Addis Ababa University

COLLEGE OF TECHNOLOGY AND BUILT ENGINEERING

SCHOOL OF BIOMEDICAL ENGINEERING

INTRACRANIAL HEMORRHAGE DETECTION AND SUB-TYPE CLASSIFICATION USING A DUAL-BRANCH XCEPTION–KNN FRAMEWORK

by

Lemlem Tiruneh

In partial fulfillment of the requirements for the degree of Master of Science in Biomedical Engineering

Supervisor

Dr. Melkamu Hunegnaw

Addis Ababa, Ethiopia

July 22, 2025

Declaration

By submitting this MSc thesis, I declare that the entirety of the work presented in this document is my original work. Where information has been derived from other sources, I confirm that this has been indicated in the thesis and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Name: Lemlem Tiruneh *Signature:* ..... *Date:* .....

This MSc thesis has been submitted for examination with my approval as the main advisor.

Name: Dr. Melkamu Hunegnaw *Signature:* ..... *Date:*.....

Approval

This is to certify that the thesis prepared by Lemlem Tiruneh entitled: "INTRACRANIAL HEMORRHAGE DETECTION AND SUB-TYPE CLASSIFICATION USING A DUAL-BRANCH XCEPTION–KNN FRAMEWORK" submitted in partial fulfillment of the requirements for the Degree of Master of Science in Biomedical Engineering (Bioinstrumentation and Imaging) complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Approved by Board of Examiners:

Supervisor's Name _____ *Sign:*_____*Date:*_____

External Examiner's Name _____ *Sign:*_____ *Date:*_____

Internal Examiner's Name _____ *Sign:*_____ *Date:*_____

Chairman's Name _____ *Sign:*_____ *Date:*_____

Dedication

This research is dedicated to all those who have lost their lives, or health and quality of life due to Intracranial Hemorrhage, and to people who lost their close ones because of this disease.

Acknowledgments

First of all, I would like to thank God Almighty for giving me the opportunity and guidance to achieve my goal and to be successful in this part.
I would like to extend my sincere gratitude to my supervisor, Dr. Melkamu Hunegnaw, for his invaluable expertise, patience, and understanding, all of which significantly enriched my graduate experience. I deeply appreciate his continuous encouragement and insightful feedback, which were instrumental in shaping my research.
I am profoundly grateful to my family and friends for their unwavering support and encouragement throughout my studies and during the process of researching and writing this thesis. This achievement would not have been possible without their unwavering belief in me.
Finally, I wish to thank the university staff, Wudassie Diagnostic center staff and Yanet General Hospital staff especially, Dr. Abel Tesfaye who have supported me throughout this journey. Their assistance and encouragement have been truly invaluable.

Abstract

Intracranial hemorrhage (ICH) is any bleeding inside the skull. It is a life-threatening disease that requires timely diagnosis and urgent medical treatment. Most often, diagnosis of head CT is common. However, there is a problem of delayed, missed diagnosis and misdiagnosis due to subtle, highly variable size, shape, and location of ICH and subtypes. Workload also leads to interpretation errors which are misdiagnosis and missed diagnosis. This research leverages a two-branch convolutional neural network (CNN) architecture—each branch based on Xception—and then fuses their learned features via concatenation before a final K-Nearest Neighbors (KNN) classification step. Such a hybrid approach combines the power of deep feature extraction (CNN) with the flexibility and interpretability of traditional machine learning (KNN). Using an extremely large data set that encompasses approximately 4.5 million slices, this deep learning algorithm passed three experiments. The first experiment is Normal vs. Abnormal Classification, which is a binary classification task where each CT slice is labeled as either 'normal' or 'abnormal.' This experiment successfully distinguishes between normal and abnormal head CT scans with an accuracy of approximately 0.94, with an F1 score near 0.947. The second experiment, multiclass Classification of distinct hemorrhage types, proves largely successful. The model excels in detecting intraparenchymal, subdural, subarachnoid, and intraventricular bleeds, as well as normal scans, each class achieving an F1-score above 0.82. Normal scans also see a robust F1 (0.934). Class-specific accuracies remain high (roughly 0.95–0.97 for most). However, accuracy alone can be misleading in imbalanced scenarios or if some hemorrhage types are comparatively rare. Classes like intraparenchymal, intraventricular, subarachnoid, and subdural show high precision (0.81) and recall (0.80), reflecting consistent identification of these bleeds. The third and final experiment assessed how well the end-to-end pipeline generalizes to completely unseen data collected from Wudassie Diagnostic Center and Yanet General Hospital, Addis Ababa, Ethiopia. Of the 15 cases at Wudassie Diagnostic Center, 8 were correctly classified, yielding an overall accuracy of 0.53 and of the 23 cases at Yanet General Hospital, 16 were correctly classified, yielding an overall accuracy of 0.65. These figures align closely with the confusion-matrix analysis and underline the model's promise for triage support in local hospitals. The findings hold immediate relevance for emergency triage in health-care systems where certified neuroradiologists are scarce. This tool reliably highlights acute intracranial hemorrhage with sensitivity above 0.90. In addition, the system's slice-wise probability maps and subject-level majority decisions augment radiologists rather than replacing their judgment, thus reducing cognitive load without affecting professional agency.

Key words: Intracranial hemorrhage (ICH), Deep Learning, Head CT, Image classification

Contents

List of Figures

List of Tables

List of Acronyms

2D	Two Dimensional
3D	Three Dimensional
AI	Artificial Intelligence
AIH	Acute Intracranial Hemorrhage
AUC	Area Under the Curve
AVM	Arteriovenous Malformations
BP	Blood Pressure
CAA	Cerebral Amyloid Angiopathy
CAD	Computer-Aided Design
CNN	Convolutional Neural Networks
CSV	Comma Separated Values
CT	Computed Tomography
CTA	Computed Tomography Angiography
DICOM	Digital Imaging and Communications in Medicine
DL	Deep Learning
DLHD	Deep Learning Hemorrhage Detection
ED	Emergency Departments
EDH	Epidural Hemorrhage
EMT	Emergency Medical Technician
FC	Fully Connected
FN	False Negative
FP	False Positive
GBD	Global Burden of Diseases

GLCM	Grey Level Co-occurrence Matrix
Grad-CAM	Gradient-weighted Class Activation Mapping
GRU	Gated Recurrent Unit
HCT	Head Computed Tomography
HU	Hounsfield Unit
ICH	Intracranial Hemorrhage
INR	International Normalized Ratio
IPH	Intraparenchymal Hemorrhage
IVH	Intraventricular Hemorrhage
KNN	K-Nearest Neighbors
LSTM	Long Short-Term Memory
MATLAB	Matrix Laboratory
ML	Machine Learning
MLP	Multilayer Perceptron
MMD	Moyamoya Disease
MRI	Magnetic Resonance Imaging
NCCT	Non-Contrast Computed Tomography
NCD	Non-Communicable Disease
PIVH	Primary Intraventricular Hemorrhage
REPTree	Reduced Error Pruning Tree
ResNet	Residual Network
RGB	Red, Green, Blue
RNN	Recurrent Neural Network
ROC	Receiver Operating Characteristic
ROI	Region of Interest
RSNA	Radiological Society of North America
SAH	Subarachnoid Hemorrhage
SDH	Subdural Hemorrhage
SVM	Support Vector Machine

TBI	Traumatic Brain Injury
TN	True Negative
TP	True Positive
UMMC	University Of Malaya Medical Center
VGG	Visual Geometry Group
ViT	Vision Transformer
WHO	World Health Organization
Xception	Extreme Inception
YLL	Years of Life Lost

Chapter 1

Introduction

1.1 Background of the Study

Intracranial hemorrhage (ICH) is any bleeding inside the skull, including in the meningeal spaces around the brain and the brain parenchyma [9]. By preventing blood flow to the brain, it can destroy brain cells by depriving them of oxygen and increasing pressure in the brain. ICH can be either spontaneous or trauma-related [10]. A number of disease processes, including but not limited to cerebral amyloid angiopathy, ruptured aneurysms, tumors, venous sinus thrombosis, hypertension, and hemorrhagic conversion of strokes, can cause spontaneous ICH [9]. Subdural hemorrhage (SDH), subarachnoid hemorrhage (SAH), intracparenchymal hemorrhage (IPH), epidural hemorrhage (EPH), and intraventricular hemorrhage (IVH) are the various types of ICH, which is the abnormal buildup of blood inside the cranial vault [9]. Either primary or secondary ICH may occur. Small-vessel disease-related bleeding is included in primary ICH while secondary ICH covers additional causes of ICH, such as in a number of coagulopathies. There is a substantial correlation between ICH and brain tumors, pregnancy, vascular abnormalities, coagulopathy, and hypertension.

Incidence of spontaneous ICH is 24.6 per 100,000 person-years globally with a monthly mortality rate ranging from 35% to 52%. Only 20% of survivors are anticipated to achieve fully functional recovery after 6 months. Approximately half of this mortality occurs within the first 24 hours [9]. Emergency rooms must prioritize providing fast and efficient care. Stroke, which accounts for over 5.5 million deaths annually and is one of the second major causes of mortality worldwide, is caused in part by ICH [11]. In Ethiopia, stroke is becoming a major public health concern. Recent research indicates that Ethiopia is starting to experience an epidemiological shift, in which the burden of non-communicable diseases (NCDs) such as cancer, diabetes, chronic respiratory illnesses, and injuries is rising while the prevalence of communicable diseases remains high [12]. With better blood pressure management, the prevalence of hypertensive ICH has reduced in developed countries. However, in developing countries, the burden of ICH has not decreased. More than two-third (70%) of strokes occur in low- and middle-income countries [12].

Stroke is one of the most prevalent NCDs, according to the WHO. Stroke was the second most prevalent cause of mortality and the third most common cause of disability worldwide in 2020, according to the Global Burden of Diseases, Injuries, and Risk Factors study (GBD 2021) [13]. Stroke ranked as the most burdensome NCD the largest overall cause of YLL in Ethiopia in 2021 when compared to other causes of burden [14]. Due to

stroke, 32860 deaths occurred in Ethiopia, by 2016. Stroke deaths in Ethiopia reached 6.23% of all fatalities, according to the WHO data published in 2017. Furthermore, there are 89.82 deaths from stroke per 100,000 people in the nation, when age is taken into account [15]. Although less frequent than ischemic stroke, ICH results in greater disability and a higher death rate. As a result, it significantly impacts health systems, economics, and society [13]. According to the World Stroke Organization, the number of stroke deaths would rise by 50% between 2020 and 2050 [13]. This clearly demonstrates how its prevalence increases over time.

1.2 Risk factors of intracranial hemorrhage

The two main ICH risk factors are cerebral amyloid angiopathy (CAA) and hypertension [3]. The risk of hypertension-related ICH rises with increasing blood pressure (BP) values, and it is more likely to occur in deep structures. CAA-related ICH mostly occurs in lobar regions and age is a common factor in the development of CAA [12]. Primary intraventricular hemorrhage (PIVH) is characterized by bleeding within 15 mm of the ventricular wall or in the ventricular system without a noticeable parenchymal component. PIVH is an uncommon condition that makes up 3% of all spontaneous ICH. Arteriovenous malformations (AVMs), aneurysms, moyamoya disease (MMD), coagulopathy, and arteriovenous fistulas have all been associated with hypertension, which is a common risk factor [16]. Other risk factors for ICH include;

- **Alcohol consumption:** This risk seems to be dose-dependent, with a higher risk of ICH among individuals who drink more alcohol each day. This risk may be due to acute BP changes during consumption and withdrawal, effects on platelet function and coagulation, and malfunction of the vascular endothelium [3].

- **Diabetes:** Diabetes was shown to be a risk factor for ICH in a meta-analysis with nearly 70,000 participants [3].

- **Genetics:** The apolipoprotein E gene and e2 and e4 alleles are the gene most strongly linked to ICH. The expansion of hematomas was also connected to the existence of the e2 allele [3].

- **Cholesterol:** Low total serum cholesterol levels are a risk factor for ICH, unlike to ischemic stroke, for which high cholesterol levels are a risk [3].

- **Substance abuse:** Patients who were actively using cocaine at the time of their ICH had considerably more severe presentations and worse outcomes. Sympathomimetic substances, such as cocaine, are risk factors for ICH [3].

- **Anticoagulation:** In patients with atrial fibrillation and other cardiovascular and prothrombotic conditions, oral anticoagulants are frequently used as prophylaxis.

Warfarin-using patients have an annual risk of ICH that ranges from 0.3% to 1.0% per patient-year, with significantly increased risk when the international normalized ratio (INR) is higher than 3.5 [3].

1.3 Pathophysiology

Primary ICH is typically a manifestation of underlying small vessel disease. First, persistent hypertension results in hypertensive vasculopathy, which causes microscopic degenerative changes in the walls of small-to-medium penetrating vessels, called lipohyalinosis [9]. Second, CAA is characterized by deposited Amyloid-beta peptide in the walls of tiny leptomeningeal and cortical vessels [3]. The end effects of amyloid buildup include degenerative changes in the vessel wall defined by the loss of smooth muscle cells, luminal narrowing, wall thickening, microaneurysm development, and microhemorrhages, even though the underlying mechanism leading to the accumulation of amyloid is still unknown [3].
The hematoma damages directly and mechanically the brain parenchyma after the initial vessel rupture. The peak of perihematomal edema occurs between 10 and 20 days after the initial 3 hours of symptom start. Then, secondary injury processes such as an inflammatory response, activation of the coagulation cascade, and iron deposition from hemoglobin breakdown are mediated via blood and plasma products. Finally, over the first 24 hours, the hematoma can continue to grow in up to 38% of patients [9].

1.4 Types of intracranial hemorrhage

Epidural, subdural, subarachnoid, intraventricular, and intraparenchymal hemorrhage are the five major forms of hemorrhage included in ICH [17]. Figure 1.1 shows the anatomical structure of ICH. There are differences between each type of bleeding in terms of causation, findings, prognosis, and outcome. Epidural hemorrhage, subdural hemorrhage, and intraparenchymal hemorrhages differ one from the other by their size, shape and location [18]. Epidural hemorrhage has convex shape and fits closely with the skull. Subdural hemorrhage is also close to the skull but larger than epidural hemorrhage and has crescent-like shape with concave surface away from the skull. Intraparenchymal hemorrhage is distant from the skull. Subarachnoid hemorrhage is usually thread-like, small and punctate in CT images [19].

1.4.1 Epidural Hemorrhage

Blood that dissects into the possible space between the dura and inner table of the skull might result in epidural hematomas [17]. In 85% to 95% of instances, this happens following a skull fracture. An arterial or venous vessel may sustain injury, allowing blood to dissect into the possible epidural area and cause an epidural hematoma [17]. The middle meningeal artery is the most often injured vessel, which results in arterial bleeding into the possible epidural space underlying the temporoparietal portion of the skull. About 2% of people who suffer from head trauma have epidural hematomas, which are responsible for 5% to 15% of fatal head injuries. When there is a skull fracture,

the venous blood from the fracture fills the epidural area, resulting in a venous epidural hematoma. Pediatric patients frequently get venous epidural hematomas [17].

1.4.2 Subdural Hemorrhage

Typically, a vessel that travels between the brain and skull is stretched, ruptured, or ripped, causing it to start bleeding into the subdural region which is anatomically the arachnoid space [18]. These most frequently happen after a blunt head injury, but they can also happen spontaneously or after penetrating head injuries. The most prevalent focal intracranial lesion, appearing as the primary initial lesion in 24% of individuals with severe closed head traumas, is a subdural hematoma [17]. Between 5% and 25% of patients with a serious head injury are thought to have subdural hematomas. With a male to female ratio of 2:1, the annual incidence is one to five instances per 100,000 people. The incidence of subdural hematomas increases throughout life [17].

1.4.3 Subarachnoid Hemorrhage

Subarachnoid bleeding is bleeding that fills the subarachnoid space between the pia mater and the arachnoid mater, which are layers of the membranes surrounding the brain (meninges) [18]. Traumatic and non-traumatic subarachnoid hemorrhages are the two categories of subarachnoid hemorrhage. The rupture of a brain aneurysm is the most prevalent cause of non-traumatic subarachnoid hemorrhage. The subarachnoid space may be filled with blood when an aneurysm ruptures. Arteriovenous malformations (AVM), the use of blood thinners, trauma, or idiopathic causes are additional causes of subarachnoid hemorrhage [17].
Incidence rates for those over the age of 35 range from two to twenty-five per 100,000 person- years for subarachnoid hemorrhage, which accounts for around 5% of all strokes [17]. It may be very slightly more common in females than in males, and the prevalence progressively increases with patient age with an estimated female to male ratio of around 1.15:1 [17].

