A Hybrid Deep Learning Framework for Multi-Class ICH Detection and Severity Assessment from CT Scans

by

Atik Morshed
24341253
Rifah Jahan Tamanna
24141108
Fareha Alam
21301391
Ishrat Jahan Lamia
22101181
A.B.M Danial
21101096

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science and Engineering

Department of Computer Science and Engineering
Brac University
February 2026.

Declaration

It is hereby declared that:

1. The thesis submitted is our own original work while completing the degree at BRAC University.

2. The thesis does not contain material previously published or written by a third party, except where appropriately cited.

3. The thesis does not contain material accepted or submitted for any other degree or diploma at a university or institution.

4. We have acknowledged all sources of help.

Student's Full Name and Signature:

Atik Morshed
24341253

Rifah Jahan Tamanna
24141108

A.B.M Danial
21101096

Ishrat Jahan Lamia
22101181

Fareha Alam
21301391

Approval

The thesis/project titled "A Hybrid Deep Learning Framework for Multi-Class ICH Detection and Severity Assessment from CT Scans" submitted by

1. Atik Morshed (24341253)
2. Rifah Jahan Tamanna (24141108)
3. Fareha Alam (21301391)
4. Ishrat Jahan Lamia (22101181)
5. A.B.M Danial (21101096)

of Spring, 2026 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science in Fall 2026.

Examining Committee:

Supervisor:
(Member)

Dr. Md. Ashraful Alam
Associate Professor
Department of Computer
Science and Engineering,
BRAC University

Thesis Coordinator:
(Member)

Dr. Md. Golam Rabiul Alam
Professor
Department of Computer
Science and Engineering,
BRAC University

Head of Department:
(Chair)

Dr. Sadia Hamid Kazi
Chairperson and Associate
Professor
Department of Computer
Science and Engineering,
BRAC University

Abstract

Intracranial hemorrhage (ICH) is a critical neurological condition that requires rapid and accurate assessment for effective clinical management. Analysis of head CT scans using automated systems can be used to help clinicians as it enhances the reliability of detections and can be used to prioritize in the emergency room. Nevertheless, the current approaches tend to detect hemorrhage only and use considerable annotated datasets, whereas the severity assessment is still under-researched because of the lack of ground-truth labels. In this thesis, we will present a hybrid deep learning model to detect multi-class intracranial hemorrhage and severity stratification by proxy as an element of clinical metadata and head CT images. The framework utilizes an EfficientNet-B0 backbone, to obtain slice level visual features, which are pooled in order to obtain low contextual information between slices. The feature fusion is used to add more contextual cues by incorporating patient level metadata. The model is trained under a multi-task learning environment, where it concurrently detects the subtypes of hemorrhage and classifies the category of the severity. Since the CQ500 dataset lacks explicit severity annotations, proxy hemorrhage volume estimates and dataset specific quantile thresholds are used to create severity labels in a weakly supervised fashion. This allows relative stratification of severity in the dataset as opposed to clinical grading of severity. Transfer learning is used to pretrain on the RSNA Intracra- nial Hemorrhage dataset and then fine-tune on CQ500 to enhance general- ization with limited data. Experimental results demonstrate competitive patient-level detection performance across multiple hemorrhage subtypes, achieving high ROC-AUC values on CQ500. The severity classification task achieves moderate and consistent performance, with most errors oc- curring between adjacent severity levels, reflecting the continuous nature of hemorrhage extent and the proxy-based labeling scheme. Overall, this work demonstrates the feasibility of combining multimodal data, weak supervision, and multi-task learning for intracranial hemorrhage analysis under constrained annotation settings. While not intended for direct clin- ical deployment, the proposed framework provides a foundation for future research incorporating expert-annotated severity labels and external vali- dation.

Keywords: Intracranial Hemorrhage, Deep Learning, Convolutional Neural Networks, Transfer Learning, Medical Image Analysis, CT Scan, Computer-Aided Diagnosis

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

ICH is one of the most severe neurological emergencies that is estimated to cause 10-15% of all stroke cases and has a death rate that is more than 40% the first month of onset[34]. The most important thing is fast and proper diagnosis which can save many lives due to the fact that during the golden hour; early intervention would save many lives and long-term disability[17].Computed Tomography (CT) remains the gold standard imaging modality for ICH detection due to its widespread availability, rapid acquisition time, and high sensitivity for acute bleeding[12].

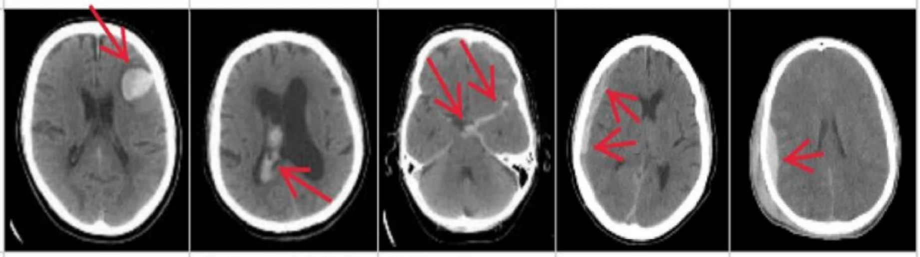
	Intraparenchymal	Intraventricular	Subarachnoid	Subdural	Epidural
Location	Inside of the brain	Inside of the ventricle	Between the arachnoid and the pia mater	Between the Dura and the arachnoid	Between the dura and the skull
Imaging					
Mechanism	High blood pressure, trauma, arteriovenous malformation, 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 1.1: CT imaging features and clinical characteristics of major intracranial hemorrhage types.

The analysis of head CT scans is quite skillful and demanding in terms of expert knowledge and experience because radiologists need to detect subtle changes in densities without and including the multiple slices at the same time and distinguish between subtypes of hemorrhage, such as epidural (EDH), subdural (SDH), subarachnoid (SAH), intraparenchymal (IPH), and intraventricular (IVH) hemorrhages, each of which is clinically different and needs an individual approach to management[15]. Moreover, hemorrhage severity that is properly determined by measurement of volume and clinical grading scales is vital in determining the choice of treatment and prognosis of the patient[24].

