1. **Epidural Hematoma**: Epidural Hematoma is an extra-axial blood collection within the potential space between the dura mater's outer layer and the skull's inner table [3]. It appears as a biconvex (lens-shaped) mass on CT scans, often localized to a specific area and is characterized by the fact that it does not cross suture lines due to the rigid nature of the sutures. Symptoms can include headaches, confusion, dizziness, and loss of consciousness. Rapid intervention is crucial, as untreated epidural hematomas can lead to increased intracranial pressure and severe brain damage.

2. **Intraparenchymal Hemorrhage**: Intraparenchymal hemorrhages are bleeding into the brain parenchyma, which results in conditions like hypertension, arteriovenous malformations, or traumatic injuries [4]. On CT scans, they present themselves as irregular areas of hyper-density, which may be surrounded by edema. The location and volume of the hemorrhage significantly influence the patient's clinical status and prognosis, with more extensive or deeper bleedings typically leading to more severe neurological disorders.

3. **Intravascular Hemorrhage**: This type of hemorrhage occurs within the vascular structures of the brain and can often be difficult to visualize on standard CT scans, as it may be indistinguishable from normal blood vessels. Advanced imaging techniques or expert interpretation are usually required for accurate diagnosis. It may occur as a result of traumatic injury, vascular malformations, or coagulopathy. Clinical implications can range from minimal symptoms to severe complications, depending on the extent and location of the bleeding.

4. **Subdural Hematoma**: A Subdural hematoma occurs when a blood vessel in the space between the skull and the subdural space of the brain is damaged [5]. On CT scans, it appears as a crescent-shaped hyperdense area that can cross suture lines. Symptoms may include headaches, confusion, or altered consciousness, which can be acute or chronic. Treatment often involves surgical intervention to evacuate the hematoma and relieve pressure on the brain.

5. **Subarachnoid Hemorrhage**: Subarachnoid hemorrhages are life-threatening and result from the accumulation of blood between the arachnoid and pia mater [6]. Subarachnoid hemorrhage often presents itself as hyperdense areas along the surface of the brain and within the sulci on CT scans. Commonly caused by ruptured cerebral aneurysms or arteriovenous malformations, it can lead to life-threatening complications such as vasospasm or rebleeding. Immediate medical attention is required to manage the underlying cause and prevent further complications.

6. **Multiple Hemorrhages**: In some cases, patients may suffer from multiple types of hemorrhages simultaneously, complicating diagnosis and management. For example, a patient may experience both subdural and subarachnoid hemorrhages following a traumatic event. Hemodynamic change after solitary ICH or vessel irregularity by untreated hypertension, cerebral amyloid angiopathy, and microhemorrhage may cause simultaneous multiple ICH [7]. This condition requires comprehensive evaluation, as the presence of multiple hemorrhages increases the risk of neurological impairment and complicates

surgical intervention strategies.

7. **Intraventricular Hemorrhage (IVH)**: Intraventricular hemorrhage (IVH) is bleeding into the fluid-filled areas, or ventricles, surrounded by the brain [8]. Premature babies are most commonly affected by this, and the risk of IVH increases with the size and prematureness of the baby. This is because the brain's blood arteries are still developing and very delicate in premature children. IVH is uncommon at birth and typically shows during the first few days of life.

8. **Cerebellar Hemorrhage**: Cerebellar hemorrhage or hematoma is a type of intracerebral hemorrhage in which the bleeding is located in the posterior fossa or cerebellum [9]. Cerebellar hemorrhages can be caused by hypertension, trauma, or blood vessel malformations. On CT, it is often localized to the posterior fossa region and can cause brainstem compression, leading to severe symptoms such as ataxia, dizziness, and altered consciousness. Surgical intervention may be necessary, especially if there is a significant mass effect.

9. **Pontine Hemorrhage**: Pontine hemorrhages are a common form of ICH and usually are a result of poorly controlled long-standing hypertension [10]. It is often associated with trauma and presents a high risk for mortality due to the critical functions regulated by the brainstem, including respiration and heart rate. On CT, it appears as a hyperdense lesion in the pons. The diagnosis is generally poor, and management focuses on supportive care and controlling the underlying causes.

10. **Traumatic Subarachnoid Hemorrhage (TSAH)**: TSAH is an adverse prognostic factor leading to progressive neurological deterioration and increased morbidity and mortality [11]. It presents as blood in the subarachnoid spaces on CT scans. While TSAH may be less catastrophic than aneurysmal subarachnoid hemorrhage, it still poses risks for seizures, meningitis, or delayed complications. Management typically involves monitoring for complications and addressing any underlying injuries.

11. **Microbleeds**: Microbleeds are small and can be detected in various clinical conditions, the most common being cerebral amyloid angiopathy and chronic hypertension [12]. They are best detected using advanced MRI techniques, showing up as small, round areas sensitive to changes. Microbleeds are linked to a higher risk of severe strokes or memory problems, so patients with known vascular risk factors need to be closely monitored.

12. **Duret Hemorrhages**: Duret hemorrhages are usually small hemorrhages within the medulla or pons caused by descending transtentorial herniation [13]. This condition is often linked to severe head trauma. Duret hemorrhages are marked by multiple small areas of bleeding scattered throughout the brainstem. Unfortunately, these types of hemorrhages can be fatal because they disrupt essential functions managed by the brainstem.

13. **Aneurysmal Subarachnoid Hemorrhage (aSAH)**: A type of stroke occurs when an intracranial aneurysm ruptures, causing bleeding in the subarachnoid space [14]. It is a life-threatening condition that requires emergency treatment, often including surgical clipping or endovascular coiling of the aneurysm. On CT, aSAH appears hyperdensity in the subarachnoid spaces, usually accompanied by vasospasm and rebleeding risks. Timely intervention is essential for improving outcomes.

14. **Cavernous Malformation Bleeds**: Cavernous Malformation is a collection of small blood vessels (capillaries) in the central nervous system that is enlarged and irregular in structure [15]. In MRI scans, these malformations appear as bright spots surrounded by darker areas, indicating the presence of blood deposits. If a patient shows symptoms or is at risk of further bleeding, treatment may involve surgical removal of the malformation.

As the global population ages and the occurrence of hypertension and related risk factors increases, ICH is becoming more common, particularly in regions with a high incidence of cardiovascular diseases [16]. ICH accounts for 10-15% of all stroke cases worldwide, though this percentage can reach as high as 24% in East Asian countries such as Japan and Korea due to various environmental and genetic factors [17].

Despite advancements in imaging technologies, diagnosing ICH remains complex. Symptoms often overlap with other neurological conditions, and the availability of expert radiologists is limited in many areas. Early and accurate diagnosis is critical to avoid worsening the condition, as delays can result in further bleeding, infection, or complications such as hydrocephalus. This challenge highlights the need for automated systems that can assist in detecting hemorrhages from brain scans with high precision.

Recent advances in artificial intelligence, particularly in deep learning, have shown promise in enhancing the accuracy and speed of medical diagnoses, including the detection of ICH. Deep learning models have been applied to CT scans to automatically detect hemorrhages, often performing comparably to or surpassing human experts in controlled settings [19]. However, several limitations still need to be addressed, such as the need for large, well-annotated datasets, variations in image quality, and model dependency on expert-provided labels [25]. Integrating machine learning into clinical practice is still an evolving field, with ongoing research required to overcome these challenges and make such systems widely usable.

1.2. Objectives

The primary objective of this study is to evaluate and compare the performance of three state-of-the-art deep learning models—InceptionV3 [21], DenseNet201 [22], and EfficientNetB4 [23]—for detecting ICH from CT scans. These models were trained using the publicly available RSNA Intracranial Hemorrhage dataset on Kaggle [24], which provides a large and diverse set of brain CT scans labeled with six different types of hemorrhages.

The models were assessed using various metrics such as accuracy, precision, recall, F1 score, and Area Under the Receiver Operating Characteristic (AUROC) curve, which measure their overall performance in three classifications (Single Hemorrhages, Multiple Hemorrhages, and No Hemorrhages). To better understand the models' behavior, we also used confusion matrices and analyzed the models' training histories to check for potential issues like overfitting. Additionally, Gradient-weighted Class Activation Mapping (Grad-CAM) images were generated to visually

inspect whether the models focused on the correct areas within the scans.

The goal is to identify which of the three models is best suited for ICH detection and provide insights into their strengths and limitations.

1.3. Scope and Limitations

Automated assistive systems in ICH identification have the potential to revolutionize if existing limitations can be appropriately addressed. Due to the shortage of ICH cases and the need for expert annotations, collecting large datasets for training and validation can be difficult and represents one of the primary challenges. Furthermore, variations in image quality and distortions can impact deep learning model performance. Moreover, unconventional CT scan techniques and variations in patient characteristics and medical conditions may affect the ability to generalize the models. These limitations highlight the necessity for future research in this sector to solve these problems and increase the accuracy and reliability of deep learning-based ICH detection systems.

The performance of each of the three selected deep learning models in this research is thoroughly examined, focusing on a comparative study of the InceptionV3, DenseNet201, and EfficientNetB4 architectures. The research explores the unique characteristics of each model's operation, focusing on how mobile inverted bottleneck layers, squeeze-and-excitation layers, and compound scaling affect the accuracy of ICH detection. Furthermore, the research explores the significance of preprocessing methods, such as Hounsfield Unit windowing and skull segmentation, in increasing model performance. This study intends to contribute to developing a strong and dependable computer-aided diagnosis system for ICH detection by thoroughly analyzing each model's advantages and disadvantages.

1.4. Outline of Thesis

This thesis is organized into seven chapters:

- **Chapter 1:** Introduction - Provides a comprehensive overview of ICH, the motivation for this research, and the challenges associated with automating ICH detection using deep learning models.

- **Chapter 2:** Literature Review - Summarizes previous work and studies in deep learning applied to medical imaging, with a particular focus on ICH detection.

- **Chapter 3:** Dataset—Provides comprehensive details about the dataset used in this study while also comparing it with similar datasets available.

- **Chapter 4:** Methods - Details the dataset preparation, preprocessing techniques, and the architectures of the InceptionV3, DenseNet201, and EfficientNetB4 models used in this study.

- **Chapter 5:** Results and Discussion - Presents the experimental results, comparing the performance of the three models in detecting ICH and their relevant differences. It also includes visualizations of model performance (accuracy, precision, recall, F1 score, AUROC curves, and Grad-CAM heatmaps).

- **Chapter 6:** Conclusion and Future Work summarizes the study's findings and suggests potential areas for future research, including improving the models and expanding the dataset.

- **Chapter 7:** References – A list of all referenced papers throughout the study.

Chapter 2: Literature Review

ICH detection primarily depended on radiologists manually evaluating CT scans before developing machine learning and deep learning capabilities. Visual examination of CT scans was used in the detection process to look for unusual patterns of density or hyperdense areas that could indicate brain hemorrhage. This manual method took time, was more prone to misdiagnosis, and needed a high level of knowledge [25]. The radiologists diagnosed ICH using various visual indicators, including blood in the Cerebrospinal Fluid (CSF). They also looked at clinical data to analyze the patient's medical history and symptoms. However, the complexity of brain architecture, the diversity in imaging quality, and the possibility of human mistakes limited this technique [26].

Some machine learning algorithms have been explored for ICH detection. For instance, a random forest-based approach demonstrated an accuracy of 91.4% sensitivity and 95.6% specificity [27]. While effective, random forests also face challenges similar to CNNs, such as the need for large datasets and considerable computational resources, making them less practical for environments with limited access to such resources. Another notable contribution to ICH detection is Aidoc's algorithm, which boasts an impressive sensitivity of 92.3% and specificity of 97.7% [28]. This algorithm performs well in general, but its accuracy drops when detecting small hemorrhages or those located in complex brain areas, making it less reliable in some instances.

Typically used in Natural Language Processing (NLP), Transformers have also been applied to ICH detection. A model based on Swin Transformers achieved a log loss of 4.3%, outperforming traditional CNN-based models [29]. However, transformer models are resource-intensive and require long training, limiting their accessibility in less-equipped medical facilities.