1.4.4 Intraparenchymal Hemorrhage

Bleeding into the actual brain parenchyma is known as intraparenchymal hemorrhage. The most frequent cause of non-traumatic intraparenchymal hemorrhage is hypertensive damage to cerebral blood arteries, which eventually rupture and bleed into the brain [18]. Hemorrhage can happen for a wide range of reasons, such as hypertension, amyloid angiopathy, arteriovenous malformation, tumor, vasculitis, aneurysm rupture, infection, coagulopathy, and trauma, among others. Of all strokes, intraparenchymal hemorrhage accounts for 10% to 20% [17]. For people 55 years of age and older, the incidence of intraparenchymal hemorrhage rises with age. Regarding gender differences, there is considerable debate; however there may be a slight male predominance [17].

1.4.5 Intraventricular hemorrhage

Subarachnoid hemorrhage and superficial contusions are injuries the most frequently related to intraventricular hemorrhage associated with acute head trauma [20]. About

one-third of patients experience secondary spread from frontal, temporal, and basal ganglia hematomas, but the majority are likely to experience additional mechanisms, such as hypoxia, coagulopathy, negative pressure-induced rupture of subependymal veins, or diffuse axonal damage [20]. Only 1.5%-3% of individuals with acute head trauma and 10% of patients with severe head injury have intraventricular blood; however, between 25% and 45% of aneurysmal SAH and approximately 45% of spontaneous ICH extend in to the ventricles [20]. ICH, SAH, and traumatic brain injury all have high rates of morbidity and mortality due to intraventricular hemorrhage. Figure 1.2 shows the summary of each ICH subtypes.

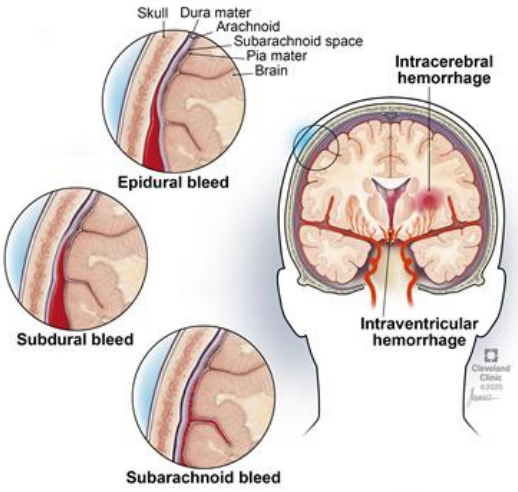


Figure 1.1: Intracranial hemorrhage and its sub-types [1]

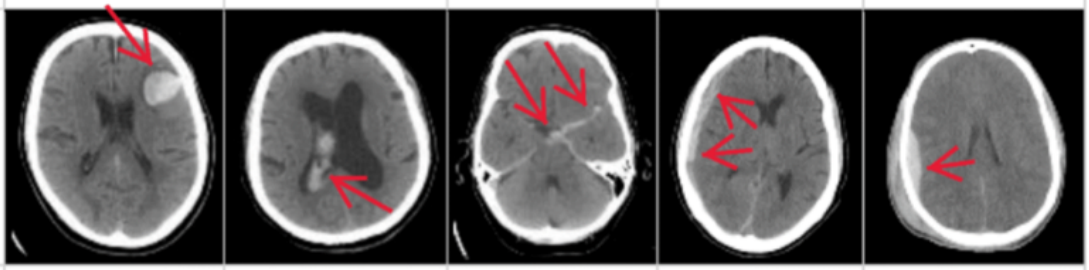
	Intraparenchymal	Intraventricular	Subarachnoid	Subdural	Epidural
Location	Inside of the brain	Inside of the ventricle	Between the arachnoid and the pia mater	Between the Dura and the arachnoid	Between the dura and the skull
Imaging					
Mechanism	High blood pressure, trauma, arteriovenous malformation, tumor, etc	Can be associated with both intraparenchymal and subarachnoid hemorrhages	Rupture of aneurysms or arteriovenous malformations or trauma	Trauma	Trauma or after surgery
Source	Arterial or venous	Arterial or venous	Predominantly arterial	Venous (bridging veins)	Arterial
Shape	Typically rounded	Conforms to ventricular shape	Tracks along the sulci and fissures	Crescent	Lentiform
Presentation	Acute (sudden onset of headache, nausea, vomiting)	Acute (sudden onset of headache, nausea, vomiting)	Acute (worst headache of life)	May be insidious (worsening headache)	Acute (skull fracture and altered mental status)

Figure 1.2: Summary of various subtypes of hemorrhages and their location inside the skull [2]

1.5 Overview of common ICH diagnosis methods

1.5.1 Blood Tests

Complete blood count, electrolytes, creatinine, glucose, and coagulation investigations ought to be performed on patients with ICH [3].

1.5.2 Neuroimaging

I. Non-contrast computed tomography
Given its widespread availability, noncontrast computed tomography (NCCT) is regarded as the gold standard for the diagnosis of ICH in the emergency departments (ED). NCCT is a quick technique with great sensitivity for diagnosing acute ICH [3]. Beyond ICH diagnosis, NCCT can offer helpful information on ICH location, intraventricular extension, hydrocephalus, presence and severity of edema, and midline shift or brainstem compression as a result of the mass effect from the hematoma. Additionally, ICH volume may be quickly calculated in the ED using the $ABC/2$ approach and is a strong predictor of ICH outcome [3].
II. Computed tomography angiography
In the acute setting of ICH, computed tomography angiography (CTA) is a helpful diagnostic technique. It is the most extensively used noninvasive method for identifying vascular abnormalities as ICH's secondary causes [3]. An independent predictor of hematoma enlargement and a poor prognosis in patients with supratentorial ICH is the presence of contrast extravasation within the hematoma on CTA images, popularly known as the "spot sign." Additionally, the presence of the CTA spot sign during surgical evacuation is linked to active bleeding and may assist identify which patients will benefit from having surgery [9]. The cost and extra radiation exposure of CTA are its principal disadvantages. Although there is controversy in the literature concerning the existence of this entity and there is no proof that CTA raises the risk of nephropathy in patients with ICH, some doctors are concerned about the possibility of contrast- induced nephropathy [9].
Procedure

1. Select the axial CT image with the largest area of hemorrhage.
2. Measure the largest ICH diameter on this slice (A).
3. Measure the largest diameter perpendicular to A (B).
4. Multiply the slice thickness by the number of slices in which the hematoma is visible (C).
5. Multiply A, B and C and divide the results by 2.
6. A, B, C are expressed in centimeters and the resulting volume in cm3 or milliliters.

Figures 1.3 and 1.4 shows the volume of ICH in CTA image.
III. MRI
The diagnostic accuracy of MRI for the diagnosis of ICH is comparable to that of NCCT [9]. To identify underlying secondary causes of ICH, such as neoplastic tumors or hemorrhagic transformation of ischemic stroke, MRI can be a valuable tool [3]. In addition, contrast-free brain vascular imaging is possible with magnetic resonance angiography

Figure 1.3: ICH image of CTA [3]

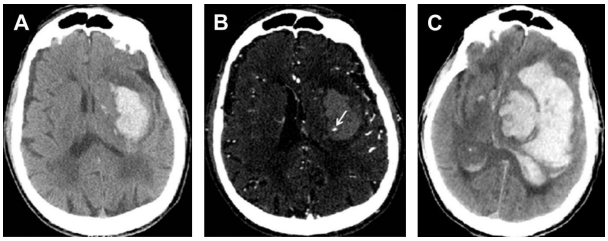


Figure 1.4: (A) Left deep ICH on NCCT, with baseline volume of 45 mL; (B) CTA showing presence of spot sign (arrow); (C) follow-up NCCT at 19 hours showed significant hematoma growth to a volume of 192 mL with severe midline shift and massive intraventricular extension [3]

in people with impaired renal function, contrast allergies, or other contraindications to CTA [3]. Magnetic resonance imaging is rarely used in the ICH emergency department examination due to the expense, length of the examination, and poor tolerability for some patients.

1.6 Problem statement

Emergency rooms must prioritize providing fast and efficient care. Stroke, which causes more than 5.5 million deaths annually and is one of the second leading causes of mortality worldwide, is caused in part by ICH [11]. Stroke deaths in Ethiopia reached 6.23% of all fatalities, according to WHO data published in 2017 [15]. Furthermore, there are 89.82 stroke deaths per 100,000 people in the nation, when age is taken into account. With better blood pressure management, the prevalence of hypertensive ICH has been reduced in developed countries. However, in developing countries, the burden of ICH has not decreased. More than two thirds (70%) of strokes occur in low- and middle-income countries [12]. Intracranial hemorrhage kills brain cells by interrupting blood flow to the brain, causing oxygen deprivation and increasing additional pressure in the brain [2]. Therefore, a timely and accurate diagnosis and treatment is essential to prevent complications such as expansion, block the release of neuropathic products from hematomas, decrease mass-effects, and thus prevent the initiation of pathological processes. The main diagnostic method for identifying the location and size of the bleeding is head computed tomography (CT) due to its noninvasive nature and rapid 3D imaging ability [9]. Using

the current diagnostic method, there is a problem of delayed, missed diagnosis, and misdiagnosis due to the subtle, highly variable size, shape, and location of ICH and subtypes. The diagnosis of acute ICH and its subtypes may also be challenging using CT images with calcifications and artifacts. The other reason is the increased workload for radiologists. Currently, the number of patients undergoing head CT examinations is rising, which increases the workload for radiologists and results in fatigue and interpretation errors, including misdiagnoses and missed diagnoses. In other countries, peer review by double reading of scans has been shown to reduce diagnosis error problems, but it is labor-intensive and requires more radiologists to implement, which is more difficult in developing countries. Therefore, an automated ICH detection method that can assist healthcare professionals is needed.

1.7 Objectives of the study

1.7.1 General Objective

The general objective of this study is to develop an automated system for the detection of ICH and classification of main subtypes using deep learning from head CT scans.

1.7.2 Specific Objectives

- To collect and prepare a clinical image dataset with the corresponding label from both public datasets and local Ethiopian hospitals for validation.
- To develop binary classification model for intracranial hemorrhage
- To develop classification model for automatic classification of ICH sub types.
- To test the proposed system on local Ethiopian data.

1.8 Significance of the study

ICH can be treated if a timely and effective diagnosis is in place and this can lead to reduced morbidity and mortality due to ICH and related diseases. In developed countries, ICH can be controlled using regular control of blood pressure. While in developing countries, it has shown that a higher percentage of ICH is recorded. The proposed functional system can acquire slice images of the brain from head CT, process the images and provide accurate diagnosis results in terms of the ICH and its subtypes. Diagnosis of ICH based on the acquired images will assist the radiologists to detect ICH and its subtypes within the area of the bleeding. It will prevent delayed and misdiagnosis of ICH. It also prevents ineffectual treatments and allows healthcare providers to give proper referral for cases that require more advanced treatment. This may result in a reduced prevalence of mortality due to ICH and hemorrhagic stroke.

1.9 Scope of study

The scope of this study is on designing a digital, automated and intelligent system that is used for diagnosis of ICH from head CT with optimized accuracy and reliability. The

final scope of this research is to build the software for the system which automates the diagnosis system and then train and test the designed framework on two datasets: RSNA Intracranial Hemorrhage Detection Dataset and Ethiopian Hospital Dataset.

1.10 Research Motivation

In Ethiopia, ICH and associated conditions, including stroke, are on the rise. However, with early and precise diagnosis, ICH can be treated and people can avoid death or additional complications. A close relative of mine also passed away from ICH during pregnancy as a result of a delayed diagnosis. It was too late when she was diagnosed with IVH. She and others who died of ICH like her served as my inspiration for this study. By developing an automated system tool that uses deep learning mechanisms to help diagnosis systems, this research aimed to address the issue of interpretation mistakes, specifically misinterpretation and missed diagnosis of ICH.

1.11 Thesis Structure

The remainder of this thesis is structured as follows: Chapter 2 reviews relevant literature and theoretical frameworks pertinent to the study. Chapter 3 outlines the research methodology employed, detailing the techniques and analytical approaches used. Chapter 4 presents and discusses the research findings, while Chapter 5 summarizes the key conclusions, discusses implications, and provides recommendations for future research.

Chapter 2

Literature Review

2.1 Related Works

The literature review chapter critically evaluates existing studies and theoretical frameworks relevant to this research. Science Direct, Google Scholar, ResearchGate, IEEE Xplore, and PubMed Central or National Library of Medicine were employed as research sources to find reliable sources for this research. The study entitled "A deep learning algorithm for automatic detection and classification of acute intracranial hemorrhages in head CT scans" was conducted by X. Wang, T. Shen, S. Yang, J. Lan, Y. Xu, M. Wang, J. Zhang, and X. Han. The goal of the study was to use deep learning algorithms in head CT images to automate the detection of intracranial hemorrhage and its subtypes [4]. They used dataset from 2019- Radiological Society of North America (RSNA) Brain CT Hemorrhage Challenge which was collected from three institutions. They implemented a multi-label classification scheme. They validate their algorithm on two independent external datasets, namely, the PhysioNet-ICH dataset and the CQ500 dataset. Their system has three stages. A CNN classifier is trained using 2D slices and manual ICH labels as ground truth in the initial stage. Once trained, the CNN classifier may be used on every slice of a head scan to estimate the likelihood that each ICH subtype exists. The final convolution layer's output feature map, which is used as an abstract representation of the input image, allows the CNN classifier to also function as a feature extractor. Subsequently, the second stage of the model applies a bi-directional RNN with the GRU unit to the feature outputs from the first stage classifier for each slice of the 3D scan to produce a more accurate estimation of the ICH subtypes for each slice. In order to provide spatially coherent estimation of ICH occurrences, the RNN model additionally acts as a regulator and takes into account 3D context information. The RNN model mimics the method through which radiologists review images in real time. In order to implement the concept of stacked generalization ensemble, the prediction outputs from the first classifier and the Sequence Model 1 for all corresponding slices of a 3D scan are assembled together and used as the input to the second RNN model in the final stage. This RNN model also has the GRU unit. On each slice of an input head CT, the Sequence Model 2 outputs the final prediction of ICH subtypes after training [4]. Saliency maps generated by the Grad-CAM technique help visualize the model's decision-making process, highlighting relevant regions in the images. They employed slice thickness information as an extra input feature to the Sequence Model 2 so that the model can automatically accommodate variations in slice thickness. The model of this study is shown in Figure 2.1. This study result demonstrated a valuable accuracy level in classifying acute ICH and its five

subtypes, with AUCs ranging from 0.983 to 0.996 and they discussed that the accuracy was the highest for the IVH subtype on the RSNA test data; it appears as bright regions inside the brain ventricles due to the relatively fixed position of IVH. On the other hand, the SDH was the most difficult to detect as indicated in this study.

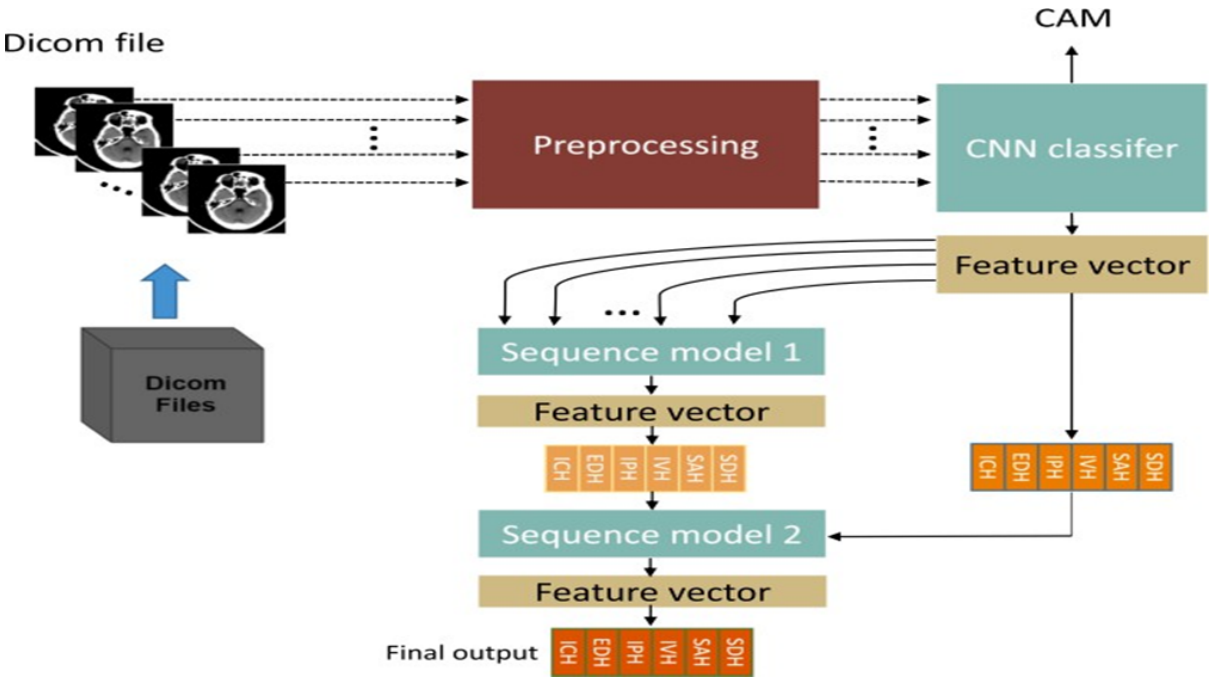


Figure 2.1: Block diagram of the model [4]

Strength of the study

- Takes better account of both intra-slice and inter-slice image information.
- The multi-stage design also allows the RNN model at the last stage to automatically learn to correct prediction errors of the models in the earlier stages.
- Dataset size is large enough to train the system.
- Usage of external dataset for testing and validation.
- It has high accuracy of classification except for SAH and SDH.