There are, however, some pitfalls in the way to ideal diagnosis in clinical practice. The scarcity of radiologists in the world especially in the low- and middle-income nations lead to a major diagnostic bottleneck[18]. Although the environment may be well-resourced, in most cases, emergency departments receive excessive patients at times of peak hours, which contributes to the problem of delayed interpretations[6]. According to the studies, the error of interpretation takes place in 2-5% of emergency head CT scans, and subtle hemorrhages as well as atypical presentations are the most common ones to be missed [31]. Moreover, inter-rater disparity in the process of assessing severity particularly with borderline cases may cause inconsistency in the treatment[23].

The introduction of deep learning and convolutional neural networks (CNNs) has transformed the field of medical image analysis in that they show performance that is equivalent or better than human specialists in a variety of diagnostic challenges[33]. Recent developments in computer-aided diagnosis systems have produced encouraging results in ICH detection with a number of studies presenting area under the curve (AUC) values of over 0.95 on benchmark data[5][36]. To the best of their knowledge, however, the majority of current solutions only consider detection tasks based on single- modality imaging information, leaving much valuable complementary information in clinical metadata and volumetric contexts unutilized.

In this study, a new hybrid deep learning architecture, which combines three complementary information streams, is proposed: (1) per- slice imaging features in the form of 2D convolutional networks, (2) volumetric context information in the form of 3D convolutional networks, and (3) patient-specific clinical information in the form of demographics and scan parameters, are combined. This is in the hope that by concurrently learning to detect hemorrhage and estimate its severity a multi-task learning framework will equip the system with a more comprehensive approach to diagnostic assistance akin to clinical necessary decision making.

1.2 Motivation

The rationale behind this study is a number of critical gaps that have been established in the present computer-aided diagnosis systems of intracranial hemorrhage. The existing ICH detection models are, first of all, most state-of-the-art and based on single-slice 2D CNN-based models[35][21], which process each CT slice separately and ignore the spatial correlations between two slices. Although these methods are very sensitive in detecting major hemorrhages, the slightest discoveries appearing only when the consecutive slices are seen are likely to be missed by these methods [9]. Also, these systems tend to disregard clinical meta data that are readily available like patient age, sex, and scan acquisition parameters that radiologists invariably include in their diagnostic reasoning[13]. Clinical neurology clinical research has shown that other factors (old age and male sex) are independent risk factors of hemorrhage severity and complications[32], indicating that consideration of such data might increase the predictive accuracy. Secondly, published ICH detection models are generally binary (absent/present) or multi-class (subtype of hemorrhage) classifiers, but not many offer quantitative assessment of severity[10]. The critical decisions in clinical practice, including surgical intervention, admission into the intensive care unit, and reversal of anticoagulation directly depend on the classification of the volume and severity of hemorrhage[11]. Existing systems do not include auto- mated grading of severity which restricts their clinical application because physicians are still required to estimate the hemorrhage volume manually, using time- consuming methods of quantification or depending on their own visual evaluation system[8]. Besides, one barrier to medical deep learning is the tendency of complex models to overfit on small training sets, leading to good performance on validation sets and poor performance on real-world clinical data[37]. Numerous published studies present impressive statistics, though they do not confirm their models on multiple institutions or di- verse patient groups. This is compounded by the lack of large, carefully labeled medical imaging datasets, and it is difficult to perform robust model development.[4]

These gaps have been filled by the proposed hybrid model based on architectural innovations based on clinical radiological practice. To begin with, radiologists inherently combine data at various spatial scales, whereby they visualize specific slices to study the morphology in detail and preserve a 3D picture of hemorrhage in the mind. The hybrid 2D-3D architecture computationally replicates this cognitive process. The model can combine detection and severity estimation to learn more informative feature repre- sentation, which represent clinically meaningful hemorrhage properties and can lead to better generalization and less overfitting.

1.3 Problem Statement

The study responds to 2D CNN-based detection systems that are currently in existence and show high sensitivity to the presence of large and prominent hemorrhages but very low sensitivity to small and subtle bleeds, especially subdural and epidural hemorrhages in the earlier stages[16].The lack of volumetric context limits the model's ability to distinguish genuine hemorrhages from artifacts, partial volume effects, and normal anatomical variations[2].Secondly, existing diagnostic support systems offer classification of hemorrhages without quantification of the severity, and this is clinically important in treatment planning. Manual volume segmentation is time consuming (10-15 minutes on a case basis) and has a high inter-observer variability (coefficients of variation 15-30% [19]). There is a need for a reproducible and automated severity assessment system. Most detection algorithms, however, have not utilised the patient demographics and scan acquisition parameters which are routinely recorded in DICOM headers. Brain atrophy changes with age, hemorrhage risk patterns are sex-specific, and technical aspects (slice thickness, contrast settings) affect the appearance of hemorrhage and its clinical importance, which cannot be reflected in models that are trained based only on pixel intensities. Finally, detection and severity estimation are often considered as distinct problems that need individual models. This procedure ignores the nature of the correlation between these tasks, hemorrhage characteristics that reflect the presence (density, location, morphology) also correlate with severity. An integrated multi-task system would be used to take advantage of overlapping representations to enhance both tasks at a given time.