With the rapid advancements in artificial intelligence, medical imaging has seen significant developments. Researchers have focused on leveraging these technologies to automate and improve the accuracy of ICH detection. One of the earliest studies by Grewal et al. [30] proposed RADnet, a convolutional neural network (CNN) model that achieved 93.5% accuracy in detecting ICH. However, RADnet required a largely annotated dataset of 12,000 CT scans to perform effectively, posing a challenge regarding data availability and expert involvement in the annotation

process. Furthermore, the large dataset requirement limits its scalability for real-world applications, especially in settings where annotated data is scarce. Similarly, the application of CNNs in ICH detection has shown remarkable results, with some models achieving sensitivity and specificity rates as high as 95.5% and 98.5%, respectively [31]. However, these deep learning models are computationally demanding and require extensive training data to generalize well. A different CNN-based approach categorized CT scans into hemorrhagic and non-hemorrhagic groups, with a model trained on 10,000 CT scans achieving 95.6% accuracy [32]. Despite its success, the small dataset size raised concerns about its generalizability to unseen data from different sources. Another study focused on transfer learning to enhance ICH detection. This study employed a pre-trained CNN model fine-tuned using 2,000 CT scans, achieving 92.5% accuracy [33]. While transfer learning reduced the need for extensive training data, the limited dataset still risks overfitting, potentially impacting the model's performance on new data.

Several studies have also explored deep learning algorithms trained on larger datasets. One such model was evaluated using the Receiver Operating Characteristic (ROC) curve and achieved an Area Under the Curve (AUC) score of 0.846 [34]. The model showed promising results but needed to improve with low-quality CT scans, highlighting the need for better preprocessing techniques. Another CNN-based approach specifically targeted acute cerebral hemorrhage detection, reaching an AUC of 0.991, suggesting that CNNs are highly effective in detecting acute ICH [35]. Nevertheless, this study acknowledged that the model might not perform well on less common types of hemorrhage, suggesting that it is still essential to consider model limitations.

Phong et al. [36] proposed a CNN-based method for brain hemorrhage detection, achieving a sensitivity of 94.1% and specificity of 96.5%. Although this approach demonstrated strong performance, the authors highlighted potential biases towards certain hemorrhage types and emphasized the need for larger labeled datasets to enhance the model's accuracy. In another study, an AI tool aimed at detecting unreported cerebral bleeding as a peer reviewer flagged 28 out of 5,585 CT scans labeled negative for ICH. Upon further review by neuroradiologists, 16 of these scans were confirmed as missed hemorrhages, demonstrating a false-negative rate of 1.6%. This finding underscores the potential of AI tools to reduce human errors in ICH detection, especially in challenging regions such as parafalcine and cerebral convexity [37].

Lee et al. [38] introduced a deep learning model that reached a sensitivity of 95.5% for detecting ICH despite challenges related to image quality variations. Similarly, Chilamkurthy et al. [39] presented a model that achieved 92.5% accuracy but relied heavily on expert annotation, indicating the model's dependence on domain-specific knowledge. This reliance on annotated data remains a standard limitation in many ICH detection studies.

Another promising approach used the Grad-CAM technique to provide visual explanations for ICH classification decisions, reaching an accuracy of 92.7% and an Area Under the ROC (AUROC) of 0.978 [40]. This technique allows clinicians to interpret the model's reasoning more transparently. However, heavy computational loads and the need for high-quality labeled datasets were notable drawbacks. A separate study employing EfficientNet technology for decision support achieved 92.7% accuracy [41]. The model struggled with detecting small or blurred hemorrhages, suggesting that additional model fine-tuning is required for more challenging cases.

To address some of the typical limitations of CNNs, a study proposed a hybrid attention-based ResNet architecture that used a Deep Convolutional Generative Adversarial Network (DCGAN) to tackle class imbalance and Principal Component Analysis (PCA) for feature reduction. This approach achieved accuracies of 99.2% for epidural hemorrhage and 96.1% for subarachnoid hemorrhage [42]. The model's performance exceeded traditional CNNs, but its reliance on one dataset and a single augmentation technique called DCGAN were cited as limitations. Another study employed Artificial Neural Networks (ANNs) along with Gray Level Co-occurrence Matrix (GLCM) and Local Binary Patterns (LBP) feature extraction techniques, achieving an accuracy of 99.3%, sensitivity of 99.5%, and an AUC of 99.84% [43]. While the results were impressive, the study acknowledged the risk of overfitting on smaller datasets. It emphasized the need for further validation on more diverse datasets to ensure the model's generalizability.

Further advancements in ICH detection came from the integration of more complex architectures. Kim et al. [44] proposed a 3D joint convolutional and recurrent neural network, achieving 94.5% accuracy. This model accounted for temporal dependencies in the CT scan slices but faced data variability and size issues. Similarly, Park et al. [45] developed a model that achieved 96.5% accuracy but remained dependent on expert annotations, which continues to be a bottleneck in

scaling deep learning models for medical applications.

The primary objective of explainable AI (XAI) is to produce AI models that are easily interpretable and understandable by humans [46]. XAI techniques have also been incorporated into ICH detection frameworks. A ResNet model combined with Grad-CAM achieved a precision of 0.81, sensitivity of 0.67, and specificity of 0.86 while providing visual explanations for clinicians by highlighting the exact regions where hemorrhage was detected [47]. The pixel-level precision of the visualizations proved valuable in confirming the presence of hemorrhage and assisting in prompt diagnosis and treatment. This method uses a convolutional neural network (CNN) to process individual CT slices and an LSTM network to integrate sequence data across slices. This model successfully balances accuracy and speed, achieving high results in the RSNA competition by lowering processing imposed via feature selection. The approach achieved a weighted mean log loss of 0.04989 in the final test set [48]. However, drawbacks include the computational cost of maintaining both CNN and LSTM layers and the loss of input data resolution, which may impair detection precision on higher-quality scans. The cHS-Heuron model evaluated data from the RSNA ICH challenge and focused on distinguishing small-volume ICH cases from larger ones. By classifying slices based on hematoma volume, this model achieved a sensitivity of 0.88 and a specificity of 0.99 for slice-wise SICH classification [49]. However, classification performance dropped slightly for more minor hematomas, where slice-wise specificity was more challenging due to the low volume of data on more minor hemorrhages, which limited the overall generalizability.

Despite the promising advancements across various studies, key challenges remain in ICH detection using deep learning models. The main issues include the need for large, annotated datasets, variations in CT scan quality, and the reliance on expert input during training. There is also a gap in understanding the comparative performance of different model architectures. This study aims to address this gap by comparing the performance of three deep learning models—InceptionV3, DenseNet201, and EfficientNetB4—in detecting ICH. This comparative analysis will provide valuable insights into each model's strengths and weaknesses and explore their potential application in clinical settings, where fast and accurate ICH detection is crucial for improving patient outcomes.

Chapter 3: Dataset

3.1. Defining the dataset

The RSNA Intracranial Hemorrhage dataset was introduced as part of a Kaggle competition hosted by the Radiological Society of North America (RSNA) in collaboration with the American Society of Neuroradiology (ASNR). The dataset was designed to promote research in automatic hemorrhage detection and classification from CT scans to assist radiologists in diagnosing intracerebral hemorrhages more accurately and efficiently. The RSNA aimed to support advancements in deep learning and computer vision techniques in medical imaging by providing a large and well-annotated dataset.

The dataset was initially released in 2019 and has been one of the most widely used datasets for research in this field since then. Though no significant revisions have been made, its extensive use across multiple studies has highlighted its robustness and usefulness for detecting and classifying various types of ICH.

The RSNA dataset includes six distinct types of intracerebral hemorrhages:

- **Epidural Hematoma**
- **Intraparenchymal Hemorrhage**
- **Intravascular Hemorrhage**
- **Subdural Hematoma**
- **Subarachnoid Hemorrhage**
- **Multiple Hemorrhages**

3.2. Types of Intracerebral Hemorrhages

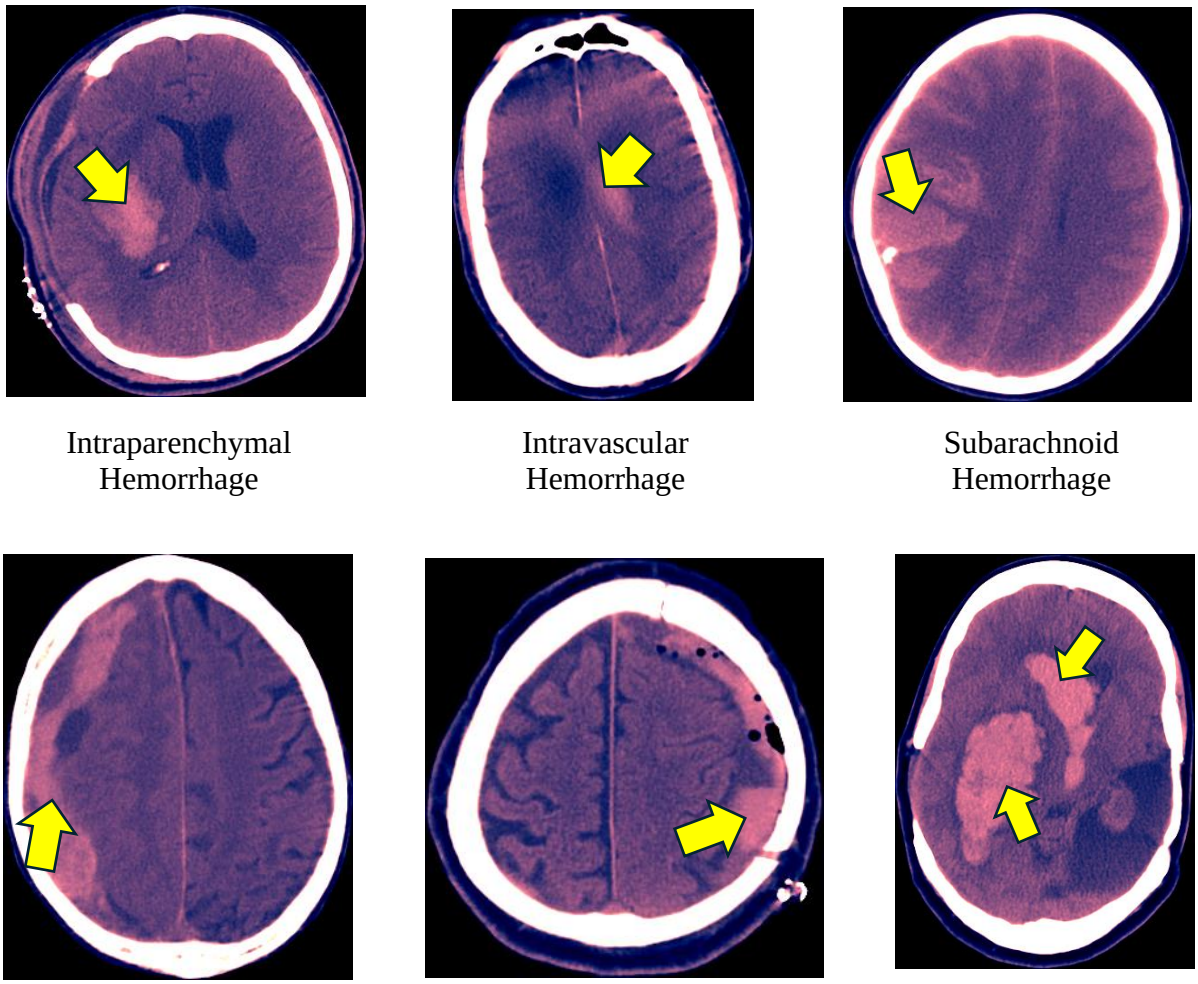


Figure 3.1: Types of Hemorrhages in the RSNA dataset

Table 3.1: Sample list of IDs and their corresponding labels in the dataset

ID	TYPE OF HEMORRHAGE	LABEL
ID_12CADC6AF	epidural	0
ID_12CADC6AF	intraparenchymal	0
ID_12CADC6AF	intraventricular	0
ID_12CADC6AF	subarachnoid	0
ID_12CADC6AF	subdural	0
ID_12CADC6AF	any	0
ID_4598FA77A	epidural	0
ID_4598FA77A	intraparenchymal	1
ID_4598FA77A	intraventricular	1
ID_4598FA77A	subarachnoid	0
ID_4598FA77A	subdural	1
ID_4598FA77A	any	1
ID_6DA8791EE	epidural	0
ID_6DA8791EE	intraparenchymal	1
ID_6DA8791EE	intraventricular	0
ID_6DA8791EE	subarachnoid	0
ID_6DA8791EE	subdural	0
ID_6DA8791EE	any	1
ID_C34F6AEC5	epidural	0
ID_C34F6AEC5	intraparenchymal	0
ID_C34F6AEC5	intraventricular	0
ID_C34F6AEC5	subarachnoid	0
ID_C34F6AEC5	subdural	0
ID_C34F6AEC5	any	0

Each CT slice is labeled with an "any" label, where '1' denotes the presence of any hemorrhage, and '0' denotes its absence. The RSNA dataset consists of 752,602 Digital Imaging and Communications in Medicine (DICOM) files, each representing an axial slice from a patient's head CT scan. We extracted the 'PatientID' and 'Z-axis' coordinates from the metadata of the DICOM files to sort the CT scans sequentially to create a complete 2D depiction of a patient's CT scan. This also helps visualize the progression of brain abnormalities across multiple slices, enhancing diagnostic accuracy.