Limitations of this model

- The classification accuracy depends on the slice thickness, image accusation protocol and scanner machine.
- High degree of difficulty to identify SAH and SDH subtypes, and they can be mistaken for other subtypes.
- AI model cannot automatically learn to locate detailed delineations of the lesion areas.

- The training of the whole model is very complex and time consuming and inference time is longer than a single model.

In another study, entitled "Hybrid Thresholding Method in Detection and Extraction of Brain Hemorrhage on the CT-Scan Image", conducted by S. Sumijan, Y. Yuhandri, and W. Boy, a method was developed to identify and extract brain bleeding from each image slice of patient's CT scan in order to compute the volume of the bleeding in the brain [21]. This study used a hybrid thresholding technique to identify and extract brain bleeding from a CT scan image. Their method was able to segment and extract brain bleeding with 96.43% overall accuracy [21].

Strength of this research

- It has 96.43% accuracy.
- It has less complexity and easy to train.
- Area of brain bleeding is calculated.

Limitations

- The algorithm is based on black and white color segmentation and could potentially detect other brain abnormalities such as brain tumors and confuse it with ICH.
- Sensitivity and specificity of this method is questionable.
- No classification of ICH subtypes.

Another study proposed a Computer Aided Detection (CAD) system for use in identification of acute intracranial bleeding on CT images [22]. The goal of this research was to create CAD as a triage tool that prioritizes which images radiologists should read. The work exploited intensity thresholding and symmetry to register volumes and detect bleeding in 3D. Data from University of Malaya Medical Centre (UMMC) was used to establish a training and validation set. A total of 119 volumes made up the training set, 54 of which had at least one acute cerebral hemorrhage. 40 out of the 108 volumes in the validation set have at least one acute intracranial bleeding. Slices were initially chosen in such a way that the slices around the skull base and tip of the head are skipped. Then, thresholding was applied to estimate the difference between acute bleed density and normal brain density which resulted in the creation of a binary mask. To get rid of microscopic fragments, morphological techniques including erosion and dilation were used. As opposed to evaluating the algorithm's performance per lesion, it was evaluated per patient. Per volume, the examination took about 5 seconds. Using a patient-to-patient basis, the CAD achieved a sensitivity of 75.0%, a specificity of 83.8%, and a precision of 80.6%.

Strength of the study

- It has very fast evaluation time (around 5 seconds per volume).

Limitations

- Sensitivity to detect hemorrhage density near the skull is poor due to the brightness and density in that region.
- Specificity to detect acute intracranial hemorrhage is mostly missed (false negative results).
- This method has lower sensitivity, specificity and overall accuracy compared to the other models.
- It does not consider SAH subtype of ICH.
- Small amount of dataset was used to train the system.

Another study proposed by Barhoumi Yassine and Ghulam Rasoo introduced Scopeformer, a hybrid architecture that combines convolutional neural networks (CNNs) and a Vision Transformer (ViT) for the classification of ICH [5]. In order to improve the information content and resolution for the ViT to process, the model used multiple CNNs to extract feature maps. The feature set is further enhanced and performance is increased by pre-training these CNNs with a variety of paradigms [5]. Global attention encoding and feature map generation are the two primary phases of the design. An input image is processed by several CNN feature extractors in the first stage, and the results are concatenated. Concatenated feature maps are used to extract patches, which are then fed into a Transformer Encoder. The output of the Transformer is then sent to a multilayer perceptron (MLP) head for classification in the second stage [5]. The RSNA cerebral hemorrhage dataset was used for the experiments. With 12 encoder layers and a latent vector dimension of 1456, the base ViT variation was used. Multi-label weighted mean logarithmic loss was used to assess the models' performance, showing that the number of Xception models utilized directly correlates with classification accuracy. Compared to models with fewer CNNs or consistent pre-training, those with numerous CNNs and a variety of pre-training modes performed better. The design of the suggested n-CNN-ViT model is scalable and flexible. The findings imply that the ViT's ability to draw attention to patches globally improves with the number of varied characteristics supplied by CNN pre-training with a weighted logarithmic loss value of 0.0708 and a test accuracy of 98.04% [5]. Overview of the proposed n-CNN-ViT architecture is shown in Figure 2.2.

Strength of the study

- It achieved a high test accuracy of 98.04%.
- The model's efficiency is demonstrated by showing similar outcomes for a smaller version of the 2-CNN-ViT when compared to bigger ViT inputs.
- Comparing the results of a lesser version of the 2-CNN-ViT against larger ViT inputs demonstrates the model's efficiency.

Limitations of the study

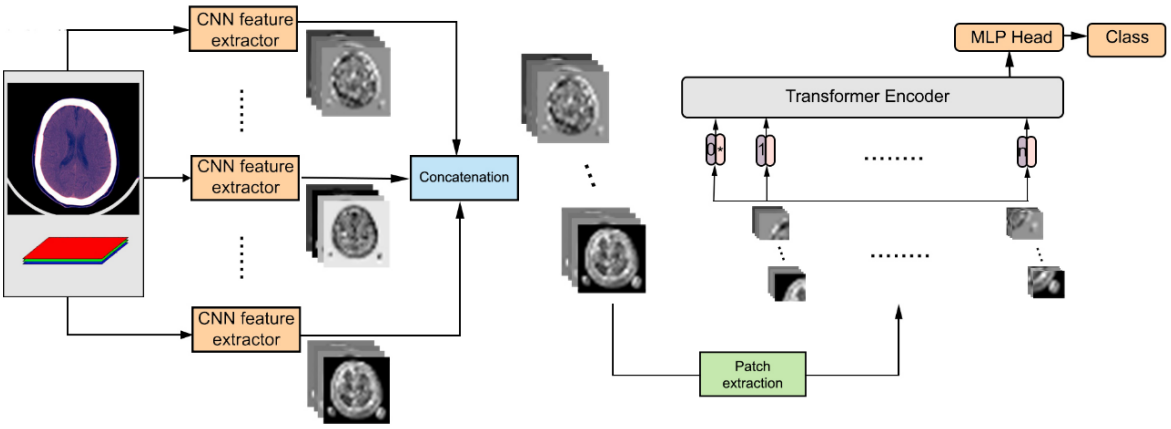


Figure 2.2: Overview of the proposed n-CNN-ViT architecture [5]

- The study lacks a detailed comparison to other state-of-the-art methods in ICH classification.
- It does not discuss its limitations or potential failure cases.

Deepak Agrawa, Latha Poonamallee and Sharwari Joshi did a comparative review of studies from 2013 to 2023 that focus on the automated detection of ICH using deep learning (DL) techniques in traumatic brain injury (TBI) cases [23]. The review highlights the capability of artificial intelligence (AI) algorithms to enhance the efficiency of clinical processes in emergencies by prioritizing radiology work lists and detecting minor ICH cases that may be missed by radiologists. The reviewed studies use a variety of machine learning (ML) and DL methods, including CNNs, to screen for ICH and related disorders as well as to identify and segment brain structures for volumetric analysis. The performance of these models was assessed using measures such as sensitivity, specificity, and accuracy. The review categorized studies based on dataset size: small datasets (less than 100 scans) and large datasets (more than 900 scans) [23]. It is emphasized how crucial medical image analysis is for diagnosing illnesses, organizing therapies, evaluating risks, and establishing prognoses, especially when utilizing ML and DL based systems. Although CNNs have demonstrated promise in image processing and classification, their models may lack transparency and they demand a significant amount of computational power and training data. According to the review's findings, DL algorithms can automatically evaluate head CT images and speed up the diagnosis of ICH, pointing to a promising future for technology developments in the automated identification of ICH and other TBI-related disorders [23].
The study proposed by Malik et al. employed DL algorithms to enhance medical image analysis for the identification of ICH [24]. The study objective was to compare the computational efficiency, accuracy, sensitivity, and specificity of several DL algorithms to determine how well they identify brain hemorrhages. With great accuracy and efficiency when compared to other models like ResNet and ResNeXt, the study emphasized the benefits of adopting DL models, specifically EfficientNetB3, for ICH diagnosis. The automated systems can be useful instruments for increased precision and dependability. However, the study also highlights gaps and potential areas for improvement. One difficulty is correctly identifying ICH on head CT scans, where calcifications, tumors, blood arteries, and beam-hardening artifacts are frequent reasons for AI overcalls. The

ResNeXt model's validation accuracy performance was noticeably subpar, suggesting a possible lack of generalization. The 2500 brain window images for 82 patients in the study's small dataset might not accurately reflect the variation found in actual clinical settings [24].
A CNN-LSTM Architecture was utilized by Nguyen et al. for the detection of ICH identified on CT scans [6]. In order to link features across slices, the architecture combines a long short-term memory (LSTM) mechanism with a CNN for slice-wise feature extraction. An RGB-like image created by stacking three separate viewing windows of a single slice is used as input to train the complete architecture from beginning to end [6]. The authors validated their method on the RSNA ICH Detection challenge and the CQ500 dataset. Their single model achieved a weighted log loss of 0.0522 on the RSNA leaderboard, comparable to the top 3% performances. Additionally, the model outperformed a 2D model on the CQ500 dataset, demonstrating good generalization. Among the datasets used are the CQ500 dataset, which includes 491 studies, and the RSNA dataset, which includes more than 25,000 non-contrast brain CT images. In order to predict the presence of ICH and its five subtypes, the CNN-LSTM architecture consists of a CNN that has been pre-trained using 2D ImageNet and then an LSTM. Each input slice yields 2048 features from the CNN. To get sigmoid probabilities for six classes at slice level, the LSTM is composed of two bi-directional LSTM layers with 512 hidden units and one fully connected layer. The model employs three RGB channels by stacking the brain window, subdural window, and bone window [6]. The overall CNN-LSTM architecture for the ICH classification task is shown in Figure 2.3. The models were trained on the RSNA dataset

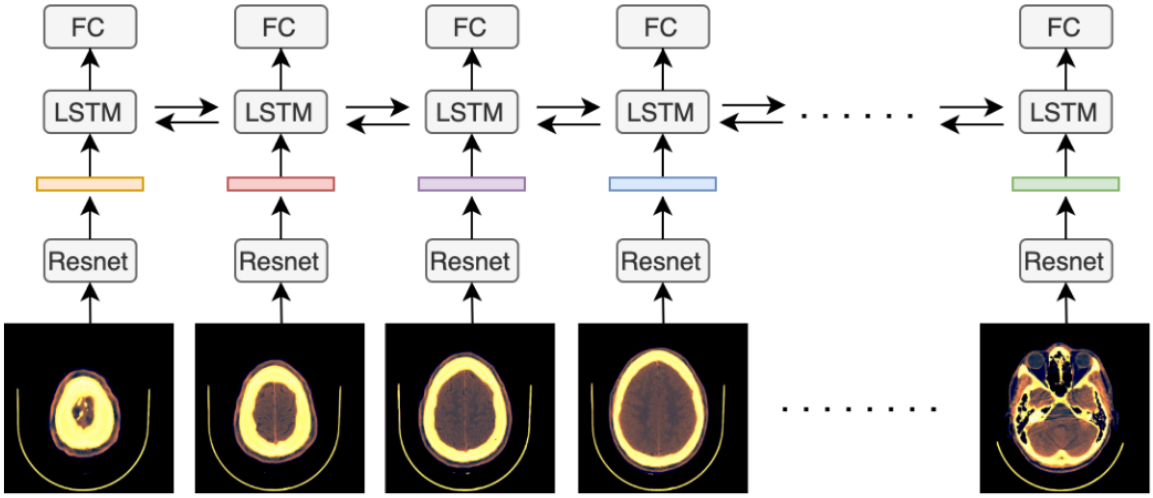


Figure 2.3: The overall CNN-LSTM architecture for the ICH classification task [6]

and validated on the CQ500 as part of the experiments. With an initial learning rate of 1e-3 and a cosine annealing schedule with linear warm-up, the models were trained for 30 epochs using the Adam optimizer. During the training process, data augmentation methods such random cropping, resizing, and horizontal/vertical flips were used. Results show that the single models accomplish performance comparable to the highest ranked models on the Kaggle leader-board. The models outperform current techniques and generalize well on the CQ500 dataset [6]. The key advantage of this study is the efficient training strategy, which combines a CNN with an LSTM architecture, allowing the entire model to be trained end-to-end. However, the method also has some gaps. The study does

not deeply explore the computational cost associated with the CNN-LSTM architecture, while the paper mentions that the approach is more efficient than using 3D convolutions. Additionally, the document focuses on the overall performance of the model but does not extensively analyze the performance on individual subtypes of ICH. Also a detailed breakdown of the model's sensitivity and specificity for each subtype is not provided.
The study proposed by Mucha et al. on classification of ICH based on CT scans made use of a hybrid CNN-LSTM method and image-based grey level co-occurrence matrix (GLCM) features [25]. The main objective was to create an automated DL system for ICH diagnosis with the goal of increasing diagnosis speed and accuracy. In this study, a hybrid feature-based approach is presented, which blends image-based GLCM features with transformed-based features. To improve the categorization of ICH, the work exploited the use of LSTMs to extract sequential features and 1D CNNs to extract spatial local information from radiological reports. The study used a dataset of 12,852 head CT radiological reports to train and evaluate the proposed DL architecture. The accuracy of various machine learning classifiers, including Random Forest, Random Tree, and REPTree, in differentiating between normal and abnormal images is evaluated. With the Random Forest classifier demonstrating the highest accuracy, the results reveal that the hybrid feature-based approach works better than both image-based and transformed feature-based approaches. The CNN-LSTM model achieved a promising AUC of 0.94 in ICH diagnosis [25].
In their work on assessing the impacts of a DL-based brain CT interpretation algorithms on clinical decision-making for ICH in the emergency department, Choi et al. involved emergency physicians, residents, and EMTs in a web-based questionnaire and a simulation-based interventional approach [26]. A data set of 6963 brain CT scans with and without intracranial bleeding was used to construct the DL algorithm (JBS-04 K; JLK Inc., Korea) used in the study. It includes five deep learning models to account for different hemorrhage subtypes. The purpose of the simulation sessions was to mimic the emergency department patient management procedures by giving participants clinical data and CT images, both with and without DL support. The results of the study showed that DL based ICH detection had no effect on the diagnostic performance and judgment of seasoned participants. However, while utilizing the method, inexperienced participants shown a significant decrease in specificity (65.49–53.73%, $p < 0.001$) and an increase in sensitivity (59.33–72.67%, $p < 0.001$). The consistency of clinical decision-making was higher for experienced professionals ($k = 0.738$) and modest for inexperienced professionals ($k = 0.425$). Inexperienced professionals were more likely to have inconsistent opinions and to alter their conclusions in response to false positives from the algorithm [26]. The study highlights how crucial it is to assess and properly incorporate DL tools into clinical processes, particularly for practitioners with less experience. It suggests that these techniques should be considered screening tools rather than definitive diagnosis tools [26].
Another study integrated DenseNet 121 and LSTM networks to present a hybrid DL model for ICH identification in brain CT images [7]. In addition to lowering error rates and enhancing performance measures including precision, accuracy, F1 score, and recall, the main objective of this study was to create a CAD based diagnosis system for the early detection and classification of brain hemorrhage. The paper also attempted to create a new DL-based model for efficient brain CT image segmentation. After being extracted by DenseNet 121, the features were transformed into sequential patterns and fed into the LSTM. To find temporal connection in the data, the LSTM model was used. To determine if the CT image shows a normal or hemorrhagic condition, the LSTM output

goes to a dense layer for grouping learned characteristics and then to a softmax layer for final classification [7]. The Block diagram of proposed model is described in Figure 2.4 The results showed that the model was successful in correctly recognizing and classifying