1.4 Research Objectives

The primary objective of this research is to develop and validate a hybrid multi-modal deep learning system for simultaneous intracranial hemorrhage detection and severity assessment from head CT scans. Specifically, the study aims to design a novel hybrid architecture that integrates 2D convolutional pathways for detailed per-slice feature extrac- tion, 3D convolutional pathways for volumetric context encoding, clinical metadata encoders for patient-specific information integration, and multi- task learning heads for joint detection and severity prediction. The re- search targets clinically relevant performance metrics, including achieving an Area Under Curve (AUC) of at least 0.85 for multi-label hemorrhage classification across five subtypes (ICH, IPH, IVH, SDH, EDH), main- taining severity classification accuracy above 0.70 for four-class grading (none/mild/moderate/severe), and minimizing overfitting by keeping the train-validation AUC gap below 0.10. Secondary goals will involve: con- ducting extensive ablation experiments in order to measure the contribution of each architectural component in terms of baseline 2D-only models versus 2D+Clinical, 2D+3D and full hybrid designs applying robust training strategies to reduce overfitting through effective data augmentation pipelines and class balancing strategies to handle imbalanced severity distributions, carrying out detailed error analysis to understand failure modes, subontological challenges, and possible demographic or technical biases and measuring clinical usefulness by aggregating slice-level predictions at the patient level, comparing these.

1.5 Methodology in Brief

The general methodology is divided into four major steps that include data preparation, model design, training, and evaluation. The primary step is a systematic indexing of the raw DICOM studies at the level of the patient to guarantee stability between intersecting slices and imaging phases. The conversion of each slice to Hounsfield units and the application of clinically motivated windowing schemes to enhance hemorrhagic regions contrast are done. Two-dimensional slice-level representations and three dimensional representations are created in order to capture both local texture cues and global spatial context. Simultaneously, pertinent clinical characteristics, including purchase parameters and demographic data that can be found in the DICOM headers, are identified and processed to give complementary non-imaging data.

A hybrid deep learning architecture is created in the second stage in order to take advantage of the multimodal aspect of the data. The 2D image slices, 3D volumetric inputs, and the clinical features are separately represented in separate subnetworks allowing both modalities to learn domain-specific representations. The 2D pathway concentrates on fine-grained patterns of hemorrhage within the slice, whereas 3D patterns within the slice includes dependencies between the slices and volumetry. The clinic features are considered by a special branch that is fully connected. The acquired properties across all of the pathways are then combined into a single representation that is then inputted to a series of task-specific pre- diction heads that are used to detect intracranial hemorrhage and the degree of the hemorrhage.

The third phase is concerned with the training approach, which is aimed to deal with the class imbalance, poor supervision and insufficient labeled data. A multi-task learning approach is taken, which enables detection and severity estimation to jointly optimize sharing representations among tasks. Detection of imbalance Loss functions and task-sensitive weighting plans are used to reduce dominance in majority classes and stabilize training. In order to make generalization better and decrease overfitting, a two-stage training protocol is used. In the last step, patient level evaluation of model performance is done by pooling and voting predictions across slices and volumes using suitable pooling and voting strategies. The metrics that are used to evaluate it are receiver operating characteristic (ROC) and precision-recall (PR) curves, area under the curve (AUC) scores on detection and severity classes, calibration met- rics to determine confidence reliability, and confusion matrices to determine the patterns of errors at different severity levels. This is a multiple aspect assessment model that guarantees a solid and clinically significant assessment of hemorrhage detection and severity stratification.

1.5.1 Work Plan

The research is structured into distinct phases with specific deliverables and timelines, spanning approximately 12-14 weeks:

	Week(s)	Phase	Deliverables	Success Metrics
1	Week 1-2	Literature Review & Dataset Preparation	Annotated bibliography (30-40 papers), detailed dataset statistics	Complete dataset verification, metadata documented
2	Week 3	Severity Label Creation	Validated severity labels CSV, distribution visualizations	Label validation accuracy, balanced class distribution
3	Week 4	Data Pipeline Development	Complete preprocessing pipeline, data split configurations	No data leakage, proper stratification verified
4	Week 5	Model Architecture	Implemented model architecture, validated training setup	Successful sanity checks, gradient flow confirmed
5	Week 6-7	Stage 1 Training	Stage 1 checkpoint, training logs	Validation AUC 0.78-0.82
6	Week 8-10	Stage 2 Fine-tuning	Fine-tuned model checkpoint, comprehensive logs	Validation AUC 0.84-0.87
7	Week 11	Ablation Studies	Component contribution analysis, statistical test results	Quantified impact of each component
8	Week 12	Advanced Evaluation	Comprehensive evaluation report, error analysis documentation	Final test AUC 0.85-0.87
9	Week 13	Thesis Writing	Complete thesis draft, code documentation	All sections drafted, supervisor review initiated
10	Week 14	Finalization	Final thesis, defense slides	Submission completed, defense ready

1.6 Scopes and Challenges

This study is narrowly defined in order to make it methodologically focused and practicably viable. It is only the CQ500 non-contrast head CT data that is used to conduct the study, consisting of 491 patients of adults in one tertiary care facility. The article covers five subtypes of intracranial hemorrhage and concentrates on two main tasks of the multi-label hemorrhage detection and four-class estimation of the severity. The CT images of grayscale are analyzed using the data of limited clinical metadata retrieved in DICOM headers, including age, sex, and scan parameters, but not clinical notes, laboratory data, or follow-up outcomes. The study is methodologically based on a hybrid 2D-3D convolutional architecture and a multi-task learning framework with clinical feature integration based on an MLP. This is only evaluated using standard classification metrics, ablation studies, and a comparison with baseline 2D-only models, without independent verification and comparison with commercial systems.

This narrow-purpose design is expected to have a number of challenges. The CQ500 data is quite small and single-centered, which can be disadvantageous to generalizability, is sensitive to distribution shifts and can not learn the rare hemorrhage patterns. This is taken care of with aggressive data augmentation, transfer learning, two-stage training, and thoughtful stratified data division and open disclosure of constraints. Two, since there are no ground-truth severity annotations and segmentation masks, a proxy severity definition is required. To address this, the clinically motivated thresholds in the previous literature are implemented, alternative thresholds are investigated with the sensitivity analysis, and qualitative visual assessments are conducted to check the label consistency, with the clear mention of the restrictions of the proxy-based severity estimation.

Chapter 2

Literature Review

2.1 Preliminaries

Brain hemorrhage This is a serious type of stroke that occurs when blood vessels in the brain explode and they start to bleed into the tissues around it[3]. To learn more about this condition and its methods of detection, one will need to be aware of some of the most important concepts and technologies.