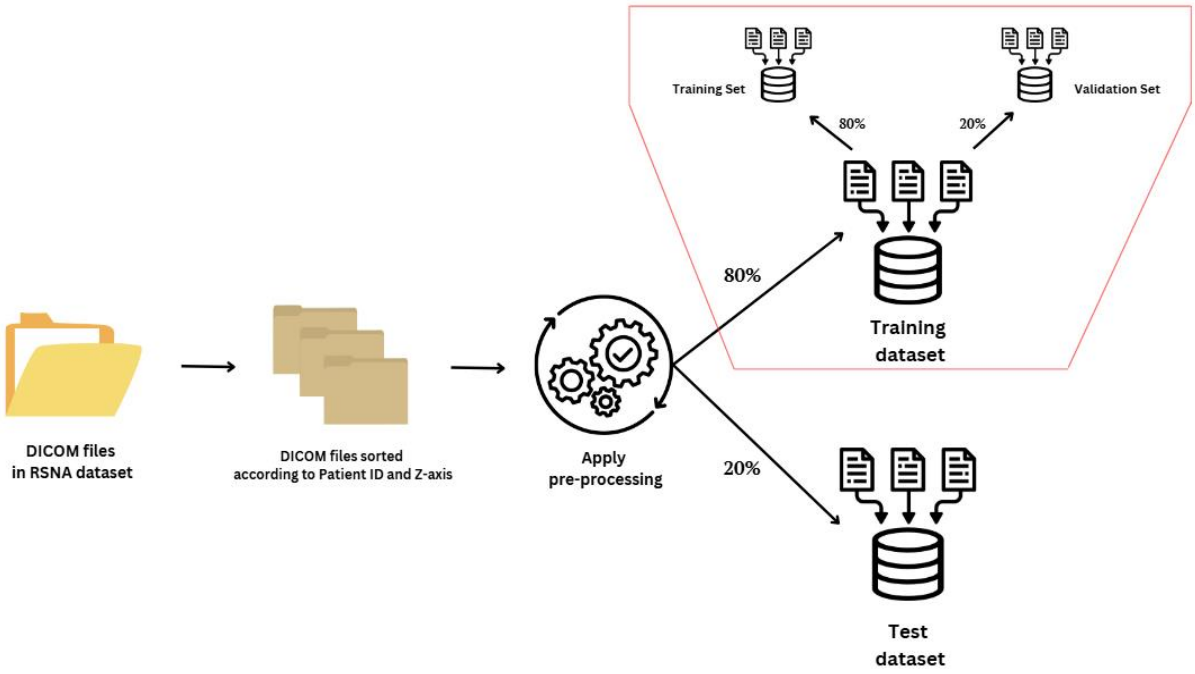


Figure 3.2: Illustration of how the dataset was split

The dataset contains 18,938 full CT scans. To split the dataset into training, validation, and test sets, 80% of the scans were allocated for training and validation, while the remaining 20% were reserved for testing. This resulted in 15,154 full CT scans for training and 3,784 for testing, ensuring a representative and unbiased evaluation split.

3.3 Handling Class Imbalance

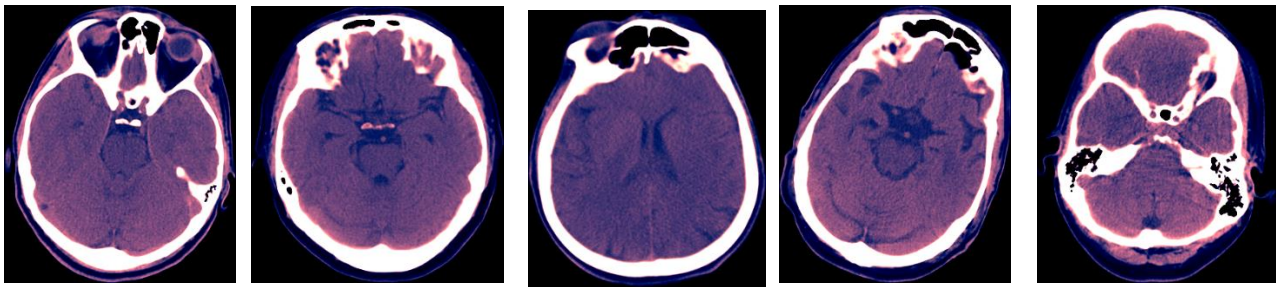
A significant challenge with the RSNA dataset is the inherent class imbalance, particularly affecting the Epidural Hematoma class, which initially contained only 1,350 CT slices. To overcome this imbalance, a duplication strategy was used for oversampling. Each Epidural Hematoma slice was duplicated exactly once, raising the total to 2,700 slices.

To create a balanced dataset for the ‘Single’ class, which contains slices depicting only one type of hemorrhage, 2,700 slices were randomly selected from each of the other hemorrhage types as well. This resulted in 13,500 CT slices in the ‘Single’ class, ensuring equal representation of all hemorrhage types.

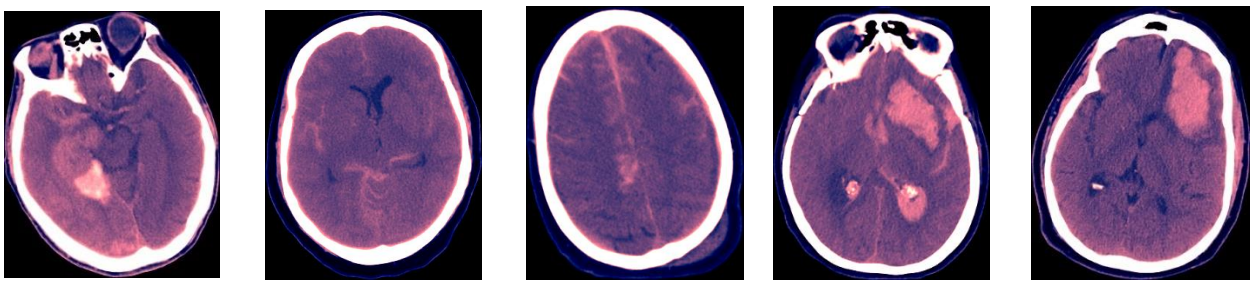
In addition to the 'Single' class, we balanced the 'Normal' and 'Multiple' classes. 13,500 slices were selected from the 'Normal' class (which contains no hemorrhage), and 13,500 slices were chosen from the 'Multiple' class (which includes slices depicting more than one type of hemorrhage). This balancing procedure was vital to prevent the model from becoming biased towards any individual class, thereby improving generalization during testing.

The DICOM files were preprocessed and converted into the more commonly used Portable Network Graphics (PNG) format to make the data suitable for deep learning models. A total of 40,500 CT slices (13,500 each from the Single, Normal, and Multiple classes) were used for training and validation, while 128,320 slices were reserved for testing. This significant testing data ensured that the final evaluation results would be robust and generalizable across clinical cases.

(a) Normal Class



(b) Multiple Class



(c) Single Class

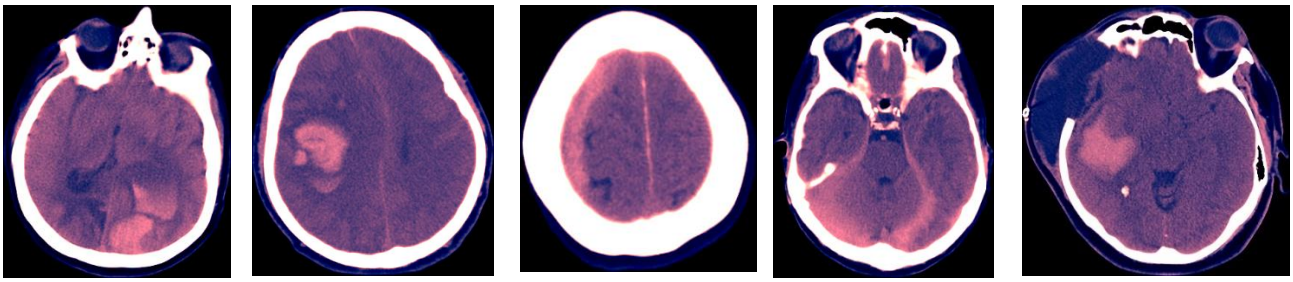


Figure 3.3: Samples of (a) Normal, (b) Multiple, (c) Single training class image

3.4 Comparison with Other Medical Imaging Datasets

Compared to other publicly available medical imaging datasets, such as the CQ500[50] or the Medical Image Computing and Computer Assisted Intervention (MICCAI)[51] datasets, the RSNA dataset offers several advantages:

- **Size**: With over 750,000 CT slices, the RSNA dataset is significantly larger than most other hemorrhage detection datasets, providing vast data for model training and evaluation.
- **Diversity**: The dataset includes seven types of hemorrhage and non-hemorrhage cases, making it highly versatile for training models to detect various conditions.
- **Labeling accuracy**: Expert radiologists annotated the dataset, ensuring high-quality, accurate labels for training deep learning models.

However, the RSNA dataset also has its limitations. For instance, the initial class imbalance required duplications to ensure fair training. Additionally, while the dataset focuses on detecting hemorrhages, it does not provide extensive metadata on other potential brain pathologies.

Despite these challenges, the RSNA dataset remains one of the most comprehensive and widely used medical imaging datasets for hemorrhage detection, which is why we used this dataset for our study.

Chapter 4: Methods

4.1 Pre-Processing and Data Augmentation

In this study, we applied multiple preprocessing and augmentation techniques to ensure the dataset's diversity, robustness, and compatibility with deep learning frameworks. Medical imaging datasets, mainly CT scans, can be subject to factors such as patient positioning and scanner variability. Therefore, preprocessing and augmentations play a critical role in improving the generalization and accuracy of deep learning models.

4.1.1 Data Augmentation

Several data augmentation techniques were applied to increase the diversity of the training set artificially. Each technique simulates real-world conditions and improves the model's generalization ability by exposing it to different variations of the same image.

The first augmentation applied was rotation. A rotation range of 10 degrees was used, meaning that each CT slice could be rotated up to 10 degrees in either direction. This mimics slight rotations during a CT scan, often due to the patient's movement or minor inconsistencies in scanner alignment. Introducing this variability helps the model recognize the same anatomical structures even when not perfectly aligned, improving its robustness to rotational differences.

Next, shearing was applied with a range of 0.05. Shear transformations shift parts of a CT slice in one direction while leaving others in place, creating a distortion that resembles a tilted view. This augmentation is significant in medical imaging, where subtle differences in perspective can occur due to changes in patient positioning or scanning angles. Shearing helps the model learn to identify key features even when they are slightly skewed, thus improving generalization.

The zoom augmentation, also set with a range of 0.05, enlarges or reduces the CT slices by a small factor, simulating variations in image magnification that might happen due to changes in scanner settings or distance. This augmentation ensures that the model can identify essential features regardless of minor variations in size, making it more resilient to differences in zoom levels across

datasets.

Another transformation applied was horizontal flipping. Since brain structures are generally symmetrical, flipping the CT slices horizontally helps the model learn features from both orientations, preventing it from becoming biased towards one specific side of the CT slice. This is particularly helpful in medical imaging, where conditions can manifest on either side of the brain.

Lastly, the fill mode was set to 'nearest,' meaning that any newly exposed pixels created by the transformations (such as rotations or shifts) were filled in using the nearest neighboring pixel values. This ensures that no critical information is lost during the transformations and helps maintain the integrity of the CT slices.

These augmentations mimic real-world variations in CT imaging, such as slight movements or adjustments in scanning protocols, making the model more resilient to different clinical conditions. Introducing such transformations during training reduces the risk of overfitting to the original dataset, ultimately enabling the model to generalize better and perform more accurately on unseen data.