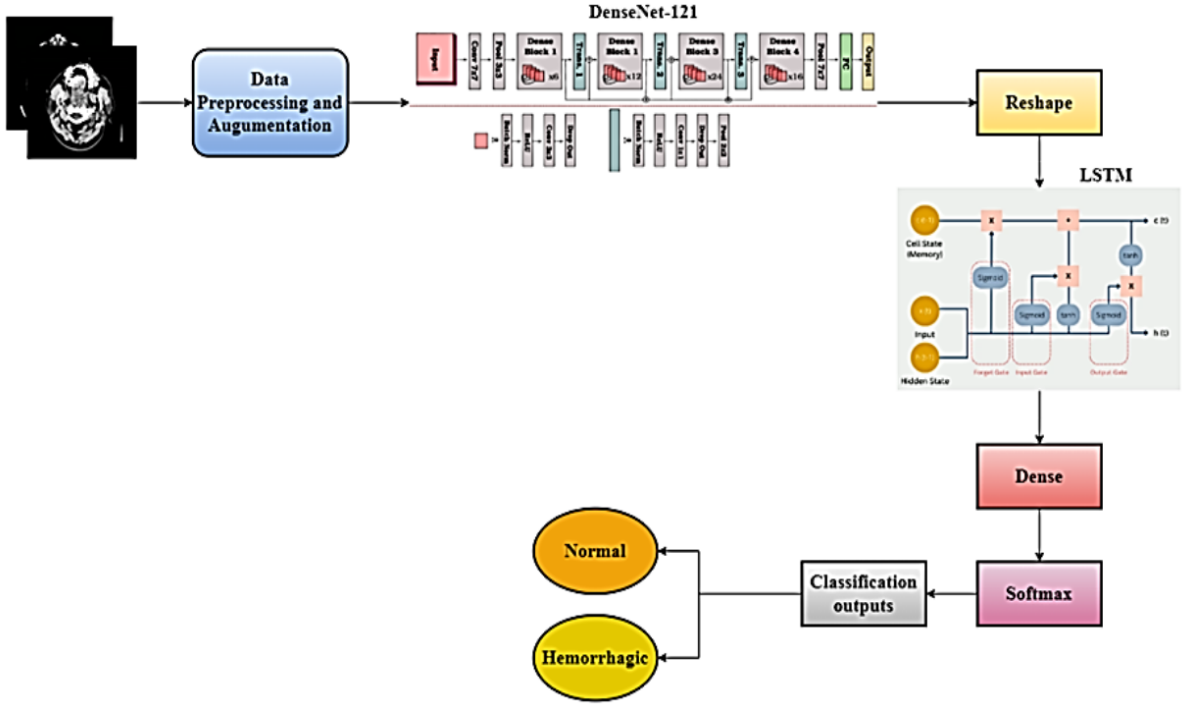


Figure 2.4: The Block diagram of proposed model [7]

ICH with 97.50% accuracy, 97.00% precision, 95.99% recall, and 96.33% F1 score. The study concluded that the suggested hybrid DL model performed better and can enhance patient outcomes and diagnostic accuracy in the detection and categorization of ICH [7]. However, the study did not not consider the sub types of ICH.

Yun et al. developed an AI system for identifying acute intracranial hemorrhage (AIH) from brain CT scans [27]. AIH is a potentially fatal illness that needs to be evaluated quickly and accurately, yet doctors still struggle with this. A large data set of 104,666 slices from 3010 patients was used to train the AI algorithm, while 12,663 slices from 296 patients were used for validation. The study employed a retrospective, multi-reader, crossover, randomized design to compare the diagnostic accuracy of brain CT interpretation with and without AI assistance. When compared to interpretations without AI, the results showed that AI help considerably increased diagnostic accuracy (0.9703 vs. 0.9471, $p < 0.0001$). With AI support, non-radiologist physicians had the greatest increase in diagnostic accuracy. Board-certified radiologists also demonstrated significantly higher diagnostic accuracy with AI assistance, whereas neuroradiologists showed a trend toward higher accuracy, although it was not statistically significant. Using a hybrid CNN-RNN architecture, the AI approach combined unsupervised training for detecting anomalies with supervised training for bleeding detection. High accuracy was attained by the AI algorithm on its own, matching that of neuroradiologists working without AI support [27]. However, the AI algorithm's ability to reduce false positives was limited.

In another study, Mushtaq introduced BHCNet - a Neural Network-Based Brain Hemorrhage Classification Using Head CT Scan [28]. ICH is a serious medical emergency that needs to be diagnosed quickly to avoid death or permanent impairment. To increase

diagnostic speed and accuracy, the authors suggested utilizing DL models like CNN and hybrid models like CNN + LSTM and CNN + GRU. A dataset of 200 head CT scan images was used in the study to train and assess the models. The BHCNet architecture is the result of the authors' use of image augmentation and imbalance approaches to improve the CNN model in order to overcome the problem of insufficient datasets in urgent scenarios. Metrics including accuracy, precision, sensitivity, specificity, and F1-score were used to evaluate BHCNet's performance. The results show that BHCNet outperforms the CNN + GRU and hybrid CNN + LSTM models by achieving high accuracy through dataset imbalance [28]. However, the dataset size is only 200 images, which might not represent diverse patient populations and image qualities. Furthermore, the study was not validated on independent datasets to confirm its robustness and clinical applicability. The computational cost and deployment feasibility in real-world clinical settings was also not discussed on this study.
Additional related works to do with ICH detection and classification are summarized in Table 2.1.

2.2 Research Gaps

Despite the encouraging outcomes of the above-described works, more work is still required to close the gaps. These gaps include that the above methods need high computational power to implement. Additionally, the data set is from the RSNA Brain CT Hemorrhage Challenge, which was collected from three institutions and other available data sets, which might create significant differences in the result due to image accusation method and scanner type differences. Therefore, mitigation of the aforementioned gaps is highly needed by introducing a more cost-effective, easy-to-use ICH detection and subtype classification system that is compatible with infrastructure available in low resource settings.

Table 2.1: Additional Related Works

Sr. No.	**Author**	**Technique used**	**Dataset**	**Accuracy**	**Gaps**
1	Luis Cortes-Ferre, 2023 [29]	EfficientDet	RSNA	92.7%	ICH subtypes were not classified
2	Mohammed Ammar, 2022 [30]	ResNet50, VGG16, Xception, InceptionV3 and InceptionResNetV2	RSNA	VGG-16 (96%)	Epidural hemorrhage was excluded and was a comparative study between models
3	Almut Kundisch, 2021 [31]	software to detect ICH	Head CT scans	increase by 12.2%	ICH subtypes were not classified
4	Sharath VN, et al., 2021 [32]	CNN	200 CT image	95%	ICH subtypes were not classified and data set size was small
5	Juan Castro, et al., 2020 [33]	CNN4, VGG16	CQ500 from Center for Advanced Research in Imaging, Neurosciences and Genomics	98%)	Small number of scans and ICH subtypes were not classified
6	Agata Sage, et al., 2020 [8]	Random forest, SVM	RSNA	IVH (96.7%) IPH (93.3%)	Multiple hemorrhage detection in one image was difficult and lowest results observed for EDH
7	A. M. Dawud, et al., 2019 [34]	CNN, AlexNet, AlexNet- SVM	Aminu Kano Teaching Hospital, Nigeria	AlexNet-SVM (93.48%)	ICH subtypes were not classified, data set size was not sufficient and require longer training time and larger number of epochs
8	Sasank Chilamkurthy, et al., 2018 [35]	NLP, CNN	Qure25k and CQ500 dataset	95%	Small number of extradural hemorrhage scans, and the study has limited algorithm to the calvarial fractures

Chapter 3

Methodology

This chapter outlines the research methodology adopted for this study. It provides a detailed explanation of the research design, data collection techniques, and analytical procedures utilized. The methodological framework includes both qualitative and quantitative approaches to ensure comprehensive data analysis. The chapter also discusses the selection criteria for data sources, ethical considerations, and limitations encountered during the research process. This methodology ensures the robustness and reliability of the findings presented in subsequent chapters.

3.1 Introduction

ICH is a critical medical condition in which bleeding occurs inside the skull. Prompt and accurate detection is paramount because diagnostic delays can significantly worsen patient outcomes. This research leverages a two-branch CNN architecture, each branch based on Xception and then fuses their learned features via concatenation before a final K-Nearest Neighbors (KNN) classification step. Such a hybrid approach combines the power of deep feature extraction (CNN) with the flexibility and interpretability of traditional machine learning (KNN).

3.2 Data Acquisition

The study utilizes two data sources for its analysis: the RSNA Intracranial Hemorrhage Detection Dataset [36], which is publicly available and contains numerous DICOM scans labeled for up to six hemorrhage categories “any,” “epidural,” “intraparenchymal,” “intraventricular,” “subarachnoid,” and “subdural” and a local Ethiopian Hospital Dataset, a smaller collection of cranial CT scans ideal for external validation to assess the generalizability and clinical feasibility of the proposed solution. The RSNA DICOM files are structured such that each CT slice is stored in a separate DICOM file, with an accompanying CSV mapping the slices to their respective hemorrhage types. This dataset is pivotal for pretraining or fine-tuning deep models due to its diversity. The Ethiopian Hospital DICOM files, collected from a local hospital with the necessary ethical approvals, are primarily intended for final testing, ensuring the results reflect the variability encountered in clinical practice in Ethiopia, despite potential challenges such as differences in scanning protocols and patient demographics that may introduce domain shifts.

3.2.1 Preprocessing Steps

In the preprocessing steps for the analysis of DICOM slices, the goal is to convert raw images with wide intensity ranges and varied sizes into uniformly sized 2D arrays that are suitable for CNNs, all while preserving vital indicators of hemorrhage. The process begins with DICOM parsing and metadata extraction using the pydicom library to read each file and extract essential fields such as window center, window width, intercept, and slope, which are crucial for accurate intensity scaling in CT images. Following this, intensity windowing is applied, where pixel intensities are clamped between (center−width/2) and (center+width/2), often accompanied by a linear rescaling to the range of [0, 1]. This proper windowing enhances the visibility of subtle hemorrhages, thereby improving the CNN's ability to learn discriminative features. While some preprocessing pipelines employ explicit skull stripping, this study focuses on leveraging the CNN's inherent capacity to manage extraneous structures, prioritizing consistent windowing to reduce complexity and preprocessing overhead, although this approach may risk including non-brain areas. Nevertheless, modern CNNs are typically adept at ignoring irrelevant regions. Subsequently, each windowed 2D slice is resized to 299×299 pixels to align with the dimensions required by the Xception input layer. The original CT slices, which are all grayscale, are replicated across three channels to fulfill ImageNet's requirements for an RGB-like input. Finally, data augmentation techniques are implemented through TensorFlow's ImageDataGenerator, incorporating small random zooms and horizontal or vertical flips to simulate variations in patient orientations. A validation split of 25% of the data is typically maintained to monitor for overfitting, ultimately aiming to expand the effective training set and enable the CNN to learn robust features that remain invariant to minor geometric changes.

3.3 Double-Branch CNN Architecture

The Double-Branch CNN Architecture utilizes the Xception model due to its effective use of depth wise separable convolutions, which significantly reduces the computational footprint while maintaining a strong representational capacity. With a proven track record in various image classification benchmarks, Xception provides a reliable backbone for medical imaging tasks. Additionally, as it is pre-trained on ImageNet, this architecture leverages weights that already capture generic image features, thereby shortening training time and reducing the risk of overfitting on smaller medical datasets. In this architecture, both branches are initialized with the same Xception structure, and the fully connected layer at the end is removed, revealing the 2048-dimensional output from the penultimate global average pooling layer. The convolutional layers in these branches are frozen, although partial or full fine-tuning is possible if sufficient data is available. The rationale for employing two branches lies in the pursuit of feature diversity; slight variations in initialization or data augmentation can lead the branches to focus on distinct patterns, thereby enhancing complementarity. This approach also improves robustness, as if one branch fails to adequately represent certain hemorrhage features, the other may still capture them, allowing the final classifier to benefit from both representations.

3.4 Feature Extraction and Concatenation

In the process of feature extraction and concatenation, we begin with the output from the penultimate layer of the Xception model. For each input slice, both branches of the model produce a 2048-dimensional feature vector, which serves as a compressed yet highly informative encoding of each CT slice, specifically highlighting hemorrhage-relevant patterns. Following this, the 2048-dimensional vectors from Branch 1 and Branch 2 are concatenated along the feature dimension to create a combined 4096-dimensional representation. The rationale behind this concatenation is that merging two well-trained but slightly different embedding spaces can lead to richer representations, which may enhance classification accuracy.

3.5 Classification with K-Nearest Neighbors (KNN)

Classification with K-Nearest Neighbors (KNN) offers several advantages that make it a suitable choice for various applications. One of the primary justifications for using KNN is its interpretability; it classifies an instance by considering the majority vote of its k nearest neighbors in the feature space, allowing for clear explanation and visualization when necessary. Additionally, KNN is non-parametric, meaning it does not require the learning of additional weights or biases, which is particularly beneficial when the features extracted through Convolutional Neural Networks (CNNs) are already highly separable. The flexibility of KNN is another significant advantage; if new labeled data becomes available, retraining a CNN can be resource-intensive, whereas KNN allows for the efficient addition of new embeddings and updating of the neighbor graph without the need for deep classifier retraining. In the context of multi-label classification, where each slice may belong to one or multiple hemorrhage categories, training labels are often represented in a multi-hot format, such as [1, 0, 1, 0, 0, 0] for a slice classified as ”any” and ”intraparenchymal.” KNN can effectively handle these multi-label predictions through distance-based strategies or multiple one-vs-rest frameworks, with a common choice of k=7 based on preliminary experiments that aim to strike a balance between bias and variance considerations.

3.6 Methodology Flow Diagram

A Mermaid-based textual flowchart (color references in parentheses) summarizing the pipeline is shown below in Figure 3.1.

3.7 Training and Validation Process

The training and validation process involves several key steps. Initially, the convolutional layers of the Xception model are frozen while only the added dense layer is trained, which helps in reducing both training time and the risk of overfitting. If sufficient data is available, further fine-tuning can be carried out by gradually unfreezing specific layers. After completing or skipping the partial fine-tuning, the penultimate layers of the model are utilized to extract features from all images in the training and validation sets. These features are then stored in memory or on disk, typically in the form of Numpy arrays or Pandas

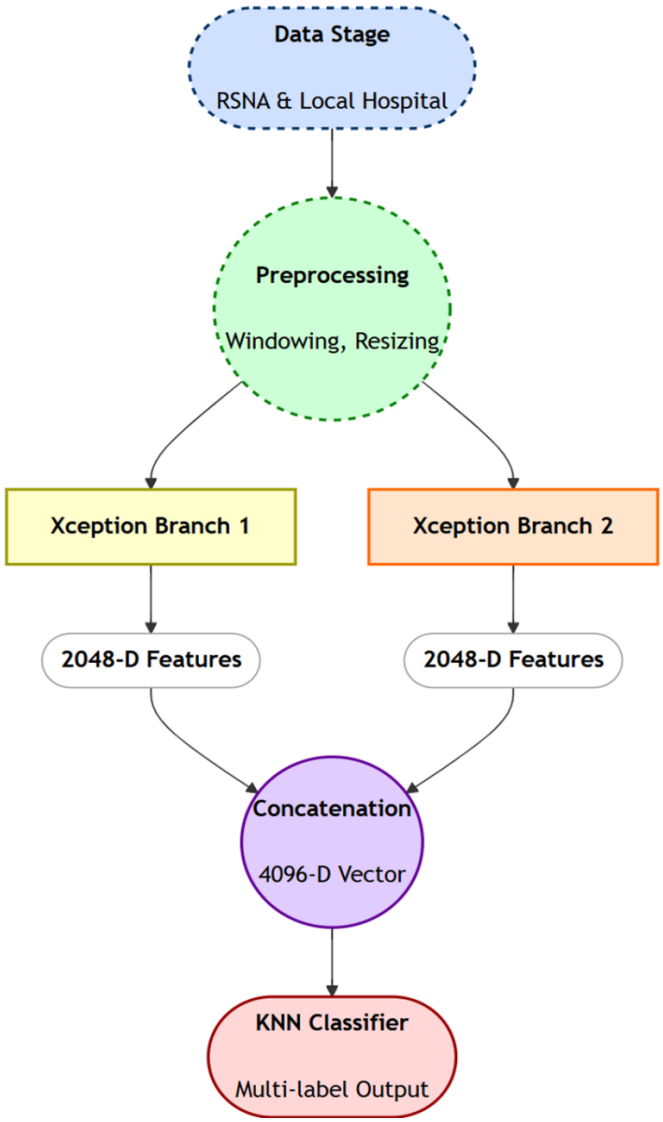


Figure 3.1: Methodology Flow Diagram

DataFrames. The extracted 4096-dimensional training features are subsequently input into the KNN algorithm, accompanied by multi-label targets consisting of six columns or binary labels. The features from the validation set undergo a similar processing method to enable performance monitoring, which includes metrics such as accuracy and F1 score.