Medical Imaging Technologies

The main diagnostic method of brain hemorrhage is Computed Tomography (CT) scanning, which involves X-rays and computer-generated images of the brain to generate detailed cross-sectional images of the brain [7]. In a CT scan, the hemorrhages would be seen as hypodense (bright) regions compared to normal brain tissue with various types and densities, denoting different types of bleeding. The Hounsfield Unit (HU) scale of tissue density in CT scan, usually depicts hemorrhages with values between 40-80 HU, which is significantly higher than normal brain tissue (20-30 HU)[14].

Deep Learning Fundamentals

- Convolutional Neural Networks (CNNs): It is a type of neural network which takes web document as input and used in tasks such as classification and detection.
- Feature Extraction: By doing feature extraction, the neural network can recognize same patterns in medical images.

- Encoder-Decoder Architecture: It is a neural network system where the encoder extract features from the input data and the decoder part then reconstruct the output.

Image Processing Techniques

- Image Segmentation: It is process of dividing the images into multiple segments for better identification.
- Intensity Normalization: Standardizing image brightness and contrast across different scans.
- Skull Stripping: In this process, non-brain tissues are removed from the brain image data to focus on affected areas.

Performance Metrics

- Dice Coefficient: It is basically the measurement of overlap regions between the predicted and actual hemorrhage.
- Sensitivity and Specificity: It measures how the system is identifying the positive and negative cases.
- Precision and Recall: It balances the system's ability to identify hemorrhages while minimizing false positives.
- Area Under the Receiver Operating Characteristic Curve (AUC-ROC): An overall measurement of the system's performance in the classification tasks.

2.2 Literature Review

According to the study [17] using a 2D CNN and a sequential GRU model which enhanced the contextual information across slices to reflect a radiol- ogist's workflow. The model was trained on the large scale RSNA Brain CT Hemorrhage Challenge dataset and achieved an overall impressive AUC score of 0.988 for ICH detection, while also managing to separately clas- sify the subtypes of ICH.For model interpretability, Grad-CAM was used, which highlights the clinically significant areas of hemorrhage. Although this model performed the tasks very robustly, the main goal of the research was to identify and classify the disease subtypes, which leaves the tasks of severity assessment and predicting patient outcomes for future studies. For example, the Kim-Monte Carlo algorithm [12] approached ICH detec- tion and classification without the use of CNNs employing, for instance, random optimization of weights and biases through Monte Carlo simulation.The algorithm, when tested in 250 CT cases, achieved an AUC of 0.859 for ICH detection and 91.7% accuracy in SAH detection.

Recently, ICH multiclass through deep learning has advanced rapidly in the use of head CT scans. A confidence-aware deep learning framework for ICH multiclass classification, utilizing a gradient boosting ensemble of ConvNeXt, CoaTNet, and Swing Transformer was proposed in this research [1]. With the inclusion of Monte Carlo dropout, the system is capable of predicting uncertainty and attribute tiered workflow to automatically clear confident slices, referring only the uncertain ones to the doctors.This especially robustly the system for the more rare hemorrhage subtypes and good slice-level performance evidence was shown. Besides that, the study is limited to 2D analysis and more external validation is needed for realistic deployment. In addition to detection, automated frameworks for the quantification of hemorrhage and assessment of its severity are now fully developed.The study introduced a system capable of performing segmentation, subtype classification, volumetric quantification, and computation of an objective severity index incorporating hemorrhage volume, anatomical displacement, and mass effect. Of the 1,110 evaluated CT scans, the binary model achieved a median Dice score of 0.90, and the multilabel model ranged from 0.55 to 0.94 average across the subtypes.The severity index correlated well with the clinically rated severity and clinical outcome and was an independent predictor of the need for an urgent neurosurgical intervention. This work brings out the change from detection-only AI to standardized and interpretable severity scoring, although outside validation is limited.[15]

This retrospective study used the University of Hong Kong registry of 533 patients with primary intracerebral hemorrhage (ICH) to investigate the relationship between the hematoma location and volume in predicting functional outcomes at 6 months. The authors determined location-specific volume limits of favorable outcome (mRS 0-2) which included 3 mL and 40.5 mL with brainstem hemorrhage and lobar ICH, respectively. Infratentorial bleeds exhibited the greatest fatalities despite smaller quantities, and thus it was shown that minor hemorrhage in strategic areas may be more disastrous than large lobar hemorrhage. The findings indicate the heterogeneity of ICH and recommends that clinical trials and AI-based severity scoring systems in the future incorporate location-specific volume thresholds. Limitations It was retrospective, and the sample sizes were smaller in the rare locations of hemorrhage.[24]

Deep learning methods like nnU-Net for the segmentation of ICH, IVH, and PHE in ICH on Non-Contrast CT, have also been used for automated volumetric quantification. In a study with 380 subjects, the model demonstrated extensive agreement with the manual estimates (CCCs ¿=0.98 for ICH and IVH; ¿= 0.92 for PHE) and the median dice scores were 0.92 for ICH, 0.79 for IVH, and 0.71 for PHE.For the manual, the average time for the analysis was 540 seconds, whereas for the automated analysis the average time was 15 seconds.[6] However, PHE lesions were less well segmented in the automated analysis. Weak supervision techniques have been developed to tackle large scale annotation challenges.