4.1.2 Windowing Techniques

CT scans are measured in Hounsfield Units (HU), with wide dynamic ranges that capture different tissues and structures. Windowing allows us to focus on critical areas without oversaturating or losing contrast in the CT slices. Each pixel in a CT slice represents a tissue density value, and radiologists rely on specific HU ranges, referred to as window width (WW) and window center (WC), to visualize different anatomical structures. Three windowing techniques were used on our dataset to enhance different anatomical regions:

- Brain Window: WC = 40, WW = 80 (used to highlight brain tissue)
- Subdural Window: WC = 80, WW = 200 (used to detect subdural structures)
- Soft Tissue Window: WC = 40, WW = 380 (used to highlight soft tissues)

The windowing process was applied to the DICOM images using the following formula:

$$window_min = \frac{2 \times WC - WW}{2}$$

$$window_max = \frac{2 \times WC + WW}{2}$$

CT scans are typically stored in the DICOM format, which can be challenging to manipulate with standard deep-learning libraries. Each windowed CT slice was scaled to an 8-bit range [0, 255] and combined into a three-channel RGB image in Portable Network Graphics (PNG) format for easy processing. The brain, subdural, and soft tissue windows were mapped to the red, green, and blue channels. This approach provided the model with rich, multi-layered structural information from a single CT slice.

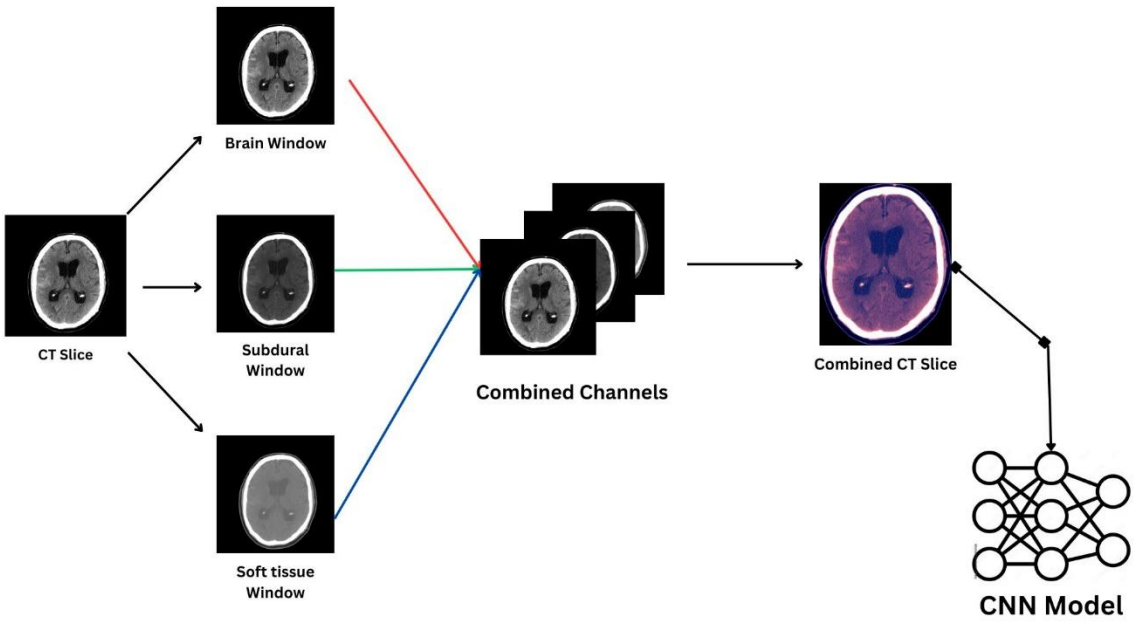


Figure 4.1: Combination of window channels to form RGB CT slice

4.1.3 Resizing and Batch Processing

Since the original CT slices were captured at 512×512 pixels, they were resized to 224×224 pixels using linear interpolation. This balanced computational efficiency while preserving the details of the CT slices, allowing for faster processing during training without significant information loss. The resized CT slices were then processed in batches of 16 slices to optimize memory usage and avoid GPU bottlenecks during training.

4.1.4 Segmentation and Skull Extraction

A segmentation algorithm was applied to extract the skull region and exclude unnecessary details from the CT slices for further refinement. The most significant contour (the skull region in this case) was identified using binary thresholding and contour retrieval. Slices, where this contour occupied less than 15% of the total CT slice area, were excluded, removing non-informative slices that could confuse the model (e.g., those containing noise or metal artifacts).

After segmentation, the skull region was isolated by drawing a bounding box around it, and all areas outside this boundary were cropped out. This provided clean, skull-focused slices with minimal background, improving the dataset's clarity and enabling the model to focus on the region's most relevant for hemorrhage detection.

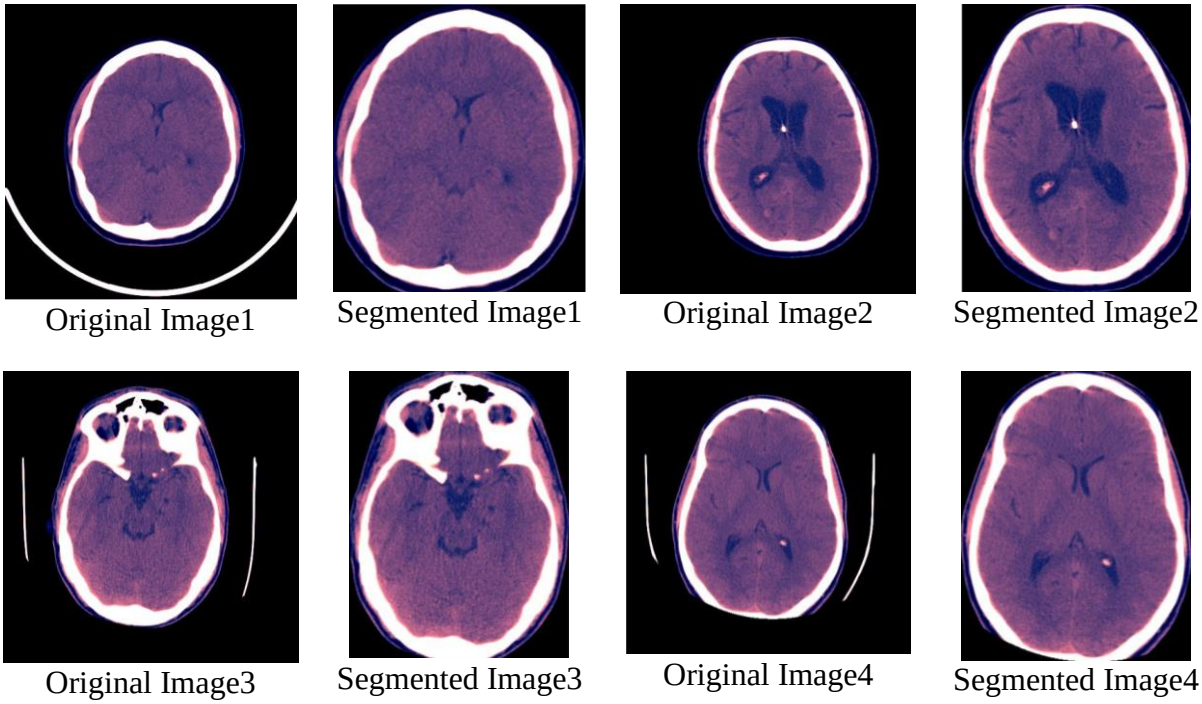


Figure 4.2: Segmentation of skull region

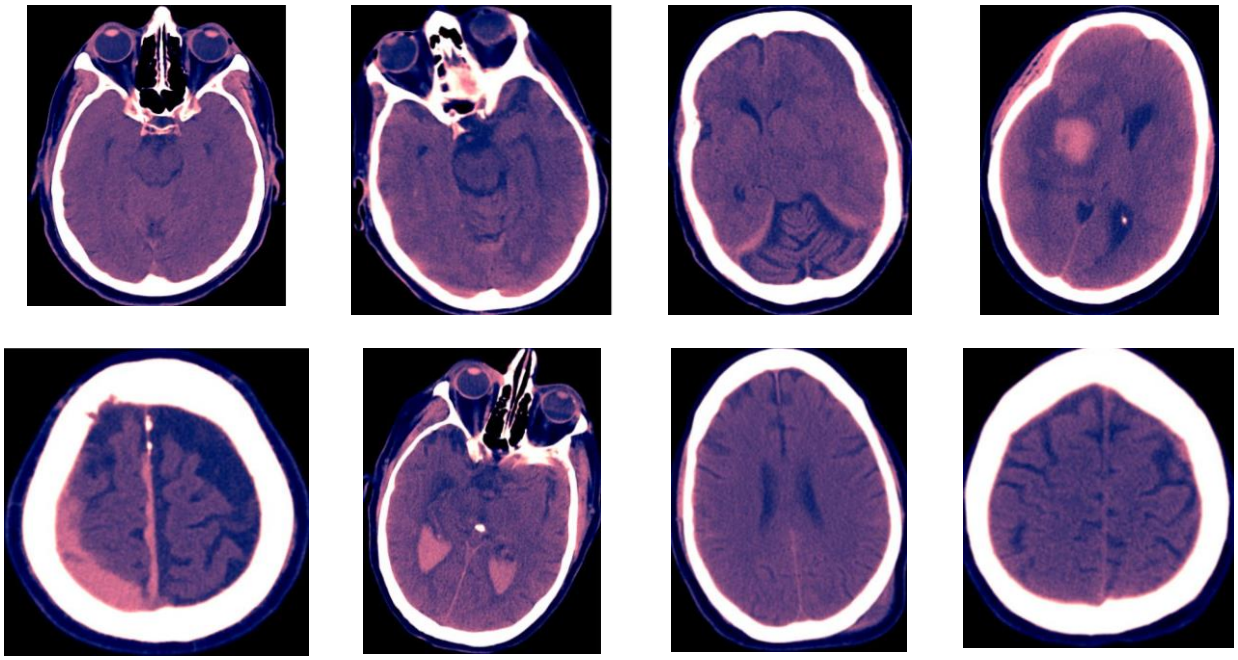


Figure 4.3: Samples of segmented CT slices used for training

4.2 Model

4.2.1 Model Layers

This study leveraged transfer learning using pre-trained models trained on the 'ImageNet' dataset. Transfer learning is a machine learning technique where a model trained on one task is reused for another related task [52]. Initially trained on large datasets like ImageNet, pre-trained models have learned general features such as edges, textures, and patterns. Fine-tuning these models on our dataset allows the network to adapt these features to medical contexts without requiring extensive training data. This technique significantly reduces training time while improving the model's ability to detect hemorrhages accurately.

The models used in this study for ICH detection were constructed with a backbone based on pre-trained Convolutional Neural Networks (CNNs), namely InceptionV3, DenseNet201, and EfficientNetB4, each proven to be effective in various image classification tasks due to their ability to extract complex features from images. To leverage the power of these models, the following layers were added on top of the CNN backbones:

- **Global Average Pooling Layer**: This layer converts the spatial dimensions of the feature maps into a single value per feature map by averaging the activations. This results in a vector that can be fed into fully connected layers, reducing the overall model complexity and overfitting risks.

- **Dense Layers**: Two fully connected layers with 512 and 256 neurons, respectively, were added, both followed by batch normalization and ReLU activation. The first dense layer with 512 neurons captures high-level abstract features, while the second with 256 neurons refines the learned features.

This consistent architecture was applied to all models to avoid bias, allowing for a fair comparison between the CNN architectures. Additionally, batch normalization was used along with ReLU activation layers to prevent overfitting and ensure that models generalized well, which normalizes the output of the dense layers and helps stabilize the learning process, improving convergence.

4.2.2 Training Procedure

All models were initially trained for 50 epochs with a learning rate of 0.001 using the Adam optimizer [53], which has shown effectiveness in adaptive learning rate optimization for deep learning models. The loss function used was categorical cross-entropy [54], commonly used for multi-class classification tasks, as the model needs to classify CT slices into different categories (Single, Multiple, or Normal).

The following callbacks were used to enhance the training process:

- **ModelCheckpoint:** This callback was used to save the best model with the lowest validation loss during training. Keeping the best-performing model ensured we did not encounter overfitting issues during further training epochs.

- **ReduceLROnPlateau:** If the validation accuracy plateaued for five consecutive epochs, this callback reduced the learning rate by 10%, allowing the model to fine-tune the weights with more minor adjustments [55]. The minimum learning rate boundary was set to 0.000001 to prevent the learning rate from stagnating too low, causing the model not to update its weights during training.

After the initial training phase, the best model was fine-tuned for 80 additional epochs but with a reduced learning rate of 0.000001, allowing it to optimize its parameters further. The same loss function and optimizer were used during this phase, but the minimum learning rate boundary was set to 0.0000000001 to ensure small but precise weight updates in the fine-tuning stage.