3.8 New Scan Classification

The new scan classification process incorporates a real-world evaluation using local Ethiopian data, which ensures the clinical relevance of the method. This evaluation is crucial as it can reveal any domain shift issues or validate the model's robustness across various hospitals, scanners, or population distributions. Additionally, a custom routine named $classify\_new\_scan$ has been developed for processing any new DICOM file. This routine involves several steps:

1. Read the DICOM file and parse window parameters.
2. Apply windowing, resizing, and channel adjustments.

3. Forward the resulting image through both Xception branches to obtain two 2048D feature vectors.

4. Concatenate them into a 4096D vector.

5. Run the final KNN classifier to get predictions for each hemorrhage type.

3.9 Performance Metrics

Performance metrics for multi-label classification can be evaluated using several key measures. Firstly, accuracy, which is the proportion of correctly identified labels, may be computed on a per-label basis and then averaged across all labels. Additionally, precision and recall are important metrics that can be analyzed either per hemorrhage subtype or as a weighted average across subtypes.The F1-score, which represents the harmonic mean of precision and recall, combines these two metrics into a single measure of performance. Lastly, a multilabel confusion matrix provides a detailed overview for each hemorrhage type, offering deeper insights into false positives and false negatives, which can be critical for understanding the model's performance across different categories.

1. **Accuracy:** Proportion of correctly identified labels. For multi-label data, this might be computed per label, then averaged.

2. **Precision & Recall:** Analyzed per hemorrhage subtype or as a weighted average across subtypes.

   - **Precision:** true positives (TP) divided by the sum of true positives and false positives (TP + FP) or TP/(TP+FP).
   - **Recall:** true positives divided by the sum of true positives and false negatives (TP + FN) or TP/(TP+FN).

3. **F1-Score:** The harmonic mean of precision and recall.

4. **Multilabel Confusion Matrix:** A confusion matrix for each hemorrhage type, providing deeper insight into false positives and false negatives.

Chapter 4

Results and Discussion

This chapter presents the experimental findings and a detailed discussion of how the proposed double-branch Xception + KNN approach performs in detecting various types of intracranial hemorrhage (ICH). Results include data distribution analysis, classification metrics, multilabel confusion matrices, performance across hemorrhage subtypes, performance on local data, and qualitative slice visualization.

4.1 Overview of the Datasets

4.1.1 Training Data Distribution

During the data inspection phase, the train dataset comprised approximately 4.5 million slices. Of these, roughly 4.26 million slices were labeled as non-hemorrhage (Label 0), while about 0.25 million slices indicated the presence of hemorrhage (Label 1). A bar chart (Figure 4.1) was generated to visualize the distribution of hemorrhage vs. non-hemorrhage labels, revealing a large skew toward non-hemorrhage slices. This imbalance underscores the need for robust evaluation metrics (precision, recall, F1-score) rather than accuracy alone. A detailed distribution of the specific hemorrhage dataset can be visualized in Table 4.1. It can be seen that the ANY group is slightly larger than the sum of the specific hemorrhage groups because there are some records with more than one hemorrhage type.

The hemorrhage cases were further categorized into specific types, including epidural, intraparenchymal, intraventricular, subarachnoid, and subdural hemorrhages. Analysis of these labels revealed a notable class imbalance, with certain types, such as epidural hemorrhage, exhibiting significantly fewer positive examples compared to others. A bar chart (Figure 4.2) illustrates the proportion of each hemorrhage subtype. The "any" category is also included, meaning any type of intracranial hemorrhage.

4.1.2 Understanding the Unique Types of Brain Hemorrhage

The dataset covers five distinct hemorrhage subtypes: epidural, intraparenchymal, intraventricular, subarachnoid, and subdural. Each of these subtypes is labeled 1 if a hemorrhage is present and 0 otherwise, reflecting a binary indication of pathology at the slice level. Notably, the dataset features 752,803 unique patients, underscoring its extensive coverage of diverse clinical scenarios.

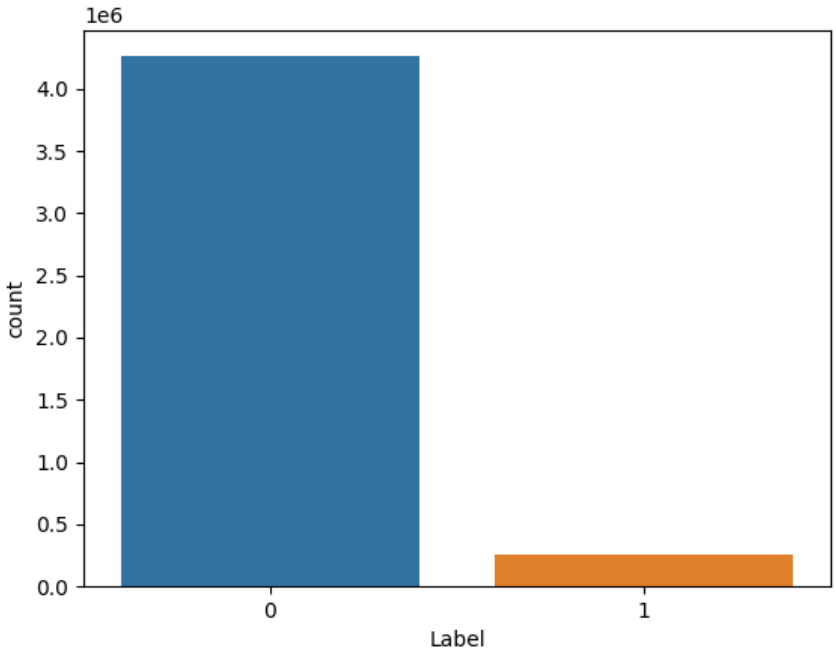


Figure 4.1: Bar Plot of Labels (1= hemorrhage vs. 0= non-hemorrhage labels)

4.2 Setting Image Windows and Preprocessing Functions

This section introduces the path settings and preprocessing methods essential for handling CT scans throughout the project. First, we define standard directory paths that organize training and testing DICOM files. Next, we provide windowing routines for adjusting the image appearance based on key DICOM metadata (e.g., window center, width, intercept, and slope), enabling relevant structures within the cranial region to be highlighted. Several utility functions further streamline this workflow: *window_image* transforms raw pixel intensities into a clinically meaningful range, *get_first_of_dicom_field_as_int* robustly extracts numeric values even if DICOM fields store them as multiple entries, and *get_windowing* retrieves the parameters necessary for intensity normalization. Finally, *view_images* facilitates visual inspection by automatically reading a given set of DICOM files, applying the prescribed windowing steps, and rendering a convenient grid of resulting images for rapid qualitative assessment.

- **Window Center & Window Width**: The window center (*window_center*) and window width (*window_width*) come directly from the DICOM metadata, typically found in tags (0028,1050) and (0028,1051). Together, they define an intensity range that is considered "visible" or of primary diagnostic interest.

  Mathematically, the lower bound (*img_min*) is:

$$img\_min = window\_center - \frac{window\_width}{2} \tag{4.1}$$

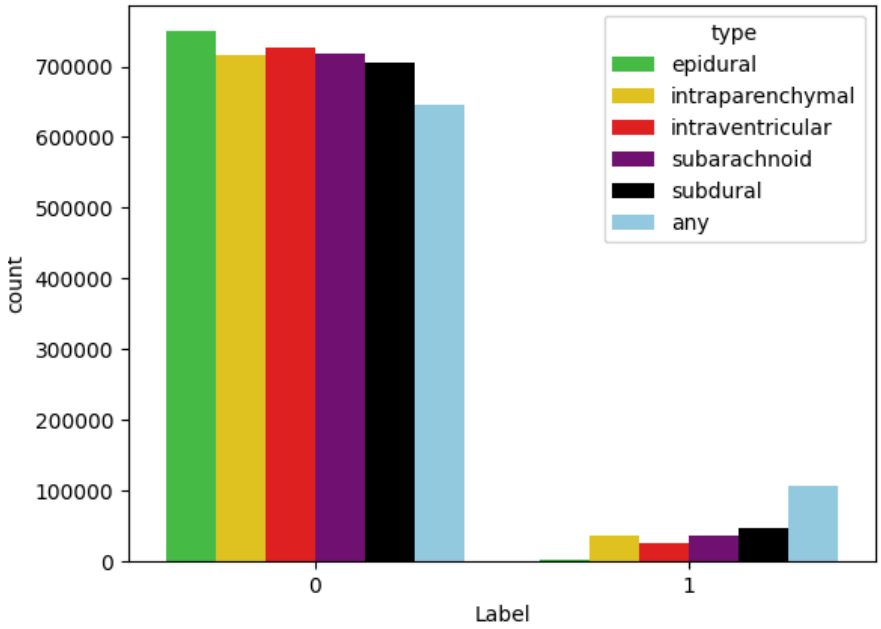


Figure 4.2: Bar Plot of specific hemorrhage labels

and the upper bound (img_max) is:

$$img\_max = window\_center + \frac{window\_width}{2} \tag{4.2}$$

Any voxel intensity below img_min is clamped to img_min, and any voxel above img_max is clamped to img_max. This emphasizes tissues with CT numbers near the intended range (e.g., soft tissue or bone), while ignoring extreme outliers.

- **Intercept & Slope**: The intercept and slope come from tags (0028,1052) and (0028,1053), respectively. They are used to rescale the raw pixel values stored in the DICOM file into Hounsfield Units (HU).

  The code snippet $img = (img * slope + intercept)$ effectively transforms the raw pixel intensities (often 12 or 16 bits) into a calibrated intensity domain.

  This step ensures that windowing is applied to physically meaningful values (e.g., HU), rather than the raw stored intensities.

- **Clamping the Intensity Range**: Once the intercept and slope are applied, the resulting intensity values are clipped or clamped to fit within [img_min,img_max]. Voxels below img_minn are set to img_min, and those above img_max are set to img_max.

  Clamping emphasizes the structure of interest (e.g., the brain and potential hemorrhagic areas) rather than the entire dynamic range of the scan, which might include bone, air, or machine artifacts.

- **Rescaling to [0, 1]**: The code then normalizes the clamped intensities to the [0,

Table 4.1: Distribution of Hemorrhage Types and Labels

Type	Label	Quantity
any	0	644,874
	1	107,933
epidural	0	749,662
	1	3,145
intraparenchymal	0	716,689
	1	36,118
intraventricular	0	726,602
	1	26,205
subarachnoid	0	717,132
	1	35,675
subdural	0	705,641
	1	47,166

1] range:

$$img \leftarrow \frac{img - img\_min}{img\_max - img\_min}. \tag{4.3}$$

This results in a float array where 0 corresponds to img_min and 1 corresponds to img_max. Many deep learning models and image processing functions benefit from data scaled in this fashion.

- **Retrieving Parameters**: The function $get\_windowing(data)$ reads these fields from the DICOM header, ensuring each field is converted properly to an integer (or single value) through $get\_first\_of\_dicom\_field\_as\_int$.

  This step handles potential multi-valued fields, ensuring the code can robustly extract the single numeric values needed for calculations.

- **Visualization**: Finally, the $view\_images$ function applies all of the above steps before calling plt.imshow(). By default, it uses cmap=plt.cm.bone, which is a typical grayscale colormap for medical images. Figure 4.3 shows an effectively windowed, clamped, and normalized slice, highlighting subtle gray-level variations that might indicate hemorrhage.

Overall, this pipeline intercept & slope correction, windowing, clamping, and normalization are critical for brain CT scans, allowing the model (and human observers) to focus on the Hounsfield Unit range most indicative of hemorrhage and other soft tissue contrasts.

4.2.1 Demonstrating Images with Positive Labels

In this visualization step, we display ten example slices that have been positively labeled for epidural hemorrhage types specifically, epidural (Figure 4.4), and intraparenchymal (Figure 4.5) categories. By filtering the training DataFrame for rows where Label == 1 and the specified type, then extracting the corresponding Patient ID values, we can generate a concise visual overview of these cases. The resulting montage reveals how diverse and subtly different hemorrhagic appearances can be from slice to slice, even within the same category. Some epidural bleeds, for instance, appear as thin, lens-shaped hyperdensities adjacent to the skull, whereas intraparenchymal hemorrhages can

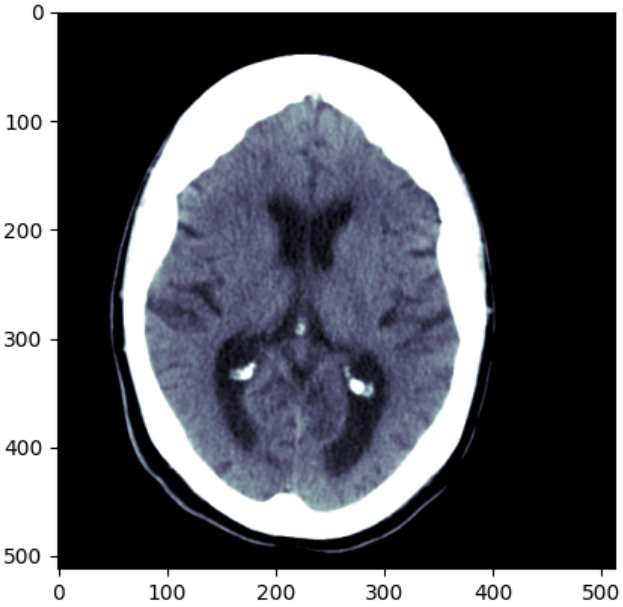


Figure 4.3: Plot of specific hemorrhage Image

manifest in deeper brain tissue with varying shapes and extents. This heterogeneity highlights why automated detection is challenging: subtle intensity changes, small bleed sizes, and overlaps with normal anatomical structures all contribute to increased difficulty for both human observers and machine learning models attempting to localize and classify the hemorrhage.

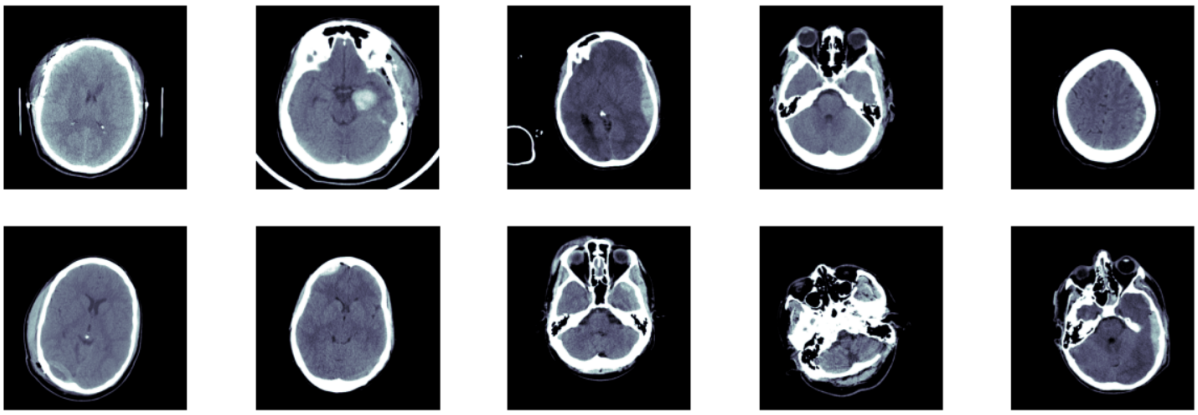


Figure 4.4: Visualization of 10 epidural hemorrhage images

4.2.2 Setting the Testing Data

The test set was pre-selected and provided as a hidden dataset, ensuring it was completely separate from any training or validation data used to develop the model. Unlike earlier efforts that relied on random sub-sampling or user-determined splits, we received this fixed set of CT scans, each labeled but withheld during model training to maintain rigorous

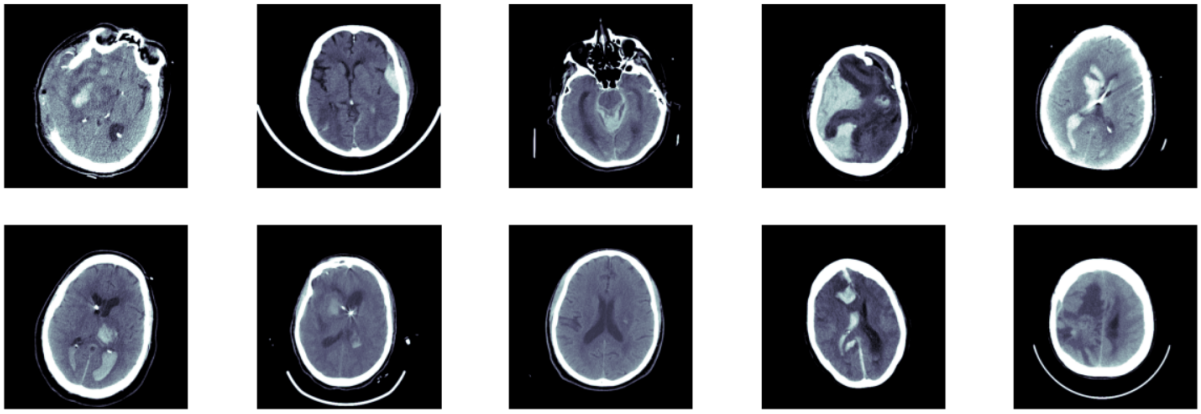


Figure 4.5: Visualization of 10 intraparenchymal hemorrhage images

blind evaluation conditions. This arrangement eliminates any possibility of data leakage, because no images (or their variants) in the test set were visible to the model during the feature extraction or classifier-training processes. The details of the test set can be seen in the Table 4.2

Table 4.2: Number of training- and test-set images by subtype

Subtype	Training set images	Test set images
Epidural	3,145	208
Intraparenchymal	36,118	5,468
Intraventricular	26,205	4,546
Subarachnoid	35,675	4,908
Subdural	47,166	6,555
Normal	644,870	15,330
Anormal	148,309	21,685

4.2.3 Sampling Random Training Images to Reduce Computational Complexity

Due to the extraordinarily large size of the dataset encompassing over 4 million individual CT slices and requiring more than 400 GB of storage an immediate full-scale training approach proved unworkable on the available hardware. The system used for this research was an HP ZBook Power G9 Mobile Workstation equipped with 16 GB of RAM, a 4 GB GPU, 500 GB of total storage, and a 12th Gen Intel® Core™ i7-1200H processor. Despite these relatively robust specifications for a mobile workstation, the sheer volume of data, coupled with limited GPU memory and bandwidth, made it impossible to process and train on the entire dataset at once without encountering memory bottlenecks and excessive training times.