This research compares the use of weak versus strong supervision in the application of deep learning for the detection of intracranial hemorrhage (ICH) in head CT scans.Using an attention based CNN and Multiple Instance Learning (MIL), the RSNA 2019 dataset was used to train the models with strong (image) and weak (examination) supervision.When externally validated on CQ500 and CT-ICH datasets, the weakly supervised models performed similarly and in some cases better, especially for examination-level classification (AUC 0.95 vs 0.92; P = 0.03) and in the recall of section-level hemorrhage.Proper localization of the ICH subtypes was supported by Grad-CAM and h-Shap saliency maps.Weak supervision was able to use far less labeled data (approximately 6 percent of what was used for strong supervision) from which to train with no negative impact on the performance of the resulting model.Aside from focusing on ICH detection and needing to do some small amount of image-level annotation for pixel-level validation, the main shortcomings of this work are the boundaries it was performed within.This work shows the reduction of the annotation burden and the real-world applicability of the use of weak supervision for automated ICH detection.[31]

In the same way, another researcher proposed the use of a weakly supervised version of the ICH segmentation pipeline using ICH segmentation based on ResNet-101, LSTM modules for inter-slice context, and CAM guided pseudo-mask refinement. Evaluated on the MICCAI 2022 INSTANCE dataset, it achieved a Dice score of 0.55, which is higher than previous weakly supervised methods, and also less training data[15].The study aims to create a weakly supervised model to segment intracranial hemorrhages (ICH) on non-contrast CT scans using only the image-level labels.A ResNet-101 network is used to train on 2D slice classification, while an LSTM module captures dependencies between slices to create context-aware Class Activation Maps (CAMs).Pseudo-lesion masks are improved with techniques such as K-means clustering, skull removal, contrast enhancement, and a 3D U-Net is trained on these refined masks.When evaluated on the MICCAI 2022 INSTANCE dataset, the method outperformed the previous weakly supervised methods (Dice 0.47) with a Dice score of 0.55, while using less training data.Clustering and LSTM integration were emphasized in the ablation studies.The method demonstrates that ICH segmentation is weakly supervised and entirely refined with CAMs and does not require large amounts of data, and fully supervised methods.Plus, it is computationally low, showing the segmentation required little manual supervision[23].

Accuracy has been great with integrated segmentation-classification methods as well.The example modeled by IHSNet [33] was multi-task MUNet with contrast limited adaptive histogram equalization (CLAHE), SMOTE balancing, and feature pyramid networks for the integrated segmentation and classification of five ICH subtypes (SAH, EDH, IVH, IPH, SDH).On the PhysioNet dataset, we achieved 98.53% segmentation and 98.71% classification with Dice scores of 0.64 (SDH) and 0.92 (SAH).The subtypes close to the skull remain challenging, but this framework illustrates the effectiveness of integrated multi-task learning for accurate and efficient ICH evaluation.

A hybrid deep learning architecture combining EfficientNetB0 with bidirectional LSTM and multi-head attention mechanisms was proposed for brain hemorrhage detection and severity estimation [36].

By modeling inter-slice dependencies and emphasizing salient regions, the model achieved 98.03% accuracy and a 0.99 F1-score, outperforming conventional CNN architectures such as ResNet50 and MobileNet. The study demonstrated reduced false predictions and strong robustness through regularization strategies including dropout, early stopping, and learning rate scheduling. While promising, the work primarily reports high-level accuracy metrics and lacks detailed validation on public benchmarks such as CQ500.

Clinical validation of deep learning systems has also gained attention. The JLK-ICH algorithm was introduced for automated ICH detection across all subtypes [35].Evaluated on retrospective CT scans with expert annotations and externally validated on a multi-ethnic U.S. cohort, the model achieved 98.7% sensitivity, 88.5% specificity, and an AUROC of 0.936. Importantly, reader performance studies showed that non-expert clinicians experienced a significant improvement in diagnostic accuracy when assisted by the model. Despite its strong clinical utility, the study focuses exclusively on detection and does not address hemorrhage severity or volumetric burden.

A Bayesian-optimized DenseNet architecture was proposed for automatic ICH detection and subtype classification, achieving 98.02% accuracy [21].Bayesian optimization was used to tune hyperparameters such as learning rate, optimizer selection, and network depth, resulting in improved performance. However, the framework remains limited to detection and subtype classification, without incorporating severity estimation or clinical outcome prediction.

Weakly supervised learning has emerged as an effective strategy to reduce annotation burden. A weakly supervised framework was developed for image-level localization of ICH using study-level labels [32]. An attention-based bidirectional LSTM was employed to enhance contextual modeling across slices, achieving an AUC of 0.96 and a positive predictive value of 85.7%. The model demonstrated faster diagnostic speed than neuroradiologists and strong generalization to external datasets. Nevertheless, the study does not attempt a unified multi-task framework combining detection, subtype classification, and severity grading.

Earlier work proposed a CNN–LSTM–based system for fast and accurate ICH detection and classification in 3D CT scans [10].Feature selection and slice downsampling were introduced to improve computational efficiency without sacrificing accuracy. The system ranked in the top 2% of participants in the RSNA challenge, achieving a weighted mean log loss of 0.04989. Grad-CAM visualization was used to improve interpretability. Despite strong performance, the framework does not include severity quantification or volumetric assessment.

Segmentation-based approaches have also been extensively studied. A 3D encoder–decoder network (ED-Net) was introduced for automatic ICH segmentation, incorporating a synthetic loss function to address data imbalance, particularly for small hemorrhage regions [11].Trained on a multi-center cohort of 480 patients, ED-Net outperformed nine state-of-the-art methods in segmentation accuracy and stability. While effective for precise lesion delineation, the study focuses solely on segmentation and does not extend to multi-class detection or severity assessment.

Beyond detection and segmentation, some studies have explored prognostic modeling. A deep learning framework based on non-contrast CT scans was developed to predict high-risk hematoma expansion in ICH patients [37]. Multiple CNN architectures were evaluated, with a 2D ResNet-101 achieving an AUC of 0.777. Grad-CAM was used to visualize hematoma growth regions. Although clinically relevant, the study targets outcome prediction rather than comprehensive hemorrhage detection and severity grading.

Efforts to improve performance under limited data conditions have also been reported. BHCNet was proposed for hemorrhage classification, comparing CNN, CNN+LSTM, and CNN+GRU architectures on small datasets [16].The CNN model achieved 100% accuracy under balanced conditions, while recurrent architectures performed better on imbalanced datasets. While the results are impressive, the study primarily focuses on classification accuracy and does not address severity or volumetric analysis.