4.3 Evaluation Metrics

Once the top-performing models were chosen, their performance was evaluated using a set of key metrics. These metrics provided insight into how well the models could classify the CT slices and comprehensively compare their effectiveness. The models were evaluated on the test dataset, which was utterly unseen during training, to assess the model's generalization ability.

The evaluation metrics included are:

- **Accuracy**: Measures the proportion of correctly classified CT slices across all classes.
- **Precision**: Focuses on the accuracy of optimistic predictions, i.e., the proportion of correctly predicted positive hemorrhage cases out of all predicted positive cases.
- **Recall (Sensitivity)**: This measure measures how well the model identifies positive hemorrhage cases among all positive cases, indicating the model's ability to detect hemorrhages.
- **F1-Score**: The harmonic mean of precision and recall, providing a single metric that balances the trade-off between accuracy and recall. This score is instrumental when there is an imbalance between positive and negative cases, providing a more balanced evaluation.

To calculate accuracy (1), precision (2), recall (3), and F1-score (4) [56], we first determine the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). TP and TN represent correctly predicted CT slices, while FP and FN represent incorrectly predicted ones. These metrics are calculated using the following equations:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1\ Score = 2 \times \frac{Precision\ \times Recall}{Precision + Recall} \quad (4)$$

The categorical cross-entropy loss function, *L*, defined by equation (5), was used to assess the classification error during training. This loss function measures the difference between the predicted probability distribution and the actual labels, helping guide the optimization process to minimize misclassification. In this equation, *N* represents the number of samples, *C* represents the number of classes, and y_{ij} is a binary indicator that takes the value of 0 or 1, indicating whether class label *j* is the proper classification for sample *i*. The variable p_{ij} The predicted probability that the *(i)*th sample belongs to the *(j)*th class, whereas the term 'log' refers to the natural logarithm.

$$L = -\sum_{i=1}^{N}\sum_{j=1}^{C} y_{ij}\, log\, (p_{ij}) \qquad (5)$$

4.3.1 ROC Curve and AUC

Additionally, the models were evaluated using the Receiver Operating Characteristic (ROC) curve and the Area Under the ROC Curve (AUROC). The ROC curve plots the True Positive Rate (TPR) against the False Positive Rate (FPR) at various threshold levels. At consecutive points along the ROC curve, TPR represents the true positive rate, and FPR represents the false positive rate. This provides a visual representation of the model's ability to distinguish between positive and negative classes.

The AUROC value (6) ranges from 0 to 1, where TPR_i and TPR_{i+1} Are the true positive rates at consecutive points while FPR_i and FPR_{i+1} Are the false positive rates at consecutive points. The AUROC is especially valuable in medical imaging tasks like hemorrhage detection, where a balance between sensitivity and specificity is crucial. A high AUROC indicates that the model distinguishes between hemorrhage and non-hemorrhage cases across various decision thresholds. The AUROC is computed using the trapezoidal rule as follows:

$$AUROC\ = \sum_{i=1}^{n-1}\left(\frac{TPR_{i+1} + TPR_i}{2}\right) \times (FPR_{i+1} - FPR_i) \qquad (6)$$

4.3.2 Grad-CAM Heatmaps

To further validate the models' predictions and ensure that they were focusing on the correct regions within the CT slices, Grad-CAM (Gradient-weighted Class Activation Mapping) heatmaps were generated. These heatmaps presented the areas in each CT slice that the model deemed most relevant for its predictions. By highlighting these regions, Grad-CAM helped determine whether the model was accurately focusing on the bleeding areas associated with hemorrhages or whether irrelevant features misled it. This method of model interpretability provided an additional layer of validation, ensuring that the predictions were accurate and clinically meaningful.

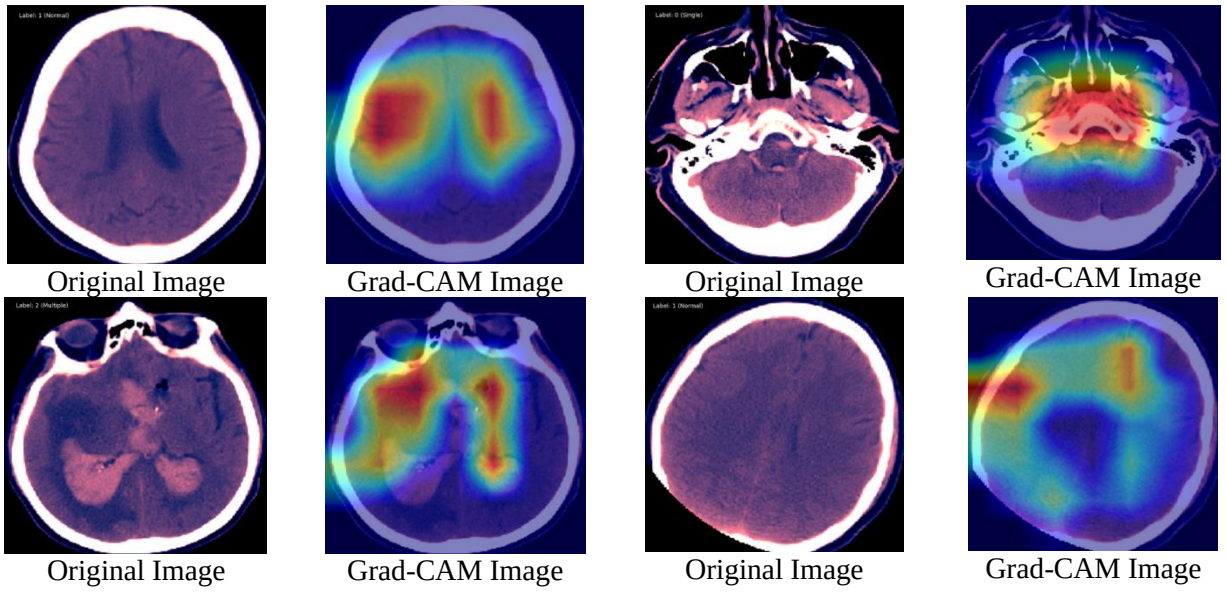


Figure 4.4: Samples of GRAD-Cam images

Utilizing these frameworks, we successfully assessed each model's performance and derived significant insights into its respective advantages and disadvantages.

Chapter 5: Results and Discussion

TABLE 5.1: Initial training Metrics

Model Names	**Initial Training**			
	Accuracy	***Precision***	***Recall***	***F1-Score***
InceptionV3	0.72	0.81	0.63	0.71
EfficientNetB4	0.78	0.83	0.72	0.77
DenseNet201	0.64	0.73	0.54	0.62

TABLE 5.2: Fine-Tuned Metrics

Model Names	**Fine Tuning Stage**			
	Accuracy	***Precision***	***Recall***	***F1-Score***
InceptionV3	0.84	0.86	0.81	0.84
EfficientNetB4	0.85	0.87	0.84	0.86
DenseNet201	0.84	0.86	0.83	0.84

Tables 5.1 and 5.2 show that EfficientNetB4 consistently outperformed InceptionV3 and DenseNet201 in the initial training and fine-tuning phases. The results highlight higher accuracy, precision, recall, and F1-score for EfficientNetB4, establishing it as the most robust model in detecting ICH from CT scans. This superior performance can be attributed to several advanced architectural features embedded within EfficientNetB4, including compound scaling, mobile inverted bottleneck (MBConv), and squeeze-and-excitation (SE) layers. These features enable EfficientNetB4 to process the input data more efficiently, leading to better outcomes than InceptionV3 and DenseNet201.

5.1 Architectural Significance of EfficientNetB4

EfficientNetB4 introduces several innovations that allow it to outperform the other models, particularly in handling high-resolution medical data:

1. **Compound Scaling**: EfficientNetB4's compound scaling is a key differentiator. Unlike other models that independently scale depth, width, or resolution, EfficientNetB4 uses a compound scaling method to balance these dimensions simultaneously. This coordinated scaling enables the model to learn complex patterns without a significant increase in computational cost:

    - **Depth scaling** enables deeper layers that capture more abstract features, which is ideal for detecting intricate patterns in CT scans.
    - **Width scaling** allows for detecting a wide range of features, which is critical when ICH presents subtle and diverse variations.
    - **Resolution scaling** improves the model's ability to capture fine-grained details, which is essential for identifying small hemorrhages in CT scans.

By simultaneously scaling these dimensions, EfficientNetB4 effectively extracts a more comprehensive feature than InceptionV3 and DenseNet201.

2. **MBConv Layers**: The mobile inverted bottleneck convolution (MBConv) layers in EfficientNetB4 further contribute to its computational efficiency. These layers optimize the flow of information by expanding feature channels before applying depth-wise convolutions and reducing them afterward. This approach minimizes the computational burden while ensuring that relevant information is retained, and unnecessary features are discarded. In comparison, DenseNet201 uses traditional bottleneck layers, which are less efficient in compressing and expanding feature channels, leading to higher computational costs.

3. **Squeeze-and-Excitation (SE) Layers**: EfficientNetB4's SE layers enhance the model's attention mechanism by recalibrating channel-wise feature responses. These layers assign greater weight to the most informative channels, which are very useful in medical imaging where subtle differences in pixel intensity may indicate critical features like hemorrhages. By improving the model's focus on relevant areas, SE layers help reduce overfitting and improve generalization, leading to higher performance on unseen data.

5.2 Comparison with DenseNet201

DenseNet201 uses dense connectivity, where each layer receives input from all previous layers. This setup promotes strong gradient flow and feature reuse, mitigating the vanishing gradient problem. While this improves feature extraction, DenseNet201 lacks the dynamic scaling and recalibration mechanisms in EfficientNetB4. DenseNet201's inability to scale depth, width, and resolution simultaneously and the use of traditional bottlenecks become a critical limitation in tasks like hemorrhage detection.

As shown in Table 1, DenseNet201's performance in the initial training phase was lower than the other models, achieving 64% accuracy with a precision of 73%, recall of 54%, and an F1-score of 62%. This is primarily due to the model's tendency to struggle with subtle features in medical images. However, DenseNet201 improved its performance after fine-tuning, achieving 84% accuracy, which is on par with InceptionV3. However, its architecture needed to be revised to handle high-resolution medical data as effectively as EfficientNetB4.

5.3 Comparison with InceptionV3

InceptionV3, built around inception modules, is known for its ability to extract features at multiple scales by applying convolutions with different filter sizes within the same module. This multi-scale feature extraction makes InceptionV3 very efficient in capturing information from varied image resolutions. However, while this architecture benefits multi-scale feature extraction, it does not offer the coordinated scaling across depth, width, and resolution that EfficientNetB4 introduces.

InceptionV3 performed well during fine-tuning, achieving 84% accuracy, 86% precision, 81% recall, and an F1-score of 84%. However, its lack of SE layers limits its ability to focus on critical areas in the CT slices. This limitation, especially compared to EfficientNetB4's recalibration ability, explains the marginally lower F1-score and recall, indicating InceptionV3's weaker ability to detect hemorrhage in complex cases.