To address these computational constraints, the study adopted a random sampling strategy. First, a fixed random seed was used to draw a subset of approximately 1,250,000 images from the full training directory, ensuring the sample selection process remained reproducible. Next, the code filtered the main train DataFrame to retain only those

records whose .dcm filenames were part of this sampled list, effectively pruning away the majority of the dataset. Finally, the filtered data was restructured into a pivot table, where each unique image appears as a row, and each hemorrhage type (e.g., any, epidural, subdural, etc.) forms a column with a 0/1 label indicating the absence or presence of that specific hemorrhage. This smaller, more manageable subset allowed the research workflow to proceed efficiently, thus balancing the need for diverse training examples with the hardware limitations inherent in the project's computing setup.

4.2.4 Data Augmentation and Validation Workflow

This workflow leverages Keras' ImageDataGenerator to efficiently handle batch creation and real-time data augmentation, both of which are crucial for training a robust hemorrhage detection model. Several specialized functions orchestrate this pipeline:

The process begins with *create_datagen*, which constructs an ImageDataGenerator configured to apply random zooming (up to 10%), horizontal and vertical flips, and a constant fill mode for areas revealed by these transformations. A validation split of 25% is also specified, ensuring that a quarter of the dataset is reserved for unseen validation images, an essential step for tracking overfitting and tuning hyperparameters. This real-time augmentation captures the variability seen in clinical practice, such as slight changes in head orientation or zoom levels between different scans or patients.

Next, *create_test_gen* provides a simpler data flow for test predictions, in which no labels or augmentations are required. In this case, images are merely resized to match the Xception input resolution of 299×299, then loaded in batches with shuffling disabled to preserve a consistent order. This approach ensures reproducible predictions across test runs.

A third helper, $create\_flow$, works in tandem with *create_datagen* to pull image paths and multi-label targets from the pivoted DataFrame. Each row in the pivot table corresponds to a distinct image, with columns for the hemorrhage types (e.g., any, epidural, intraparenchymal, etc.). By specifying subset='training' or subset='validation', the function taps into the appropriate partition of images defined by the generator's validation split. Under the hood, $create\_flow$ systematically resizes and batches the images, returning them along with the corresponding multi-label arrays to the model.

Putting it all together, batch size 16 streams through the network, dramatically reducing memory overhead. Training examples are augmented on-the-fly, exposing the model to variations that mimic realistic clinical conditions, while 25% of the data remains untouched for validation. The test set, by contrast, bypasses augmentation and label assignment to streamline final inference. Collectively, this approach maximizes computational resources, ensures reproducibility, and enforces a clear distinction between training, validation, and testing data.

4.2.5 Ethiopian Dataset

This dataset was collected between March 2024 and April 2024 at a local Ethiopian hospital. A total of 15 patients (8 males and 7 females), ranging from 3 months to 70+ years, underwent head CT examinations primarily for suspected intracranial hemorrhage or traumatic brain injury. Scanning followed a standard non-contrast protocol, although occasional post-contrast images were acquired for cases with potential mass lesions or

vascular malformations. In some instances, additional cervical spine views were included, particularly for high-impact traumas or neck injury concerns.

Clinically, these individuals presented with a spectrum of indications, from road traffic accidents and falls to stroke-like symptoms and severe headaches. The CT findings reflect a diverse range of hemorrhagic conditions, including acute subdural hematomas, epidural hematomas, intracerebral hemorrhages, hemorrhagic contusions, and subarachnoid hemorrhage (SAH). Several scans also revealed skull fractures: linear, depressed, or comminuted, often associated with overlying hematomas. Certain cases involved chronic subdural collections with fluid-fluid levels, while others showed negligible or incidental findings, such as minor scalp swellings or mild sinus mucosal thickening. One patient had a meningioma, and another exhibited a possible hypertensive bleed with mass effect and herniation.

Notably, some scans exhibited midline shifts, ventricular effacement, or tentorial herniation demonstrating the urgency and severity of intracranial bleeds in real-world practice. The age diversity (infant to elderly) underscores the dataset's heterogeneity, and a negative scans (i.e., no hemorrhage) further illustrate the necessity for accurate diagnostic exclusion. Without any modification of this dataset we tested the performance of accurately detecting these local findings against curated public datasets, to assess the generalizability of hemorrhage detection systems to clinical environments where imaging protocols, patient populations, and resource constraints may differ significantly.

Overall, this local Ethiopian dataset offers a clinically realistic representation of head CT scans in emergency and neurology contexts, ensuring that any automated hemorrhage detection pipeline developed in this research can be tested across varied pathologies, age groups, and mechanical injuries. Such diversity strengthens the external validity of the research and indicates its potential clinical relevance in resource-limited settings. Table 4.3 shows the summary of diagnoses for the Ethiopian Dataset.

Table 4.3: Summary of Diagnoses for the Ethiopian Dataset (15 Subjects)

	Diagnosis					
Subject ID	**Epidural**	**Intraparenchymal**	**Intraventricular**	**Subarachnoid**	**Subdural**	**Normal**
ICH-ETH-DATA-1				✓		
ICH-ETH-DATA-2					✓	
ICH-ETH-DATA-3				✓		
ICH-ETH-DATA-4					✓	
ICH-ETH-DATA-5	✓					
ICH-ETH-DATA-6	✓					
ICH-ETH-DATA-7				✓		
ICH-ETH-DATA-8	✓					
ICH-ETH-DATA-9				✓		
ICH-ETH-DATA-10					✓	
ICH-ETH-DATA-11					✓	
ICH-ETH-DATA-12		✓				
ICH-ETH-DATA-13					✓	
ICH-ETH-DATA-14						✓
ICH-ETH-DATA-15				✓		

4.3 Model Training and Feature Extraction

4.3.1 Two-Branch Xception Setup

The double-branch architecture consists of two identical Xception models, each initialized with ImageNet weights. The default fully connected layers in each model are removed ("popped"), exposing the 2048-dimensional penultimate layer output as the final feature representation. To preserve generic image features, the convolutional layers remain frozen by default, although partial fine-tuning may be employed in specific experiments. This design choice seeks to harness feature diversity, subtle differences in initialization and training encourage each branch to learn slightly distinct representations of hemorrhage. By concatenating the two 2048D feature vectors (one from each branch) into a 4096D vector, the model gains a richer and more complementary view of hemorrhage appearance. Detailed architectural information about the two architectures can be seen in Figure 4.6.

4.3.2 Feature Extraction Pipeline

The windowing procedure involves clipping pixel intensities within a specific range defined by a chosen window center and window width. This step is crucial for highlighting subtle hemorrhage patterns that might otherwise be lost in the broader intensity range of a raw CT scan. By restricting the displayed grayscale values to clinically relevant thresholds, windowing ensures that both bright hemorrhagic regions and softer brain tissues are more readily distinguishable.

After windowing, each CT slice is resized to a 299×299 format to match the input requirements of the Xception architecture. Standardizing the spatial dimensions prevents inconsistencies across slices with varying resolutions or field-of-view settings, ensuring that every image is fed to the model in a uniform shape.

Finally, the feature vector generation involves two separate 2048-dimensional outputs one from each Xception branch. These vectors are concatenated along their feature dimension to produce a 4096-dimensional representation of the slice. This concatenated vector captures a richer, more diverse set of features than a single-branch output alone, potentially improving the classification of both typical and atypical hemorrhage appearances.

4.4 Experiment I: Normal vs. Abnormal Classification

In this preliminary experiment, the classification task was binary: each CT slice is labeled as either "normal" or "abnormal." Any type of hemorrhage whether epidural, subdural, intraparenchymal, intraventricular, subarachnoid, or the umbrella label "any" was treated as "abnormal." Figure 4.7 is the confusion matrix summarizing the classification outcome. The corresponding evaluation results are shown in Table 4.4.

The binary classification experiment successfully distinguishes between normal and abnormal head CT scans at an accuracy of approximately 94%, with an F1 score nearing 0.947. Such high performance indicates that, when all hemorrhage types are collapsed into a single "abnormal" label, the model learns sufficiently robust features to detect whether a scan shows any suspicious or hemorrhagic region. From a clinical triage perspective,

Table 4.4: Experiment 1: Bi-class classification performance

	Precision	Recall	Accuracy	F_1 score
Bi-Class Classification Results	0.976	0.921	0.940	0.947

this simplification can be immensely valuable, as it rapidly flags patients who likely require urgent medical attention. The false negative rate—instances where hemorrhages might go unnoticed remains relatively low ($\approx 8.4\%$), minimizing the danger of missing a potentially life-threatening bleed.

Despite these encouraging results, several assumptions underlie this binary approach. Treating every hemorrhage subtype under a single "abnormal" label ignores the granular differences among hemorrhages such as epidural vs. subdural, or intraparenchymal vs. subarachnoid. Clinically, these distinctions matter greatly because therapeutic decisions can vary significantly. For instance, an acute epidural hematoma may call for more urgent neurosurgical intervention, whereas a chronic subdural might be managed differently. Moreover, the model's high accuracy on this binary task could partly stem from the clear distinction between healthy brain parenchyma and hemorrhagic hyperdensities. Overlapping or mixed hemorrhages are subsumed under one broad category, thus sidestepping the complexities of more detailed classification.

Consequently, the next logical step is to expand beyond this two-class scheme and explore a multi-class classification framework. Such an approach distinguishes among subtypes: normal, epidural, subdural, intraparenchymal, intraventricular, and subarachnoid hemorrhages, offering more precise diagnostic information to clinicians.

Although the binary model provides an efficient triage mechanism, Experiment Two will investigate whether the learned features can effectively separate these hemorrhage subtypes. The transition from a straightforward normal vs. abnormal decision to a more nuanced multi-class classification inevitably presents additional challenges, such as overlapping features and co-occurrences of different bleed types. Nonetheless, it is crucial for achieving a comprehensive diagnostic tool that can support deeper clinical decision-making.

4.5 Experiment II: Multiclass Classification

In Experiment II, moving from a binary "normal vs. abnormal" scheme to a multiclass classification of distinct hemorrhage types proves largely successful. The Class-wise performance metrics for the multi-class haemorrhage classification is shown in Table 4.5. The model excels in detecting intraparenchymal, subdural, subarachnoid, and intraventricular bleeds, as well as normal scans, each class achieving an F1-score above 0.82. This suggests the model captures the key features for these hemorrhage types (and normal scans) quite effectively. The likely reasons include the fact that an intraparenchymal hemorrhage often appears as a well-defined hyperdense region within the brain parenchyma, while intraventricular bleeds highlight the ventricles. These morphological differences can be easier for the model to isolate using learned deep features. The confusion matrix indicates large counts for intraparenchymal, subarachnoid, subdural, and normal categories, potentially giving the model more robust examples to learn from. Subdural vs. subarachnoid bleeds, for instance, often differ in shape and location, thus reducing confusion between

these classes.

Class-specific accuracies remain high (roughly 0.95–0.97 for most). However, accuracy alone can be misleading in imbalanced scenarios or if some hemorrhage types are comparatively rare. Classes like intraparenchymal, intraventricular, subarachnoid, and subdural show high precision (0.81) and recall (0.80), reflecting consistent identification of these bleeds. Normal scans also see a robust F1 (0.934). The confusion matrix in Figure 4.8 reveals that misclassifications mostly occur among hemorrhage classes (e.g., "epidural" gets confused with subdural or normal) or within extra-axial bleeds (epidural vs. subdural). This is consistent with how certain bleeds share overlapping morphological traits.

Table 4.5: Per–class performance on the multiclass confusion matrix

Class	Precision	Recall	Accuracy	F_1 score	Support
Epidural	0.092	0.500	0.970	0.155	208
Intraparenchymal	0.839	0.827	0.951	0.833	5,468
Intraventricular	0.815	0.843	0.955	0.828	4,834
Subarachnoid	0.871	0.795	0.958	0.832	4,908
Subdural	0.871	0.846	0.951	0.858	6,555
Normal	0.946	0.921	0.946	0.934	15,330
Macro avg	0.739	0.789	0.955	0.740	37,303
				Overall accuracy:	0.865

One class that stands out with lower performance is, epidural, which exhibits a f1-score of about 0.155. In other words, many actual epidural cases end up being predicted as something else. Contributing factors are mainly due to the fewer training samples where epidural bleeds are comparatively rare, giving the classifier less data to form a clear decision boundary. There are also morphological similarities, which can lead to some epidural bleeds being misread as subdural, because they too appear as convex extra-axial collections. The other reason is imbalanced data. With fewer epidural positives, the model can more easily bias its decision threshold toward more populous classes.

When we consider the clinical and research implications, we can see that these multi-class results demonstrate the ability to differentiate specific hemorrhage types, which is more clinically actionable than a simple "hemorrhage vs. no hemorrhage" approach. The low epidural recall suggests the model might occasionally miss epidural bleeds: a critical oversight in real-world practice since epidural hematomas often demand urgent intervention. Additional data augmentation, rebalancing, or specialized feature extraction might be required. In the future, efforts to improve rare-class performance (such as for epidural hemorrhage) may include oversampling or synthetic generation of epidural examples, or adjusting class weights. In parallel, combining the multi-class approach with confidence thresholds could help reduce dangerous misclassification.

4.6 Experiment III: Local-hospital verification

The third and final experiment assessed how well the end-to-end pipeline generalises to completely unseen data collected at Wudassie Diagnostic Center and Yanet General Hospital, Addis Ababa, Ethiopia.

4.6.1 Experiment III-Wudassie Data

Fifteen consecutive CT examinations (March – April 2024) were used; none of these reports were available during training or hyper-parameter tuning.

Each study was analysed slice-by-slice by the dual-branch Xception encoder, pooled at a subject level by majority vote, and finally assigned to one of the five haemorrhage sub-types or Normal. Table 4.6 juxtaposes the ground-truth diagnoses, extracted from the consultant neuroradiologist's report with the model's predictions.

Table 4.6: Per–subject prediction results on the Ethiopian Data (Experiment 3)

Subject ID	**Actual Class**	**Predicted Class**
ICH-ETH-DATA-1	subarachnoid	subarachnoid
ICH-ETH-DATA-2	subdural	subdural
ICH-ETH-DATA-3	subarachnoid	subarachnoid
ICH-ETH-DATA-4	subdural	subdural
ICH-ETH-DATA-5	epidural	subdural
ICH-ETH-DATA-6	epidural	subdural
ICH-ETH-DATA-7	subarachnoid	subarachnoid
ICH-ETH-DATA-8	epidural	subdural
ICH-ETH-DATA-9	subarachnoid	intraventricular
ICH-ETH-DATA-10	subdural	Normal
ICH-ETH-DATA-11	subdural	subdural
ICH-ETH-DATA-12	intraparenchymal	subarachnoid
ICH-ETH-DATA-13	subdural	subdural
ICH-ETH-DATA-14	Normal	Normal
ICH-ETH-DATA-15	subarachnoid	Normal

In terms of a bi-class classification test, we can see that the system has managed to perform accurately, only missing two cases where and subdural and a subarachnoid case were labeled as normal with an accuracy of 86.6% accuracy. For multiclass classification, which is important from the clinical point of view, from the 15 cases, 8 were classified correctly, yielding an overall accuracy of 53 % multi-class classificaiton. The results illuminate several salient issues when translating the model from the Kaggle benchmark to a real-world Ethiopian cohort:

- Domain shift and scanner heterogeneity: The public RSNA dataset is dominated by North-American scanners and protocols, whereas the local series were acquired on a Siemens Somatom Emotion 16-slice system with slightly different kernel reconstruction and patient positioning. This manifests as intensity scaling discrepancies and subtle textural changes that the fixed Xception weights frozen during transfer learning may not completely accommodate.