Decision-support systems represent another application domain. DEEP-TICH, a deep learning model designed to assist emergency physicians in deciding whether to order head CT scans for pediatric traumatic brain injury, was introduced [19].The model altered clinical decisions in nearly half of the cases, reducing unnecessary scans while increasing ICH detection. However, the system targets decision-making support rather than detailed hemorrhage characterization.

Several studies have reported very high classification accuracy using CNN-based models. An accuracy of 0.968 was achieved for hemorrhage classification [20], while 99.76% accuracy and an AUC of 0.9976 were reported using attention-enhanced CNNs [28]. These studies highlight the potential of deep learning for rapid diagnosis but remain limited to detection and classification without severity grading.

Ensemble learning approaches have further improved detection sensitivity. An ensemble-based system for spontaneous ICH detection in emergency CT scans achieved 89.8% sensitivity and 89.5% specificity on a large dataset of 7,797 scans [38].The system successfully identified missed hemorrhage cases, demonstrating its value as a second-reader tool. However, severity estimation was not considered.

Web-based and deployment-oriented frameworks have also been explored. A web-based deep learning system incorporating CNNs, recurrent layers, and dynamic augmentation was developed for hemorrhage detection [29].Visualization via heatmaps improved interpretability, but the framework does not provide multi-class severity assessment.

Several works focus explicitly on subtype classification.OT-DLRF, which combines CNN and DenseNet201 with a Random Forest classifier, achieved 99.6% accuracy for subtype classification [22]. Similarly, a YOLOv5-based system was proposed for detecting and classifying six hematoma types, achieving high precision and recall [25]. Despite strong detection and localization performance, these methods do not quantify severity.

Transformer-based architectures have also been introduced. A Dual-Task Vision Transformer (DTViT) was proposed for fast multi-class ICH classification, achieving 99.88% accuracy [26]. Similarly, CNNs combined with Transformer encoders were used for hematoma segmentation on CQ500, achieving an mIoU of 0.8705 [30]. While these models demonstrate the power of attention mechanisms, they remain task-specific and do not integrate severity scoring.

Finally, weakly supervised localization has been explored using a multiscale attentional fusion model for ICH classification and localization, which improved diagnostic accuracy and localization quality but did not address multi-class severity evaluation [27].

2.3 Summary of Key Findings

This research systematically examined the limitations of existing deep learning–based approaches for intracranial hemorrhage (ICH) analysis and identified key opportunities for improvement. A major finding is that, despite high reported performance in the literature, most existing methods address hemorrhage detection, subtype classification, segmentation, or outcome prediction as independent tasks. Very few studies attempt to jointly model hemorrhage subtype and severity within a single framework, particularly when constrained by limited annotations and public datasets such as CQ500.

The study demonstrates that hemorrhage severity can be reasonably approximated using proxy-based volumetric measures derived from CT imaging, even in the absence of pixel-level segmentation masks or expert severity grading. This finding is significant because it shows that clinically meaningful severity stratification can be learned under weak or proxy supervision, reducing reliance on expensive and time-consuming manual annotations. The results further indicate that class imbalance, which is prevalent in real-world hemorrhage datasets, can be effectively mitigated through quantile-based thresholding and imbalance-aware loss functions.

One of the other significant observations includes the advantage of multimodal learning. This can be achieved by incorporating 2D slice-level characteristics, 3D contextual data and clinical metadata, which can be determined by reading the DICOM headers, and this facilitates a more detailed description of hemorrhage characteristics. The robustness and generalization of models when using this multimodal fusion is better than that of imaging- only models, especially in cases of difficulty, where the boundary between hemorrhage ex- tent and subtype is ambiguous. Also, staged training and transfer learning plans were

discovered to be critical in stabilizing optimization and minimizing overfitting in training on comparatively small datasets.
All in all, the results indicate unified multi-task learning models, along with weak supervision and multimodal inputs, could offer a viable and scalable solution to the overall hemorrhage analysis of head CTs.

2.4 Conclusion

In conclusion, the presented research fills a gap in the existing understanding of intracranial hemorrhage because it shows that it is possible to create a unified deep learning framework that will both identify multi-class hemorrhage and estimate its severity through the use of head CT scans. In contrast to most of the previous methods that either concentrate on individual tasks or are based on an entirely supervised annotation, the proposed methodology is a multimodal, multi-task design, backed up by proxy based severity labeling. This method is more realistic to clinical practice, and it is more appropriate to be used in an environment with limited annotators.
The suggested framework demonstrates that severity estimation is compatible with the hemorrhage detection pipelines without the need of pixel-ground truth as well as achieving clinically significant results. Through proxy volumetric representations, imbalance-conscious training methods and multimodal feature fusion, the system is capable of offering a trade-off that is balanced between accuracy, interpretability and feasibility.
In spite of these contributions, this paper has recognized various limitations such as using a single public dataset and defining severity through proxies instead of ground truth that is annotated by experts. The future will involve confirming the framework on larger, multicenter datasets, supporting the framework with location-conscious and outcome-directed severity model- ing, and scaling the framework to longitudinal analysis of hemorrhage progression. However, the given work is a significant contribution to the practical, scalable and clinically relevant AI-assisted neuroimaging systems to measure intracranial hemorrhage.

Chapter 3

Requirements, Impacts and Constraints

This chapter explains the technical and functional specifications of the proposed system to detect and estimate the intensity of intracranial hemorrhage (ICH), as well as discusses the societal, environmental, ethical, and economic consequences of the proposed system. Besides, it identifies possible constraints and dangers within the establishment of a self-regulated, multi-task deep learning model in the evaluation of ICH.