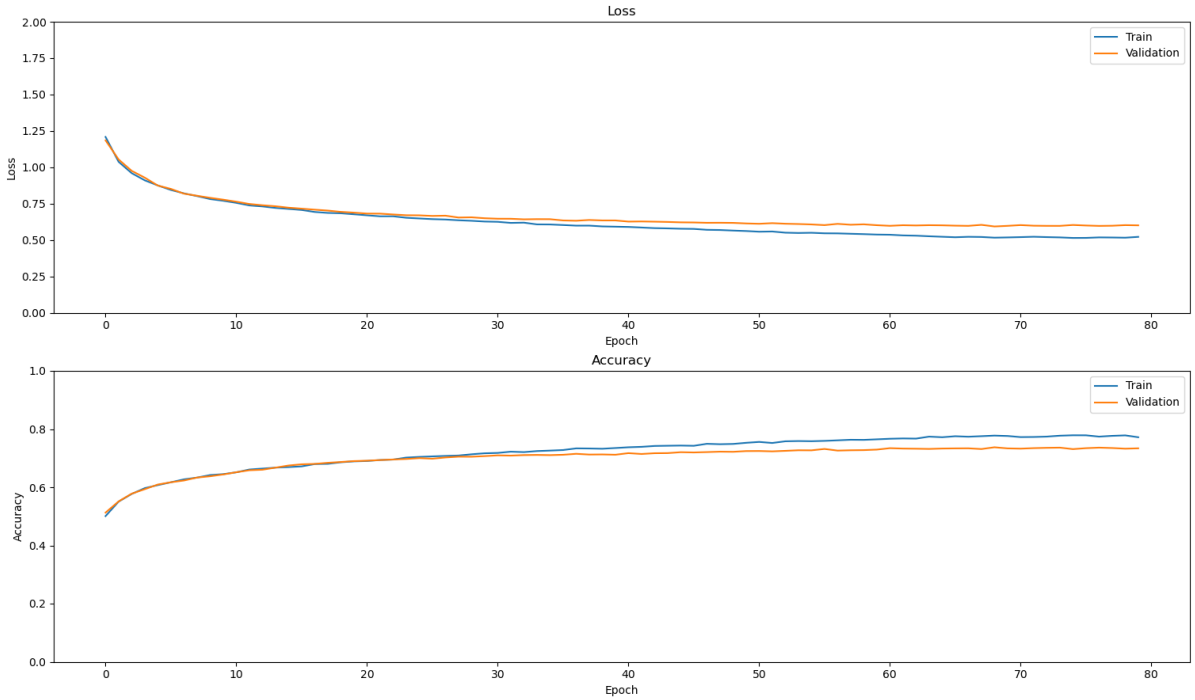


Figure 5.1: EfficientNetB4 plot history

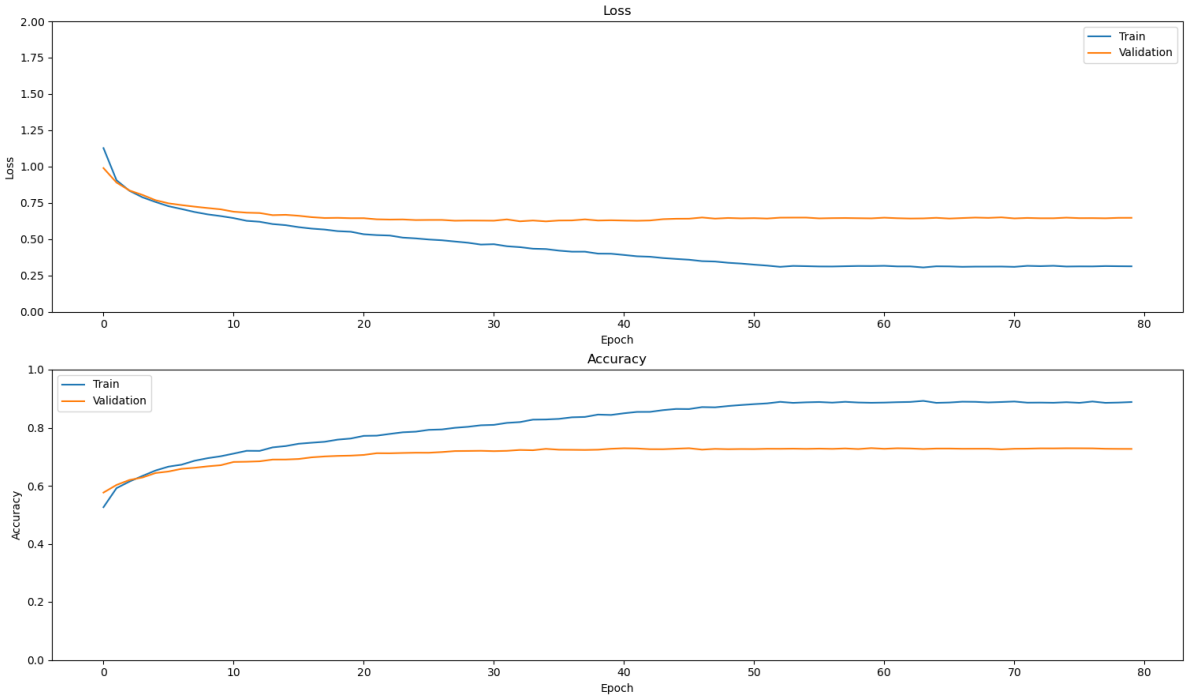


Figure 5.2: DenseNet201 plot history

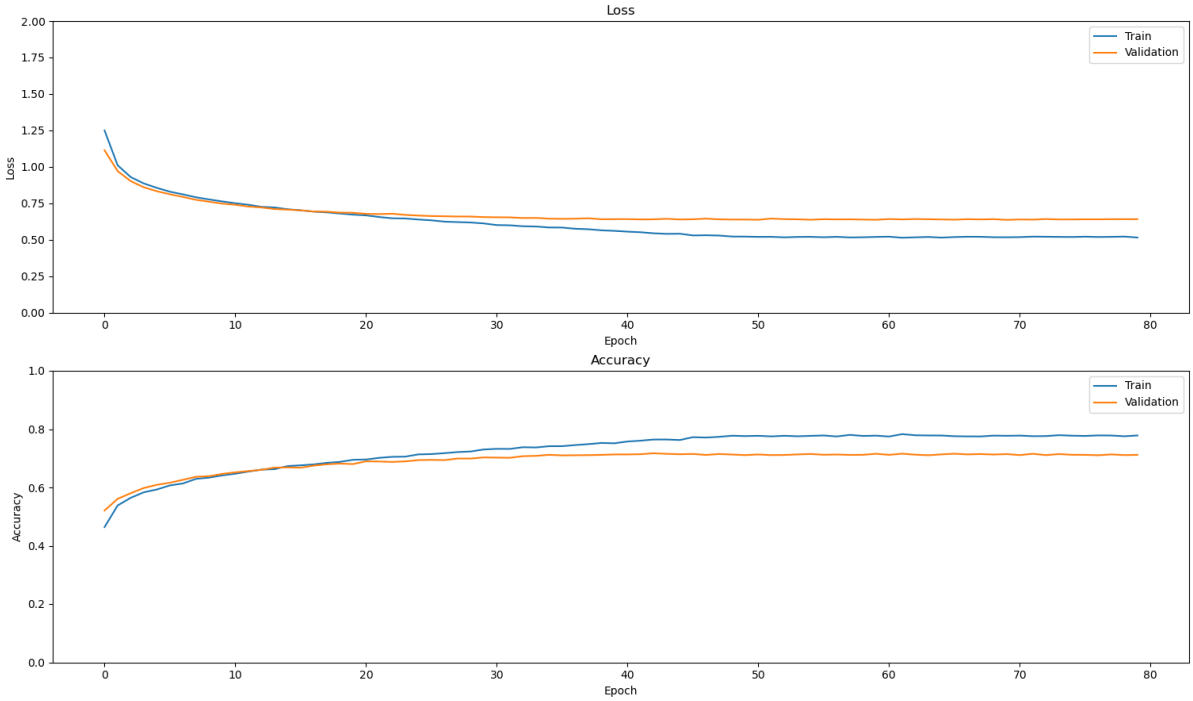


Figure 5.3: InceptionV3 plot history

5.4 Performance Analysis

The three models' architectural design differences manifest clearly in their performance metrics, as illustrated in Figures 5.1, 5.2, and 5.3. EfficientNetB4 demonstrated a **4%** difference between training and validation accuracy, suggesting strong generalization and minimal overfitting. In comparison, InceptionV3 exhibited a **6%** gap, and DenseNet201 displayed a more significant **10%** gap, indicating a higher tendency towards overfitting.

The performance metrics across all three models reflect the strengths and weaknesses of their respective architectures. EfficientNetB4 demonstrated superior performance across all evaluation metrics, achieving 78% accuracy in the initial training phase and 85% after fine-tuning. This was accompanied by comparatively higher precision, recall, and F1 scores, reinforcing its robustness in detecting hemorrhages.

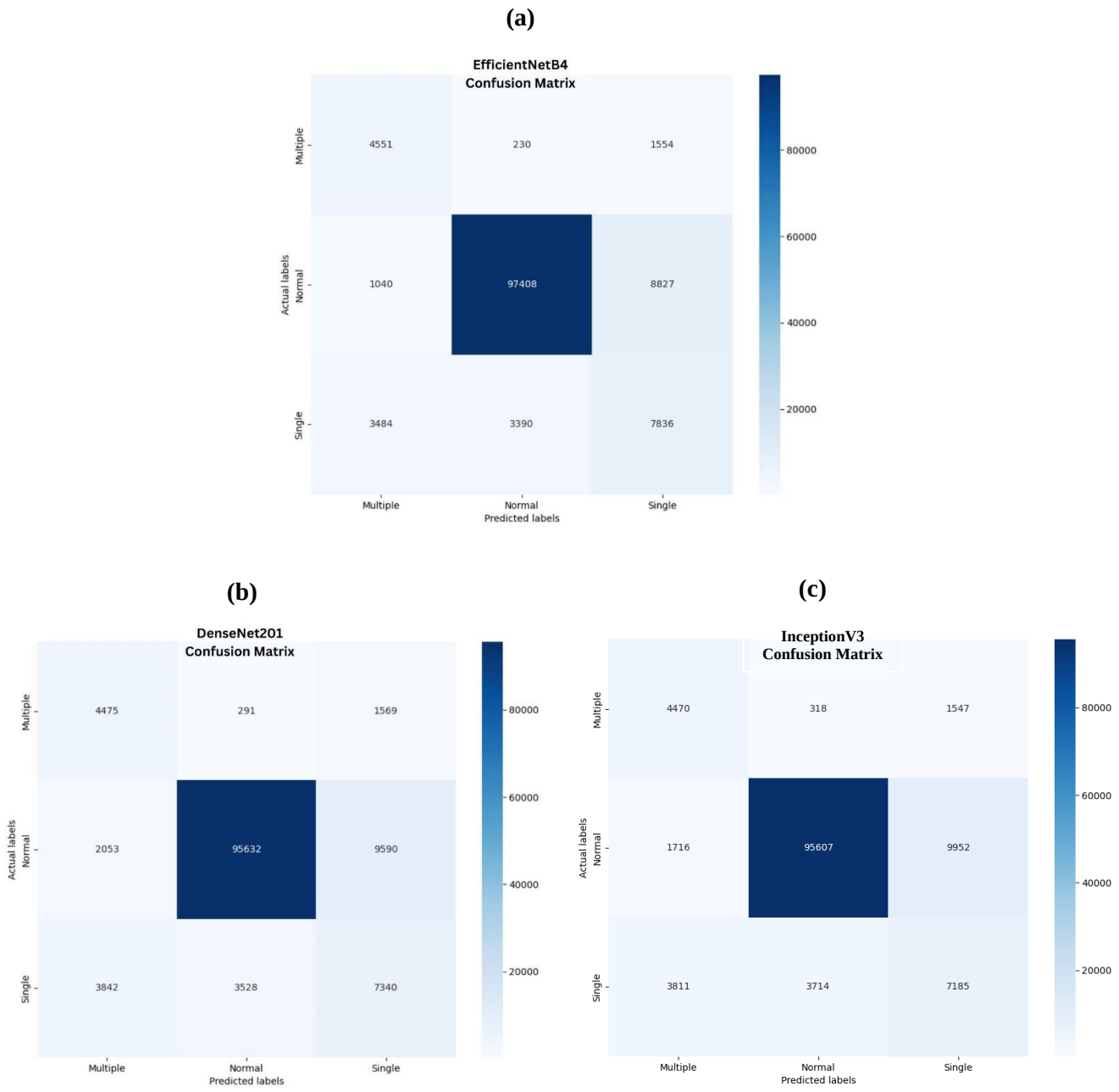


Figure 5.4: Confusion Matrices of (a) EfficientNetB4, (b)DenseNet201, (c) InceptionV3

The confusion matrices in Fig 5.4 further emphasize EfficientNetB4's accuracy in correctly classifying positive cases of hemorrhage with minimal false positives and false negatives. We can see that the total number of misclassified CT slices in the confusion matrix of EfficientNetB4 is significantly lower than in the other models. In comparison, DenseNet201 and InceptionV3 exhibited more false negatives, indicating a weaker ability to detect hemorrhage in some instances.