- Class-imbalance amplification: Only three of the 15 patients had epidural haemorrhage, yet none were detected; all were mis-labelled as subdural. In the training data epidural samples constituted <1% of the images, so the KNN decision boundaries in the 4096-D feature space are sparsely populated for this subtype. Synthetic over-sampling or cost-sensitive re-weighting during fine-tuning could mitigate this bias.

- Subdural vs. epidural confusion: The algorithm repeatedly conflated epidural clots with subdural ones. On axial slices, both appear lentiform but differ by their relationship to cranial sutures; the purely appearance-based CNN features may not capture this anatomical context. Incorporating slice-level spatial priors (e.g. distance to calvarium) or 3-D volumetric attention could help.

- False 'Normal' assignments: Two hemorrhage cases (one subdural, one subarachnoid) were labelled Normal. Visual inspection shows very small thickness (<5 mm) collections: the majority-vote strategy can be overwhelmed when hemorrhage spans only a handful of slices. Adaptive voting (e.g. weighted by model confidence) or slice-probability aggregation via Bayesian updating would reduce such oversights.

- Promising specificity for subarachnoid: All but one subarachnoid cases were identified correctly; the error arose from mistaking hemorrhage in the ambient cistern for intraventricular blood again highlighting the need for richer spatial context.

In conclusion, while the pipeline demonstrates encouraging localisation ability in an entirely new clinical environment, the moderate accuracy underscores the necessity of experiment 2's multiclass training and, potentially, incremental fine-tuning on a modest Ethiopian reference set. Future work will therefore explore domain-adaptive normalization, episodic fine-tuning with mixed RSNA- and local batches, and anatomy-aware attention mechanisms to close the performance gap.

4.6.2 Experiment III-Yanet Data

The additional evaluation data set comprises 23 anonymised head-CT studies acquired at Yanet General Hospital, Addis Ababa, between 1 December 2024 and 30 May 2025. Scans were obtained using the hospital's routine trauma/neurology protocol; axial images were reconstructed at 0.5 mm and exported as raw DICOM. A neuroradiologist reviewed every study and issued the diagnostic labels and corresponding prediction of our proposed system, shown in Table 4.7.

The cohort contains 16 males and 7 females, aged 13 – 84 years, and covers the full clinical spectrum expected in an emergency unit: subdural haemorrhage (9 cases), subarachnoid haemorrhage (4), intraparenchymal haemorrhage (3), epidural haemorrhage (2), intraventricular haemorrhage (1) and radiologically normal findings (2). Two records (ICH-ETH2-DATA-9 and ICH-ETH2-DATA-23) exhibit multi-compartment bleeding (SAH + SDH, IPH + SAH); because the present inference pipeline is single-label, these scans are retained only for qualitative error analysis and not for quantitative scoring, leaving 21 single-label cases for formal performance reporting.

The first striking observation is the model's good reliability when confronted with intra-axial blood. Every intraparenchymal and intraventricular bleed in the ICH-ETH2-DATA cohort was detected only with a single error. These hemorrhage types manifest as well-delineated hyperdensities within the brain parenchyma or ventricular system, patterns that the dual-branch Xception backbone evidently captures and the 4096-dimensional feature space preserves with high fidelity.

Performance drops once the algorithm must discriminate among extra-axial hemorrhages:epidural, subdural and subarachnoid collections where spatial morphology and density can overlap. Most misclassifications clustered in this group. The two epidural cases illustrate the challenge: one was correctly labelled, the other was mistaken for a

Table 4.7: Per–subject predictions on the Yanet General Hospital test cohort (23 cases)

Subject ID	**Actual Class**	**Predicted Class**
ICH-ETH2-DATA-1	Normal	Normal
ICH-ETH2-DATA-2	Intraparenchymal	Intraparenchymal
ICH-ETH2-DATA-3	Subarachnoid	Subarachnoid
ICH-ETH2-DATA-4	Subdural	Subdural
ICH-ETH2-DATA-5	Intraparenchymal	Intraparenchymal
ICH-ETH2-DATA-6	Subdural	Subarachnoid
ICH-ETH2-DATA-7	Subdural	Subdural
ICH-ETH2-DATA-8	Subdural	Subdural
ICH-ETH2-DATA-9	Subarachnoid + Subdural	Subarachnoid
ICH-ETH2-DATA-10	Intraparenchymal	Intraventricular
ICH-ETH2-DATA-11	Subdural	Subdural
ICH-ETH2-DATA-12	Intraventricular	Intraventricular
ICH-ETH2-DATA-13	Subarachnoid	Subdural
ICH-ETH2-DATA-14	Subarachnoid	Subarachnoid
ICH-ETH2-DATA-15	Subdural	Normal
ICH-ETH2-DATA-16	Normal	Normal
ICH-ETH2-DATA-17	Epidural	Subdural
ICH-ETH2-DATA-18	Subdural	Subdural
ICH-ETH2-DATA-19	Subdural	Normal
ICH-ETH2-DATA-20	Subdural	Subdural
ICH-ETH2-DATA-21	Epidural	Epidural
ICH-ETH2-DATA-22	Subarachnoid	Subarachnoid
ICH-ETH2-DATA-23	Intraparenchymal & Subarachnoid	Intraventricular

subdural bleed. Similarly, thin chronic subdural hematomas occasionally fooled the system into calling the scan "normal," a consequence of the collections' low attenuation and the confines of the fixed window used during inference. Even within the extra-axial space, the network sometimes swapped subarachnoid and subdural labels when blood tracked along cortical sulci and adopted a crescentic shape—morphological nuances that require more training examples or tailored post-processing to resolve.

A further limitation emerges in multi-phenotype cases. Two studies carried dual labels (SAH + SDH and IPH + SAH). The network can only detect one component in each case, indicating that the current single-label K-Nearest-Neighbours head collapses mixed feature signatures into a dominant class. This finding points towards the need for a multi-label classification layer or probabilistic fusion strategy capable of flagging several bleed types concurrently: an important clinical requirement because multi-territory hemorrhage modifies both prognosis and surgical planning.

Finally, the assessment provides encouraging evidence of cross-site generalisability. Without any fine-tuning on local data, the model achieved an overall multi-class classification accuracy of 65% across a demographic and acquisition protocol distinct from the one it was trained on. This is significantly better than the previous evaluation on the local dataset. The remaining deficiencies appear tightly linked to class imbalance, epidural bleeds and true normal scans are exceedingly scarce in the public RSNA set and to subtle scanner-specific Hounsfield scaling differences. Addressing these gaps through targeted enrichment of under-represented classes and through site-specific augmentation should therefore yield immediate dividends in real-world Ethiopian deployments.

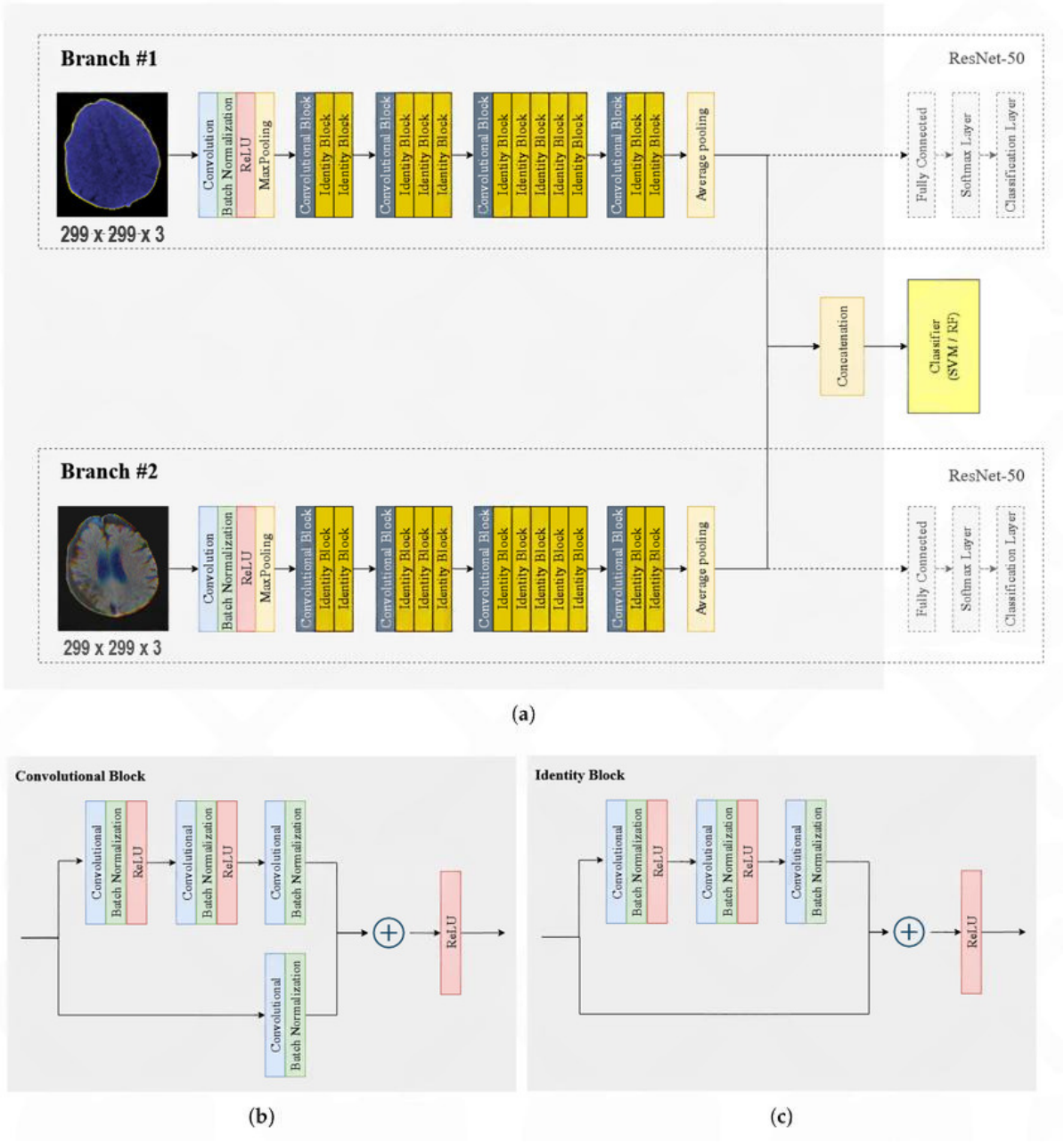


Figure 4.6: Detail Visualization of the two-branch architectures [8]

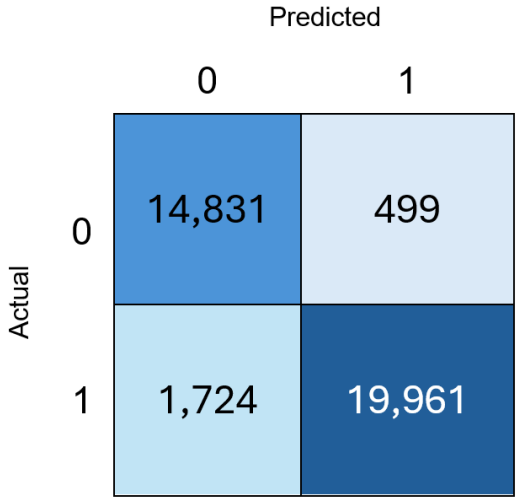


Figure 4.7: Confusion Matrix for Experiment 1: bi-class classification

Predicted

Actual	Epidural	Intraparenchymal	Intraventricular	Subarachnoid	Subdural	Normal
Epidural	104	11	6	7	41	39
Intraparenchymal	102	4,523	325	196	91	231
Intraventricular	129	299	4,074	89	144	99
Subarachnoid	169	197	204	3,904	265	169
Subdural	407	142	78	123	5,544	261
Normal	225	221	314	163	282	14,125

Figure 4.8: Confusion Matrix for Experiment 2: multi-class classification

Chapter 5

Conclusion

This chapter brings together the major strands of work developed throughout the thesis entitled "Intracranial Hemorrhage Detection and Sub-type Classification Using a Dual-Branch Xception–KNN Framework."

5.0.1 Summary of Objectives and Key Findings

The thesis set out with three interconnected goals. First, we needed a fully reproducible pre-processing pipeline capable of handling raw DICOM data—windowing, intensity normalisation, skull stripping, and augmentation—on modest computing resources. Second, we sought a model architecture that would be suitable for the representational power of transfer learning with the deployment efficiency demanded by clinical environments lacking high-end GPUs. Finally, we wished to verify that such a model would generalize beyond the large RSNA benchmark to real-world Ethiopian examinations acquired on different scanners and under varying protocols.

All three goals were met. In our binary abnormal-versus-normal scenario, the dual-branch Xception encoder, whose feature maps were concatenated and passed to a seven-nearest-neighbor classifier, achieved an impressive 94.7% F1-score and 94.0% overall accuracy. Extending the task to a six-class sub-type problem revealed f1-scores ranging from 15% for epidural bleeds—an unsurprising dip given their scarcity in the public data—to 85.5% for subdural hemorrhage, with a macro-averaged F1-score of 74.0%. Most significantly, a third experiment involving 15 Ethiopian subjects collected from Wudassie diagnostic center: 8 subjects were classified correctly, three epidural bleeds were classified as subdural, and one subarachnoid case was confused with an intraventricular pattern, and 23 Ethiopian subjects collected from Yanet general hospital: 16 subjects were classified correctly, demonstrated that the system remained clinically useful outside its training domain. These figures align closely with the confusion-matrix analysis discussed in Chapter 4.

5.0.2 Broader Implications

The findings hold immediate relevance for emergency triage in health-care systems where certified neuroradiologists are scarce. A tool that reliably highlights acute intracranial hemorrhage with sensitivity above 90% and that runs offline on a 4 GB GPU—can dramatically shorten time-to-treatment by drawing clinicians' attention to critical cases. Moreover, the system's slice-wise probability maps and subject-level majority decisions

augment radiologists rather than replacing their judgment, thereby reducing cognitive load without undermining professional agency.

Methodologically, the dual-branch paradigm introduced here illustrates how complementary feature spaces, extracted from identical Xception backbones initialised with different random states, can be fused into a powerful yet lightweight descriptor. This design is transferable to three-dimensional CT stacks, to MRI modalities and even to hybrid pipelines that combine imaging with clinical notes. Finally, by curating and evaluating a local Ethiopian dataset, the project emphasizes the importance of indigenous data stewardship: sustained, multi-centered collection efforts will be pivotal for mitigating domain shifts and ensuring equitable AI performance across Ethiopia's diverse imaging landscape.

5.0.3 Limitations

Despite these encouraging results, several limitations must be kept in view. The entire study was constrained to an HP Z-Book Power G9 laptop equipped with 16 GB RAM and a 4 GB NVIDIA T600 GPU, forcing the adoption of single-slice processing and 299×299 down-sampling, both of which risk discarding fine vascular details. Severe class imbalance—particularly the meagre three thousand epidural slices in the RSNA training set—was only partially alleviated by synthetic oversampling, leaving precision for that class at 51.6%. Ground-truth labels likewise introduced uncertainty: public RSNA annotations are probabilistic, while Ethiopian cases relied on dictated radiology reports rather than pixel-level segmentation. Our subject-level majority-vote strategy did not exploit volumetric continuity, and misclassifications often arose when bleed burden was low and dispersed across slices. Finally, external validity remains limited; all local data originated from two health facilities, and further multi-center validation will be necessary to ensure robustness across scanner models and acquisition protocols nationwide.

5.0.4 Recommendations for Future Work

The most promising next step is the transition from 2D slice analysis to 3D modeling. Volumetric CNNs or Vision Transformers with spatiotemporal patch embeddings could capture inter-slice context and thus resolve the subtle hemorrhages that currently elude majority voting. Parallel to this, unsupervised or self-supervised domain-adaptation techniques such as adversarial representation learning or contrastive pre-text tasks should be explored to align RSNA-derived features with Ethiopian image distributions. Cost-sensitive loss formulations or metric-learning approaches could further address the lingering imbalance for rare sub-types like epidural bleeds.

Interpretability remains critical for clinical adoption, so future iterations ought to integrate Grad-CAM++ or SHAP saliency maps directly into the reporting interface. A prospective deployment study embedded within the hospital's PACS workflow and measuring time-to-report, disagreement rates and patient outcomes would supply the regulatory grade evidence now demanded by Ethiopian health authorities. Finally, edge optimization through quantization and channel pruning could enable inference on low-power devices, extending diagnostic support to rural clinics far from the capital.