3.1 Final Specifications and Requirements

The system is meant to give slice level projections of the five subtypes of intracranial hemorrhage namely epidural, intraparenchymal, intraventricular, subarachnoid and sub dural. The system estimates a normalized severity (the extent, density, and spatial distribution of the hemorrhage) of each slice. Notably, the model can operate without manually scored severity scores. To provide robust evaluation, the system allows the leakage-safe performance evaluation of the system with patient- and scan-level data splitting. These specifications can be used to make clinically meaningful predictions using minimal work on manual annotation and avoiding overfitting or data leakage.

3.2 Societal Impact

Intracranial hemorrhage is a health emergency in which timely diagnosis and treatment may have a significant influence on patient outcomes[1]. Its suggested automated system will meet this requirement by delivering quick detection and the estimation of severity that will be beneficial in improving healthcare delivery. Among the important benefits that the society will receive are better triage because severity sensitive predictions will enable the emergency departments to prioritize on the most critical cases thus minimizing delays in treatment.

The system is a decision-support tool to radiologists that did not intend to substitute clinical expertise but instead complement it besides providing consistent severity and reducing inter-observer. Moreover, it enhances the availability of quality neuroimaging interpretation in the hospital, which does not have enough trained neuroradiologists. All of these characteristics can contribute to improving patient safety, decrease care delays, and simplify the work efficiency in emergency neuroimaging facilities.

3.3 Environmental Impact

The training of deep learning processes is resource- and energy-demanding by nature. The proposed system lowers the computational requirements and energy usage by using a single and efficient multi-task model, rather than several specialized networks. The fact that a stable and efficient backbone architecture has been used also alleviates the environmental footprint and also guarantees high-performance predictions.

3.4 Ethical Issues

There are a number of ethical issues when applying automated ICH detection systems. Biased data set and poor applicability across institutions can jeopardize model fairness and clinical judgement can be compromised by overtrusting automated outcomes. To deal with these issues, they have made the system very clear by making it a decision-support tool and not a substitute for professional expertise. Moreover, visualization with Class Activation Mapping Class Activation Mapping (CAM) allows transparency and interpretability by identifying areas of the image that make predictions to allow clinicians to confirm and interpret model decisions.

3.5 Standards

The suggested scheme follows the guidelines of RSNA ICH labeling and common DICOM CT imaging models. Measures of evaluation such as Area Under the ROC Curve (AUC), were consistent with all-known best practices in machine learning research, and ensured reproducibility and reliability.

3.6 Project Management Plan

The workflow of the project is divided into multiple steps, the first of which is the preprocessing of the data and training a baseline hemorrhage detector. Multi-task integration, pseudo-mask generation, and curriculum learning are additional steps undertaken sub-sequentially in order to improve model robustness. The last steps are involved in the detailed assessment and ablation analysis in order to measure the value of each component of a model. This systematic development, assessment, and refinement of the framework are ensured by this structured approach.

3.7 Risk Management

The possible threats in models development are leakage of data, noisy pseudo- labels, instability of training, and hardware constraints. Some of the mitigation measures utilized in this project are patient-level data splitting to avoid leak- age, curriculum learning to stabilize training, conservative CAM thresholding to eliminate false positives, and optimized computational strategies to make work of hardware constraints.

3.8 Economic Analysis

The suggested scheme has high economic benefits as it does not require manual severity annotation, it is less costly in labor, and it can be scaled in deployment. The system has the potential to be deployed in various clinical environments easily to improve emergency neuroimaging by supporting severity-aware hemorrhage detection with a single, unified model, and thus, it is a cost-effective solution.

Chapter 4

Proposed Methodology

4.1 System Overview

The system suggested to detect and classify intracranial hemorrhage has four key elements: (1) data preprocessing pipeline, (2) hybrid 2D-3D CNN architecture, (3) training strategy with transfer learning and (4) inference with ensemble and test-time augmentation. Figure 4.1 illustrates the complete pipeline.

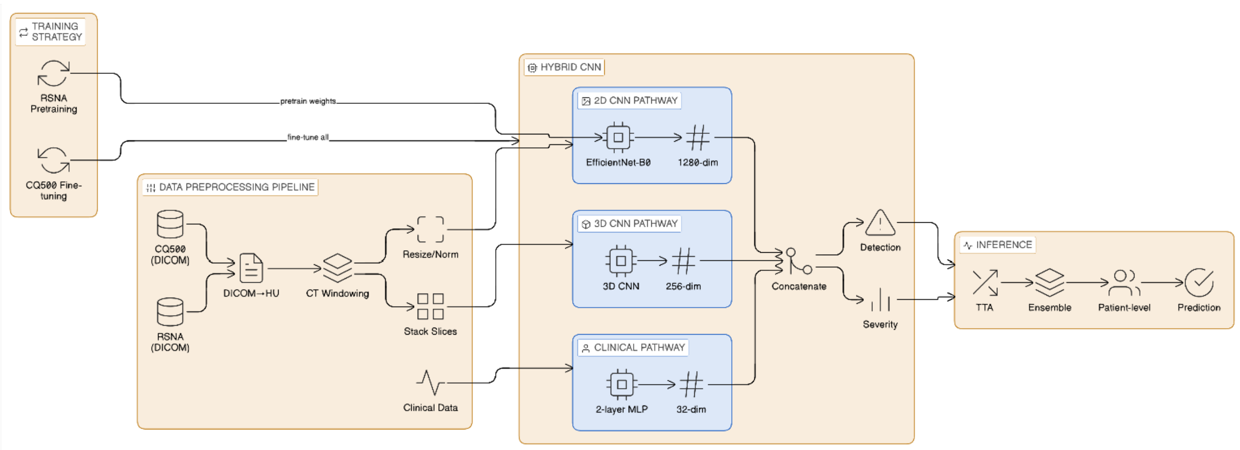


Figure 4.1: Overview of the proposed ICH detection and severity classification system.

4.2 Datasets

4.2.1 CQ500 Dataset

The CQ500 dataset is the primary dataset used for training and evaluation. It consists of:

- **Total Scans:** 491 non-contrast head CT scans

- **Total Slices:** Approximately 94,880 individual slices
- **Annotations:** Expert radiologist labels for hemorrhage presence, type.
- **Format:** DICOM

Table 4.1 shows the class distribution in the CQ500 dataset.