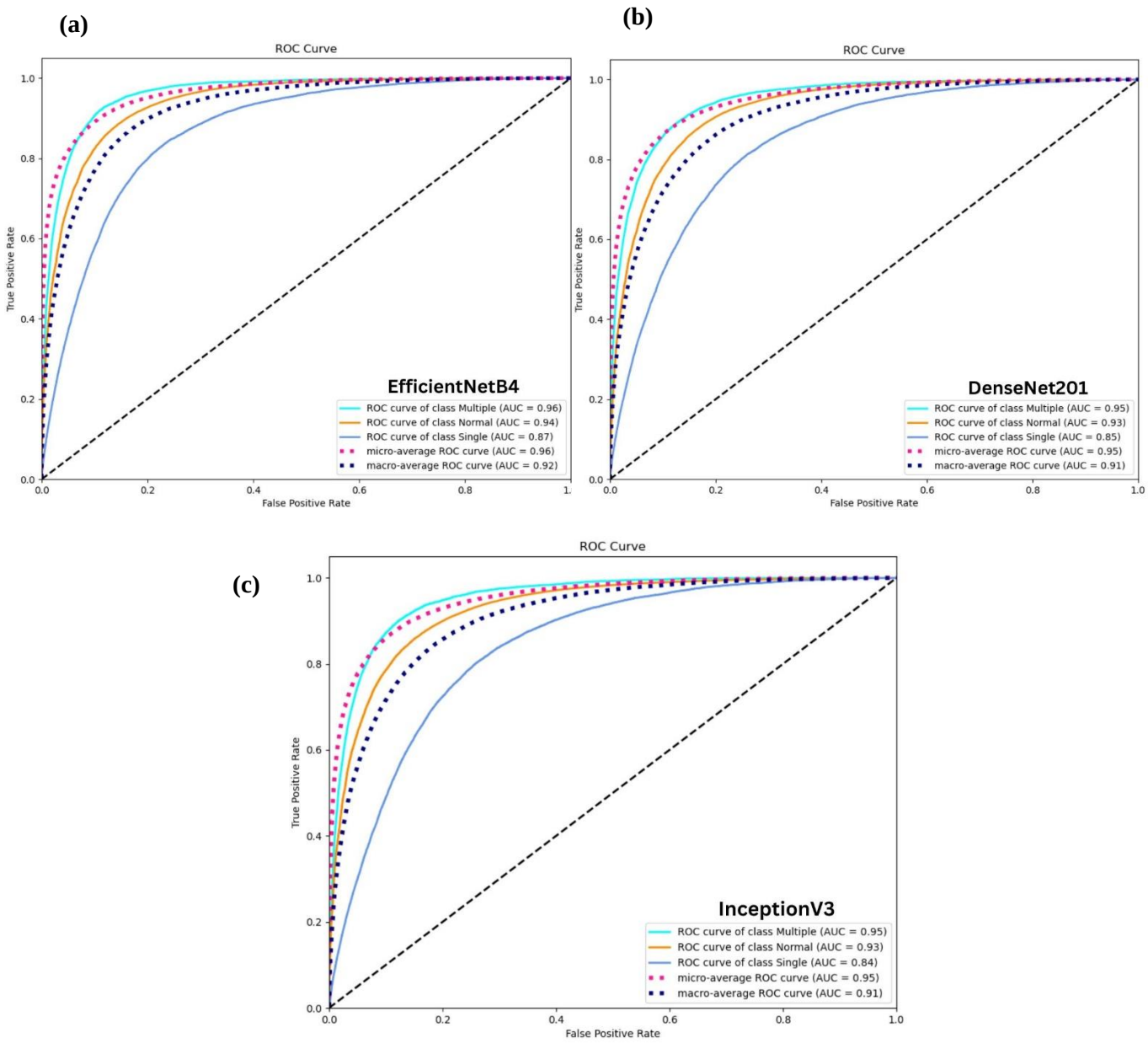


Figure 5.5: ROC curves of (a) EfficientNetB4, (b) DenseNet201, (c) InceptionV3

The ROC curves in Fig 5.5, including the micro and macro averages, offer additional insight into each model's classification ability at varying thresholds. EfficientNetB4's ROC curve demonstrated a steeper initial rise, with an AUROC value that surpassed both InceptionV3 and DenseNet201 on all accounts. This suggests that EfficientNetB4 excelled in general accuracy and distinguishing between positive and negative classes across various decision thresholds, which is crucial for medical tasks requiring high sensitivity.

Finally, the Grad-CAM heatmaps were used to see where the models were focused during decision-making. EfficientNetB4's heatmaps concentrate on the regions of the CT scans related to hemorrhage, even in cases where the visual signs were hard to see, indicating that the model focused on the right areas. On the other hand, DenseNet201 and InceptionV3 often highlighted larger, irrelevant areas across the whole CT scan, missing the key regions. This lack of focus led to higher false positive and false negative cases in these models. So, while DenseNet201 and InceptionV3 had high accuracy, their results were less reliable than EfficientNetB4, which consistently delivered more accurate and dependable outcomes.

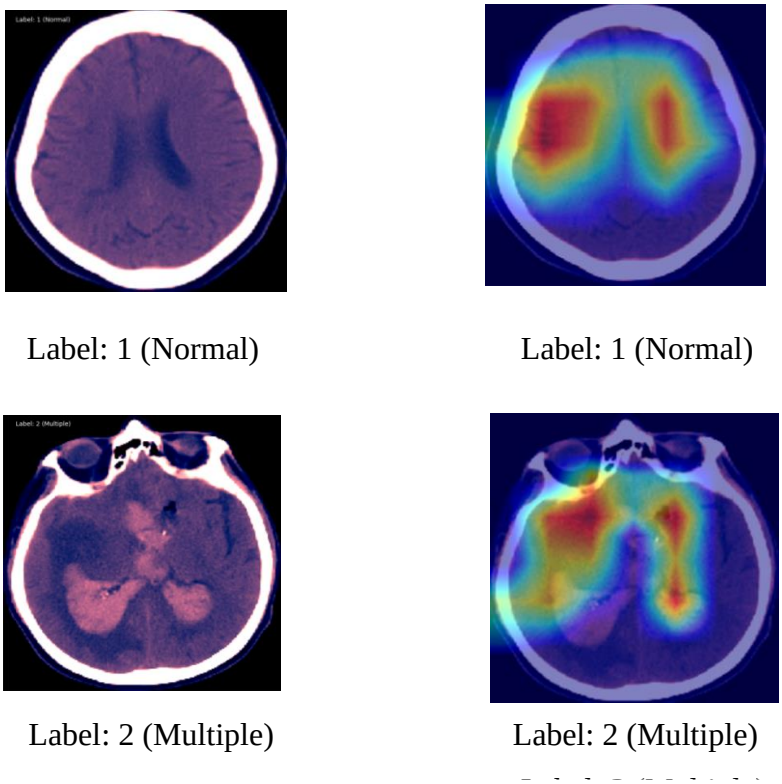
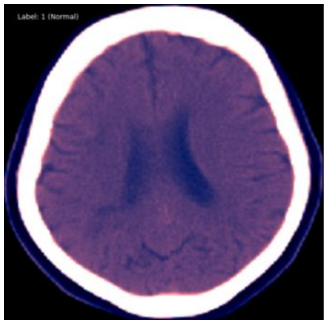
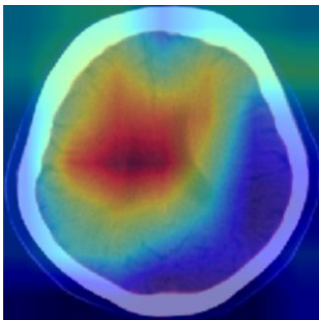


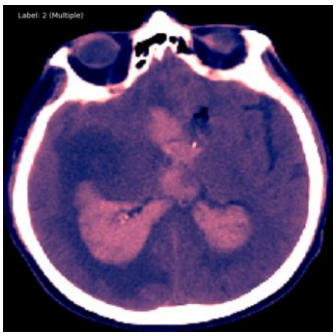
Figure 5.6: EfficientNetB4 Grad-CAM Heatmaps on Test Set



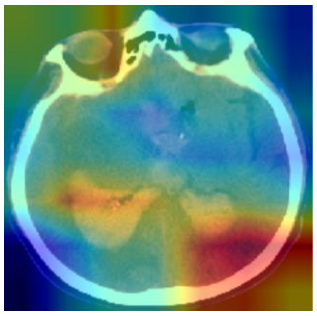
Label: 1 (Normal)



Label: 1 (Normal)

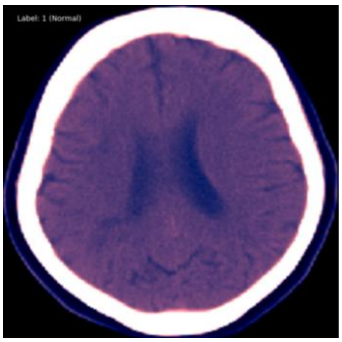


Label: 2 (Multiple)



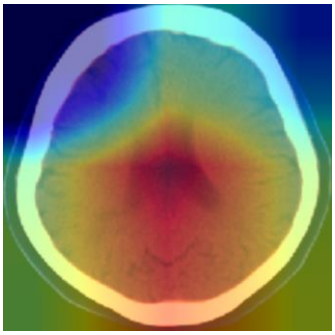
Label: 2 (Multiple)

Figure 5.7: DenseNet201 Grad-CAM Heatmaps on Test Set

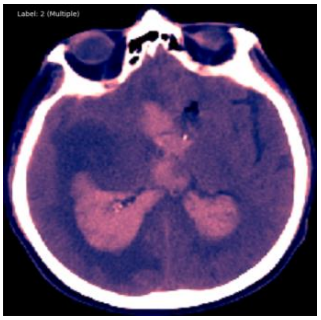


Label: 1 (Normal)

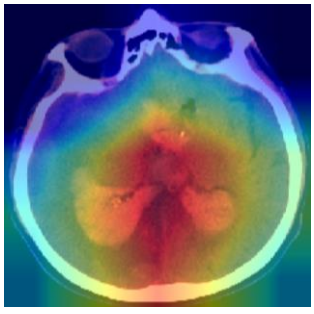
F



Label: 1 (Normal)



Label: 2 (Multiple)



Label: 2 (Multiple)

Figure 5.8: InceptionV3 Grad-CAM Heatmaps on Test Set

Chapter 6: Conclusion and Future Work

6.1 Summary

This thesis presents an in-depth exploration of deep learning models for detecting intracerebral hemorrhage from computed tomography scans. The primary objective was to evaluate and compare three advanced CNN architectures—InceptionV3, DenseNet201, and EfficientNetB4—in automatically detecting ICH. Utilizing the RSNA Intracranial Hemorrhage dataset, the models were trained, validated, and tested under real-world conditions.

EfficientNetB4 consistently outperformed the other models, achieving an accuracy of 85% and a precision of 87%. This performance is attributed to its advanced architectural innovations, such as squeeze-and-excitation (SE) layers, compound scaling, and mobile inverted bottleneck (MBConv) layers. The results highlight the importance of using advanced data augmentation techniques such as rotation, shearing, zooming, and horizontal flipping, along with preprocessing methods like HU windowing and skull segmentation, significantly improving generalization and robustness in model performance.

6.2 Challenges and Limitations

The experiments were conducted on a desktop computer with an Intel® Core™ i5-7500 Central Processing Unit (CPU) operating at a clock speed of 3.40 GHz. The system was running a 64-bit operating system and featured an x64-based processor. It was also equipped with an NVIDIA GeForce RTX 3060 Ti graphics card, which includes 4,864 Compute Unified Device Architecture (CUDA) cores, 152 Tensor cores, and 38 Ray Tracing (RT) cores. Additionally, the computer had 16 gigabytes (GB) of Random Access Memory (RAM). All coding and development tasks were performed using Microsoft Visual Studio. The limitations that we have faced in this study are as follows:

- **Limited Training Data Due to Resources**: Due to computational limits, we used only 40,500 of the available 752,602 CT slices for training and validation, with 128,320 reserved for testing. This restricted sample size may have limited the model's overall performance.

- **Class Imbalance**: The dataset was imbalanced, with the epidural hematoma class having significantly fewer examples than other hemorrhage types. This limited the model's generalization ability and required us to use duplication to increase the sample size for training.
- **Distorted or Unusable Images**: Some dataset images were distorted or unsuitable for training. Although we removed these images, they could still cause issues for future studies using this dataset.
- **Loss of Detail from Resizing**: Images were resized from 512x512 to 224x224 pixels to fit within our computational resources, which may have caused a loss of fine details important for accurately detecting hemorrhages.
- **Partial Patient Scans**: Only a subset of slices per patient was used, meaning some scans did not capture all relevant brain areas of an individual patient. This partial coverage could limit the model's ability to detect every sign of hemorrhage across a full brain scan.

6.3 Future Directions

Our study provides valuable insights by integrating advanced preprocessing techniques and a detailed comparative analysis of the three deep learning models, establishing a strong foundation for future research to improve ICH detection. In future work, we plan to address the current limitations by increasing the number of slices used for training and enhancing class balance to improve model generalizability across all hemorrhage types. We aim to use higher-resolution images to retain more fine-grained details for accurate detection. Additionally, we have conducted our study based on three classifications: Single, Multiple, and No Hemorrhage. In the future, we plan to expand these classifications to detect specific types of hemorrhages and improve patient diagnosis.