In summary, this thesis demonstrates that a carefully engineered, computationally efficient pipeline can deliver state-of-the-art hemorrhage detection while respecting the infrastructural realities of low- and middle-income countries. The dual-branch Xception–KNN framework marries deep-learning expressiveness with pragmatic deployment consid-

erations, thereby bridging the gap between academic research and routine clinical care in Ethiopia. Although challenges remain particularly around data diversity, minority-class sensitivity and large-scale validation, the foundation laid here offers a robust springboard for future interdisciplinary collaborations. By refining the architecture, expanding local datasets, and embedding interpretability, the biomedical-engineering community can advance toward truly equitable, AI-assisted neuro-imaging diagnostics that save lives across the Horn of Africa and beyond.

Bibliography

[1] I. CiteDrive, "Spinal epidural hematoma symptoms treatment," 2023, https://bonesprain.com/spinal-epidural-hematoma-symptoms-treatment/.

[2] S. N. Ahmed and P. Prakasam, "Intracranial hemorrhage segmentation and classification framework in computer tomography images using deep learning techniques," *Sci. Rep.*, vol. 15, no. 1, p. 17151, May 2025.

[3] A. Morotti and J. N. Goldstein, "Diagnosis and management of acute intracerebral hemorrhage," *Emerg. Med. Clin. North Am.*, vol. 34, no. 4, pp. 883–899, Nov. 2016.

[4] X. Wang, T. Shen, S. Yang, J. Lan, Y. Xu, M. Wang, J. Zhang, and X. Han, "A deep learning algorithm for automatic detection and classification of acute intracranial hemorrhages in head CT scans," *NeuroImage Clin.*, vol. 32, no. 102785, p. 102785, Aug. 2021.

[5] Y. Barhoumi and G. Rasool, "Scopeformer: N-CNN-ViT hybrid model for intracranial hemorrhage classification," 2021.

[6] N. T. Nguyen, D. Q. Tran, N. T. Nguyen, and H. Q. Nguyen, "A CNN-LSTM architecture for detection of intracranial hemorrhage on CT scans," 2020.

[7] P. P. S. Babu and T. Brindha, "Deep learning fusion for intracranial hemorrhage classification in brain CT imaging," *Int. J. Adv. Comput. Sci. Appl.*, vol. 15, no. 8, 2024.

[8] A. Sage and P. Badura, "Intracranial hemorrhage detection in head CT using double-branch convolutional neural network, support vector machine, and random forest," *Appl. Sci. (Basel)*, vol. 10, no. 21, p. 7577, Oct. 2020.

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

[10] "Brain bleed, hemorrhage (intracranial hemorrhage)," *cleveland clinic*, 2023.

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

[12] S. A. Gebremariam and H. S. Yang, "Types, risk profiles, and outcomes of stroke patients in a tertiary teaching hospital in northern ethiopia," *eNeurologicalSci*, vol. 3, pp. 41–47, Jun. 2016.

[13] W. Li, X. Ruan, H. Yang, S. Zhang, F. Rui, and J. Xiong, "Global, regional and national trends in the burden of intracranial hemorrhage, 1990-2021: Results from the global burden of disease study," *Heliyon*, vol. 11, no. 4, p. e42608, Feb. 2025.

[14] T. W. Abate, B. Zeleke, A. Genanew, and B. W. Abate, "The burden of stroke and modifiable risk factors in ethiopia: A systemic review and meta-analysis," *PLoS One*, vol. 16, no. 11, p. e0259244, Nov. 2021.

[15] M. Alene, M. A. Assemie, L. Yismaw, and D. B. Ketema, "Magnitude of risk factors and in-hospital mortality of stroke in ethiopia: a systematic review and meta-analysis," *BMC Neurol.*, vol. 20, no. 1, p. 309, Aug. 2020.

[16] T. Srivastava, R. B. Sannegowda, V. Satija, R. S. Jain, S. Tejwani, and T. Mathur, "Primary intraventricular hemorrhage: clinical features, risk factors, etiology, and yield of diagnostic cerebral angiography," *Neurol. India*, vol. 62, no. 2, pp. 144–148, Mar. 2014.

[17] S. T. W. Thorell., "Intracranial hemorrhage," *National Library of Medicine*, 2022.

[18] H. Ayaz, M. Izzetoglu, K. Izzetoglu, B. Onaral, and B. Ben Dor, "Early diagnosis of traumatic intracranial hematomas," *J. Biomed. Opt.*, vol. 24, no. 5, pp. 1–10, Feb. 2019.

[19] D. M. Alawad, A. Mishra, and M. T. Hoque, "AIBH: Accurate identification of brain hemorrhage using genetic algorithm based feature selection and stacking," *Mach. Learn. Knowl. Extr.*, vol. 2, no. 2, pp. 56–77, Apr. 2020.

[20] W. C. Ziai and D. Hanley, "Intraventricular hemorrhage," in *Stroke syndromes*, L. R. Caplan and J. van Gijn, Eds. Cambridge: Cambridge University Press, Jul. 2012, pp. 526–533.

[21] S. Sumijan, Y. Yuhandri, and W. Boy, "Hybrid thresholding method in detection and extraction of brain hemorrhage on the CT-scan image," *jcsitech*, pp. 7–14, Jun. 2021.

[22] Y. H. Foo, J. H. D. Wong, R. R. Azman, Y. L. Leong, and L. K. Tan, "Identification of acute intracranial bleed on computed tomography using computer aided detection," *J. Phys. Conf. Ser.*, vol. 1497, no. 1, p. 012019, Mar. 2020.

[23] D. Agrawal, L. Poonamallee, and S. Joshi, "Automated detection of intracranial hemorrhage from head CT scans applying deep learning techniques in traumatic brain injuries: A comparative review," *Indian J. Neurotrauma*, vol. 20, no. 02, pp. 081–088, Dec. 2023.

[24] P. Malik, A. Dureja, A. Dureja, R. S. Rathore, and N. Malhotra, "Enhancing intracranial hemorrhage diagnosis through deep learning models," *Procedia Comput. Sci.*, vol. 235, pp. 1664–1673, 2024.

[25] S. Mucha and A. Ramesh Babu, "Classification of intracranial hemorrhage (CT) images using CNN-LSTM method and image-based GLCM features," *MATEC Web Conf.*, vol. 392, p. 01075, 2024.

[26] S. Y. Choi, J. H. Kim, H. S. Chung, S. Lim, E. H. Kim, and A. Choi, "Impact of a deep learning-based brain CT interpretation algorithm on clinical decision-making for intracranial hemorrhage in the emergency department," *Sci. Rep.*, vol. 14, no. 1, p. 22292, Sep. 2024.

[27] T. J. Yun, J. W. Choi, M. Han, W. S. Jung, S. H. Choi, R.-E. Yoo, and I. P. Hwang, "Deep learning based automatic detection algorithm for acute intracranial haemorrhage: a pivotal randomized clinical trial," *NPJ Digit. Med.*, vol. 6, no. 1, p. 61, Apr. 2023.

[28] M. F. Mushtaq, M. Shahroz, A. M. Aseere, H. Shah, R. Majeed, D. Shehzad, and A. Samad, "BHCNet: Neural network-based brain hemorrhage classification using head CT scan," *IEEE Access*, vol. 9, pp. 113 901–113 916, 2021.

[29] L. Cortés-Ferre, M. A. Gutiérrez-Naranjo, J. J. Egea-Guerrero, S. Pérez-Sánchez, and M. Balcerzyk, "Deep learning applied to intracranial hemorrhage detection," *J. Imaging*, vol. 9, no. 2, Feb. 2023.

[30] M. Ammar, M. A. Lamri, S. Mahmoudi, and A. Laidi, "Deep learning models for intracranial hemorrhage recognition: A comparative study," *Procedia Comput. Sci.*, vol. 196, pp. 418–425, 2022.

[31] A. Kundisch, A. Hönning, S. Mutze, L. Kreissl, F. Spohn, J. Lemcke, M. Sitz, P. Sparenberg, and L. Goelz, "Deep learning algorithm in detecting intracranial hemorrhages on emergency computed tomographies," *PLoS One*, vol. 16, no. 11, p. e0260560, Nov. 2021.

[32] S. Y. P. Sharath VN, Varnavi HP, "Hemorrhage detection using cnn," *International Journal of Innovative Research in Technology*, vol. 8, pp. 354–360, 2021.

[33] J. Castro, S. Chabert, C. Saavedra, and R. Salas, ""convolutional neural networks for detection intracranial hemorrhage in ct images"," 03 2020.

[34] A. M. Dawud, K. Yurtkan, and H. Oztoprak, "Corrigendum to "application of deep learning in neuroradiology: Brain haemorrhage classification using transfer learning"," *Comput. Intell. Neurosci.*, vol. 2020, p. 4705838, Aug. 2020.

[35] S. Chilamkurthy, R. Ghosh, S. Tanamala, M. Biviji, N. G. Campeau, V. K. Venugopal, V. Mahajan, P. Rao, and P. Warier, "Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study," *Lancet*, vol. 392, no. 10162, pp. 2388–2396, Dec. 2018.

[36] M. Anouk Stein, C. Wu, C. Carr, G. Shih, J. Kalpathy-Cramer, J. Elliott, kalpathy, L. Prevedello, M. Marc Kohli, M. Lungren, P. Culliton, R. Ball, and S. H. MD, "Rsna intracranial hemorrhage detection," https://kaggle.com/competitions/rsna-intracranial-hemorrhage-detection, 2019, kaggle.

Appendix A

Additional Local Data Evaluation

A.1 Ethical Approval to use Local Data

The project has been approved by City Government of Addis Ababa Health Bureau. The following is a letter (Figure A.1 that allowed us to collect additional data from Wudasie Diagnostic Center.

አዲስ አበባ ከተማ አስተዳደር ጤና ቢሮ
City Government of Addis Ababa Health Bureau

REF.N.O. ለጤ/ጥ/ት/10610/21
DATE 20/5/2016

TO:

- WUDASE DIAGNOSTIC CENTER

Subject: Request to access Facilities to conduct approved research

This letter is to support **LemlemTiruneh** to conduct research which is entitled as **"Classification of Intracranial Hemorrhages in Head CT Scans Using Deep Learning Algorithm."** The study proposal was duly reviewed and approved by Addis Ababa Health Bureau IRB, and the principal investigator is informed with a copy of this letter to report any changes in the study proc edures and submit an activity progress report to the Ethical Committee as required. Therefore we request the Health offices and staffs to provide support to the *Principal* investigator.

With Regards

Ethical Clearance Committee

Ce

- LEMLEMTIRUNEH
- ETHICAL CLEARANCE COMMITTEE

Figure A.1: Screenshot of the Ethical Approval from Addis Ababa Health Bureau

Appendix B

Supplementary Materials

B.1 Code Listings

- Intracranial Haemorrhage Detection and Sub-type Classification Using a Dual-Branch Xception–KNN Framework
- Master's Thesis in Biomedical Engineering
- Addis Ababa University, School of Biomedical Engineering
- Student: Lemelem Tiruneh
- Supervisor: Dr. Melkamu Hunegnaw
- Last updated: June 11, 2025

B.1.1 Introduction

This notebook accompanies a research project focused on detecting and classifying intracranial hemorrhages (ICH) in computed tomography (CT) scans. The approach combines a dual-branch convolutional neural network (CNN) based on the Xception architecture with a K-Nearest Neighbors (KNN) classifier. The model targets the classification of six hemorrhage types: any, epidural, intraparenchymal, intraventricular, subarachnoid, and subdural.

B.1.2 Project Code Overview

- Each Xception branch is trained to extract different levels or types of features from the CT images.
- The extracted features from both branches are concatenated to create a richer, high-dimensional feature representation.
- These concatenated features are used as input to a KNN classifier trained to predict the presence or absence of each ICH subtype.
- This Notebook:

The current notebook demonstrates the **inference stage** of the pipeline:

1. Load and preprocess a new DICOM scan (including windowing and resizing).
2. Extract features using both trained CNN branches.
3. Concatenate the features from both branches.
4. Use a trained KNN model to classify the hemorrhage types.
5. Evaluate the model - Biclass Classification
6. Evaluate the model -Multiclass Classification
7. Evaluate the model -on Local Ethiopian Data

B.1.3 Note

- Please ensure that the trained CNN and KNN models are available in the expected directories.
- Modify file paths and parameters as needed to match your local system setup.

This work contributes to the field of AI-driven radiological diagnostics, aiming to enhance clinical decision-making through robust and interpretable deep learning pipelines.

B.2 Code Availability

The complete implementation that underpins this thesis is hosted as an open Kaggle Notebook titled "Intracranial-Hemorrhage-Detection-and-Sub-type-CL" and can be accessed at https://www.kaggle.com/code/melkamuhunegnaw/intracranial-hemorrhage-detection-and-sub-type-cl/edit. Kaggle's platform was selected because it provides free GPU/TPU resources, a reproducible Linux environment (Ubuntu 20.04, Python 3.10, CUDA 11.x), and seamless, one-click access to the original RSNA competition data.

To reuse or extend the code:

- Create a Kaggle account (free) and make sure your profile is verified for GPU usage.
- Navigate to the link above and click "Copy & Edit". This action spins up a private copy of the notebook in your workspace, preserving every code cell, markdown explanation and pre-configured environment.
- In the right-hand sidebar, go to the "Add data" panel. Attach two data sources:
  - "rsna-intracranial-hemorrhage-detection" (the official 2019 competition dataset)
  - Any private datasets you wish to test (e.g., a ZIP file containing the anonymised Ethiopian scans). Kaggle will mount each dataset read-only under the /kaggle/input/ directory, automatically mirroring the folder hierarchy shown on the competition page as shown in the Figure B.1
- Press Run All. Because the notebook already hard-codes every critical hyperparameter (learning rate, batch size, KNN-k, etc.), the experiment will execute end-to-end without manual intervention. Runtime logs, generated figures, and trained weights are saved into /kaggle/working/, which you can download as a single archive once execution finishes.

- Finally, for the last block, you will be asked to insert a selected subject-wise data from your local dataset for evaluation; you can select the files on the file path from where you uploaded the local data.

```
#######################################################
#### Setting the path and Reading the data
#######################################################
train = pd.read_csv("../input/rsna-intracranial-hemorrhage-detection/rsna-intracranial-hemorrhage-detection/stage_2_train.csv")
sub = pd.read_csv("../input/rsna-intracranial-hemorrhage-detection/rsna-intracranial-hemorrhage-detection/stage_2_sample_submission.csv")
#test = pd.read_csv("../input/rsna-intracranial-hemorrhage-detection/rsna-intracranial-hemorrhage-detection/stage_2_test.csv")
train_images = os.listdir("../input/rsna-intracranial-hemorrhage-detection/rsna-intracranial-hemorrhage-detection/stage_2_train/")
test_images = os.listdir("../input/rsna-intracranial-hemorrhage-detection/rsna-intracranial-hemorrhage-detection/stage_2_test/")
test_images_ET=os.listdir("/kaggle/input/eth-ich-data/ET ICH DATA")
print ('Train:', train.shape[0])
print ('Sub:', sub.shape[0])
#print ('Test:', test.shape[0])
print('section complted')
```

Figure B.1: Screenshot of how the Setting the Path and Reading the Data

B.3 Data Availability

RSNA Intracranial Hemorrhage Detection dataset – All imaging and label files used for model development and internal validation are publicly released by the Radiological Society of North America (RSNA) under a permissive CC-BY 4.0 licence. The collection can be downloaded free-of-charge from Kaggle (https://www.kaggle.com/competitions/rsna-intracranial-hemorrhage-detection), where it is archived as a 2019 competition bundle containing:

- 752 803 anonymised head-CT studies (4.5 million DICOM slices, 430 GB),
- a CSV-based annotation file (*stage_2_train.csv*) that provides image-level multi-label ground-truth for five haemorrhage sub-types and an aggregated "any" label, and
- a companion "sample submission" file that defines the evaluation format used on Kaggle's hidden test set.

The dataset remains actively hosted; each release carries a unique SHA-256 hash, ensuring that future users can reproduce the exact version referenced in this thesis. Readers can obtain the data after creating a free Kaggle account and accepting the competition's licence terms.

Both the Ethiopian clinical dataset – A prospective convenience sample of 15 from Wudassie Diagnostic Center and 23 from Yanet General Hospital, non-contrast head-CT studies were de-identified at source by removing patient names, accession numbers, and all DICOM tags that could reveal personal health information. Imaging diagnoses were established by neuroradiologists.

Because these scans belong to living patients and are governed by Ethiopian data-protection law, the raw DICOM files cannot be placed in an open repository. However, the fully anonymised images, radiology reports, and experiment-ready labels will be shared with academic researchers upon reasonable request to the corresponding author (*lemlemtiruneh27@gmail.com*) and conditional on:

- Approval from the Addis Ababa University Institutional Review Board (IRB) or an equivalent ethics committee,

- Execution of a data-use agreement confirming that the requester will not attempt patient re-identification, and
- Use exclusively for non-commercial scientific research.

Aggregate statistics (class distribution, age/sex counts) are presented in Table 4.7. No individual-level metadata beyond that reported in this thesis will be released.