Future directions include training lightweight models for clinical scenarios where fast diagnosis is crucial. Furthermore, instead of analyzing 2D CT slices, 3D convolutional networks could capture volumetric features across CT slices, potentially improving accuracy in complex cases. We also plan to adopt a multimodal model approach to integrate additional data types, addressing the growing need for multimodal data fusion and further improving diagnostic accuracy. This study lays a solid foundation for future research in the domain, with a clear path toward more accurate, reliable, and comprehensive AI-driven diagnostic tools.

Chapter 7: References

[1] M. Diringer, "Intracranial Hemorrhage," in *Bradley's Neurology in Clinical Practice*, 7th ed., A. Aminoff, J. Daroff, G. Gronseth, and M. J. S., Eds. Philadelphia, PA, USA: Elsevier, 2020, pp. 1515-1522

[2] J. A. Caceres and J. N. Goldstein, "Intracranial hemorrhage," *Emergency Medicine Clinics of North America*, vol. 30, no. 3, pp. 771–794, Aug. 2012

[3] A. Khairat and M. Waseem, "Epidural hematoma," *StatPearls - NCBI Bookshelf*, Jul. 31, 2023

[4] S. Tenny and W. Thorell, "Intracranial hemorrhage," *StatPearls - NCBI Bookshelf*, Feb. 17, 2024

[5] N. Website, "Subdural haematoma," *nhs.uk*, Nov. 18, 2021

[6] E. Ziu, M. Z. K. Suheb, and F. B. Mesfin, "Subarachnoid hemorrhage," *StatPearls - NCBI Bookshelf*, Jun. 01, 2023

[7] J.-S. Seo, T.-K. Nam, J.-T. Kwon, and Y.-S. Park, "Multiple spontaneous simultaneous intracerebral hemorrhages," *Journal of Cerebrovascular and Endovascular Neurosurgery*, vol. 16, no. 2, p. 104, Jan. 2014

[8] "Intraventricular hemorrhage," *Johns Hopkins Medicine*, Oct. 28, 2024

[9] M. A. Fischer and J. M. Das, "Cerebellar hematoma," *StatPearls - NCBI Bookshelf*, Jun. 17, 2023

[10] B. Rasuli and F. Gaillard, "Pontine haemorrhage," *Radiopaedia.org*, Apr. 20110

[11] D. P. Griswold, L. Fernandez, and A. M. Rubiano, "Traumatic Subarachnoid Hemorrhage: A scoping Review," *Journal of Neurotrauma*, vol. 39, no. 1–2, pp. 35–48, Feb. 2021

[12] O. Kargiotis et al., "Cerebral microbleeds: incidence, imaging characteristics, common and uncommon causes," *Journal of Neurosonology and Neuroimaging*, vol. 10, no. 2, pp. 80–94, Dec. 2018

[13] J. A. Gadde, B. D. Weinberg, and M. E. Mullins, "Neuroimaging of patients in the intensive care unit," *Radiologic Clinics of North America*, vol. 58, no. 1, pp. 167–185, Oct. 2019

[14] "Prescribed drug use and aneurysmal subarachnoid hemorrhage incidence | Neurology," *Neurology*

[15] "Cavernous malformations," Johns Hopkins Medicine, May 13, 2019

[16] J. C. Hemphill et al., "Guidelines for the management of Spontaneous Intracerebral hemorrhage," *Stroke*, vol.

46, no. 7, pp. 2032–2060, May 2015

[17] Q. Bai, Z. Sheng, Y. Liu, R. Zhang, V. W. Yong, and M. Xue, “Intracerebral haemorrhage: from clinical settings to animal models,” *Stroke and Vascular Neurology*, vol. 5, no. 4, pp. 388–395, Mar. 2020

[18] S. J. An, T. J. Kim, and B.-W. Yoon, “Epidemiology, risk factors, and clinical features of intracerebral hemorrhage: an update,” *Journal of Stroke*, vol. 19, no. 1, pp. 3–10, Jan. 2017

[19] D. Rajashekar and J. W. Liang, “Intracerebral hemorrhage,” *StatPearls - NCBI Bookshelf*, Feb. 06, 2023

[20] D. Woo et al., “Risk factors associated with mortality and neurologic disability after intracerebral hemorrhage in a racially and ethnically diverse cohort,” *JAMA Network Open*, vol. 5, no. 3, p. e221103, Mar. 2022

[21] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the inception architecture for computer vision,” *arXiv.org*, Dec. 02, 2015

[22] G. Huang, Z. Liu, V. D. M. Laurens, and K. Q. Weinberger, “Densely connected convolutional networks,” *arXiv.org*, Aug. 25, 2016

[23] M. Tan and Q. Le V., “EfficientNet: Rethinking model scaling for convolutional neural networks,” *arXiv.org*, May 28, 2019

[24] A. Stein et al., "RSNA Intracranial Hemorrhage Detection," *Kaggle*, 2019

[25] M. Maghami, S. A. Sattari, M. Tahmasbi, P. Panahi, J. Mozafari, and K. Shirbandi, “Diagnostic test accuracy of machine learning algorithms for the detection intracranial hemorrhage: a systematic review and meta-analysis study,” *BioMedical Engineering OnLine*, vol. 22, no. 1, Dec. 2023

[26] J. Y. Lee, J. S. Kim, T. Y. Kim, and Y. S. Kim, “Detection and classification of intracranial haemorrhage on CT images using a novel deep-learning algorithm,” *Scientific Reports*, vol. 10, no. 1, Nov. 2020

[27] L. Cortés-Ferre, M. A. Gutiérrez-Naranjo, J. J. Egea-Guerrero, S. Pérez-Sánchez, and M. Balcerzyk, "Deep Learning Applied to Intracranial Hemorrhage Detection," *Journal of Imaging*, vol. 9, no. 2, p. 37, Feb. 2023

[28] A. Kundisch et al., “Deep learning algorithm in detecting intracranial hemorrhages on emergency computed tomographies,” *PLoS ONE*, vol. 16, no. 11, p. e0260560, Nov. 2021

[29] A. ElZemity et al., "A Transformer-Based Deep Learning Architecture for Accurate Intracranial Hemorrhage Detection and Classification," *2023 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT)*, Sakheer, Bahrain, 2023, pp. 215-220

[30] M. Grewal, M. M. Srivastava, P. Kumar, and S. Varadarajan, "RADnet: Radiologist level accuracy using deep learning for hemorrhage detection in CT scans," *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, Washington, DC, USA, 2018, pp. 281-284

[31] M. Seyam, T. Weikert, A. Sauter, A. Brehm, M.-N. Psychogios, and K. A. Blackham, "Utilization of artificial intelligence–based intracranial hemorrhage detection on emergent noncontrast CT images in clinical workflow," *Radiology Artificial Intelligence*, vol. 4, no. 2, Feb. 2022

[32] Kundisch, A., Hönning, A., Mutze, S., Kreissl, L., Spohn, F., Lemcke, J., Sitz, M., Sparenberg, P., & Goelz, L. (2021). Deep learning algorithm in detecting intracranial hemorrhages on emergency computed tomographies. *PloS one*, 16(11).

[33] R. Kumar "Intracranial hemorrhage detection using deep learning and transfer Learning," *NORMA@NCI Library*

[34] M. R. Arbabshirani et al., "Advanced machine learning in action: identification of intracranial hemorrhage on computed tomography scans of the head with clinical workflow integration," *Npj Digital Medicine*, vol. 1, no. 1, Jan. 2018

[35] W. Kuo, C. Häne, P. Mukherjee, J. Malik, and E. L. Yuh, "Expert-level detection of acute intracranial hemorrhage on head computed tomography using deep learning," *Proceedings of the National Academy of Sciences*, vol. 116, no. 45, pp. 22737–22745, Oct. 2019

[36] T. D. Phong, H. N. Duong, H. T. Nguyen, N. T. Trong, V. H. Nguyen, T. V. Hoa, and V. Snasel, "Brain Hemorrhage Diagnosis by Using Deep Learning," in *Proceedings of the 2017 International Conference on Machine Learning and Soft Computing (ICMLSC '17)*, Ho Chi Minh City, Vietnam, 2017, pp. 34–39

[37] B. Rao, V. Zohrabian, P. Cedeno, A. Saha, J. Pahade, and M. A. Davis, "Utility of Artificial Intelligence tool as a prospective radiology peer reviewer — Detection of unreported intracranial hemorrhage," *Academic Radiology*, vol. 28, no. 1, pp. 85–93, Feb. 2020

[38] H. Lee et al., "An explainable deep-learning algorithm for the detection of acute intracranial haemorrhage from small datasets," *Nature Biomedical Engineering*, vol. 3, no. 3, pp. 173–182, Dec. 2018

[39] S. Chilamkurthy et al., "Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study," *The Lancet*, vol. 392, no. 10162, pp. 2388–2396, Oct. 2018

[40] S. Kaur and A. Singh, "A new deep learning framework for accurate intracranial brain hemorrhage detection and classification using Real-Time collected NCCT images," *Applied Magnetic Resonance*, vol. 55, no. 6, pp.

629–661, Jun. 2024

[41] L. Cortés-Ferre, M. A. Gutiérrez-Naranjo, J. J. Egea-Guerrero, S. Pérez-Sánchez, and M. Balcerzyk, “Deep learning applied to intracranial hemorrhage detection,” *Journal of Imaging*, vol. 9, no. 2, p. 37, Feb. 2023

[42] A. Hussain et al., “An Attention-Based ResNET Architecture for Acute Hemorrhage Detection and Classification: Toward a Health 4.0 Digital Twin Study,” *IEEE Access*, vol. 10, pp. 126712–126727, Jan. 2022

[43] B. A. Mohammed et al., “Multi-Method diagnosis of CT images for rapid detection of intracranial hemorrhages based on deep and hybrid learning,” *Electronics*, vol. 11, no. 15, p. 2460, Aug. 2022

[44] H. Ye et al., “Precise diagnosis of intracranial hemorrhage and subtypes using a three-dimensional joint convolutional and recurrent neural network,” *European Radiology*, vol. 29, no. 11, pp. 6191–6201, Apr. 2019

[45] Y.-H. Li, L. Zhang, Q.-M. Hu, H.-W. Li, F.-C. Jia, and J.-H. Wu, “Automatic subarachnoid space segmentation and hemorrhage detection in clinical head CT scans*,” International Journal of Computer Assisted Radiology and Surgery*, vol. 7, no. 4, pp. 507–516, Nov. 2011

[46] G. P. Reddy and Y. V. P. Kumar, "Explainable AI (XAI): Explained," *2023 IEEE Open Conference of Electrical, Electronic and Information Sciences (eStream)*, Vilnius, Lithuania, 2023, pp. 1-6

[47] K. H. Kim, H.-W. Koo, B.-J. Lee, S.-W. Yoon, and M.-J. Sohn, “Cerebral hemorrhage detection and localization with medical imaging for cerebrovascular disease diagnosis and treatment using explainable deep learning,” *Journal of the Korean Physical Society*, vol. 79, no. 3, pp. 321–327, May 2021

[48] M. Burduja, R. T. Ionescu, and N. Verga, “Accurate and Efficient Intracranial Hemorrhage Detection and Subtype Classification in 3D CT Scans with Convolutional and Long Short-Term Memory Neural Networks,” *Sensors*, vol. 20, no. 19, p. 5611, Oct. 2020

[49] “NCCT Image-based Automated intracranial hemorrhage classification using a deep learning model: Is the intracranial hemorrhage classification of very small volume hematoma cases reliable?”, *NCCT*

[50] “Non-contrast head/brain CT CQ500 Dataset,” *Academic Torrents*

[51] M. Toikkanen, D. Kwon, and M. Lee, “ReSGAN: Intracranial Hemorrhage Segmentation with Residuals of Synthetic Brain CT Scans,” in *Lecture notes in computer science*, 2021, pp. 400–409

[52] N. Donges, “What is transfer learning? Exploring the popular deep learning approach.,” *Built In*, Aug. 15, 2024

[53] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv.org*, Dec. 22, 2014

[54] M. Tan and Q. Le V., “EfficientNet: Rethinking model scaling for convolutional neural networks,” *arXiv.org*, May 28, 2019

[55] L. N. Smith, “Cyclical learning rates for training neural networks,” *arXiv.org*, Jun. 03, 2015

[56] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 770-778