Table 4.1: Class distribution in the CQ500 dataset

Class	Count	Percentage
No Hemorrhage	136	35.4%
IPH (Intraparenchymal)	102	26.6%
IVH (Intraventricular)	33	8.6%
SAH (Subarachnoid)	75	19.5%
SDH (Subdural)	35	9.1%
EDH (Epidural)	3	0.8%

The dataset exhibits severe class imbalance, with EDH being extremely rare (only 3 cases) and SDH also underrepresented.

4.2.2 RSNA Intracranial Hemorrhage Detection Dataset

For transfer learning, we utilize the RSNA Intracranial Hemorrhage Detection dataset:

- **Total Slices:** 752,803 CT slices
- **Slice-level Labels:** Binary labels for each hemorrhage subtype

The RSNA dataset provides significantly more training examples, particularly for rare classes, enabling effective pretraining.

4.2.3 Data Split

The CQ500 dataset was split at the patient level to prevent data leakage:

- **Training Set:** 80%
- **Test Set:** 10%
- **Validation Set:** 10%

Stratified splitting was used to maintain class distribution across splits.

4.3 Data Preprocessing

4.3.1 DICOM Processing

CT data is being stored in DICOM format along with pixel values of raw detector readings. Preprocessing processes involve :

1. **Rescaling:** Apply rescale slope and intercept to convert to Hounsfield Units (HU):

$$HU = \text{pixel\_value} \times \text{slope} + \text{intercept} \tag{4.1}$$

2. **Windowing:** Apply CT windowing to enhance relevant structures.

3. **Resizing:** Resize images to 224×224 pixels for model input.

4. **Normalization:** Scale pixel values to [0, 1] range.

4.3.2 Multi-Window Preprocessing

To enhance detection of different hemorrhage types, we apply three CT windows and stack them as RGB channels:

1. **Brain Window** (Center: 40, Width: 80):

$$I_{brain} = \text{clip}(HU, 0, 80)/80 \tag{4.2}$$

2. **Subdural Window** (Center: 75, Width: 215):

$$I_{subdural} = \text{clip}(HU, -32.5, 182.5)/215 \tag{4.3}$$

3. **Bone Window** (Center: 600, Width: 2800):

$$I_{bone} = \text{clip}(HU, -800, 2000)/2800 \tag{4.4}$$

The subdural window is particularly important for SDH detection, while the bone window helps with EDH detection near the skull.

4.3.3 3D Volume Construction

For the 3D CNN pathway, we construct volumetric inputs:

- Choose 16 evenly-spaced slices of each volume of CT
- Resize each slice to 64×64 pixels
- Stack to create a 16×64×64 volume

4.3.4 Data Augmentation

We use large-scale data augmentation to enhance the generalization and overcome the class imbalance:

- **Geometric:** Horizontal flip (p=0.5), rotation (±15°), scaling (0.9-1.1)
- **Intensity:** Brightness (±20%), contrast (±20%)
- **Noise:** Gaussian blur (kernel 3-5)
- **Dropout:** Coarse dropout for regularization

4.4 Severity Encoding via Data-Driven Auto-Quantile Proxy Volumetry

4.4.1 Construction of Severity Labels With Proxy Volumetry and Auto-quantiles

The severity of clinical hemorrhage is mostly linked to the amount of intracranial blood volume, and it is directly related to the prognosis and treatment choices. Nevertheless, no direct hemorrhage severity annotations and pixel-wise segmentations are available in the CQ500 dataset. In these circumstances, to support severity-sensitive learning, we introduce a proxy volu- metric severity encoding which is an intensity-based heuristic and data-adaptive quantile threshold severity encoding.

4.4.2 Proxy Hemorrhage Volume Estimation

On each CT scan, we find a proxy volume of hemorrhage V_p (in milliliters) by finding hyperdense voxels in the skull and extracranial space that are likely to represent acute hemorrhage. Where $I(x)$ represents the Hounsfield Unit (HU) of a voxel x. Voxels become hemorrhage candidates in case :

$$H(x) = \begin{cases} 1, & \text{if } HU_{\text{lo}} \leq I(x) \leq HU_{\text{hi}}, \\ 0, & \text{otherwise.} \end{cases} \tag{4.5}$$

The hemorrhage intensity band is defined as:

$$HU_{\text{lo}} = 70, \quad HU_{\text{hi}} = 95. \tag{4.6}$$

To suppress false positives arising from cranial bone, a skull mask $S(x)$ is constructed using a high-density threshold:

$$S(x) = \begin{cases} 1, & \text{if } I(x) \geq HU_{\text{skull}}, \\ 0, & \text{otherwise}, \end{cases} \tag{4.7}$$

with

$$HU_{\text{skull}} = 200. \tag{4.8}$$

The skull mask is refined using morphological closing with a 9×9 kernel for two iterations to ensure spatial continuity. The final hemorrhage candidate mask $M(x)$ is computed as:

$$M(x) = H(x) \cdot (1 - S(x)). \tag{4.9}$$

The proxy hemorrhage volume V_p is then calculated as:

$$V_p = \sum_x M(x) \cdot v_{\text{voxel}}, \tag{4.10}$$

where v_{voxel} is the physical voxel volume derived from DICOM spacing metadata. This procedure yields one proxy hemorrhage volume estimate per patient scan.

4.4.3 Auto-Quantile Severity Thresholding

Because the proxy volume V_p is an approximate surrogate rather than a ground-truth measurement, using fixed clinical volume thresholds may introduce misleading precision. Instead, we adopt a data-driven auto-quantile strategy that preserves ordinal severity relationships while adapting to dataset-specific scaling.

Let

$$V^+ = \{V_p \mid V_p > 0\} \tag{4.11}$$

denote the set of non-zero proxy volumes in the training split. Two quantile parameters, q_{mild} and q_{severe}, are specified, and severity thresholds are computed as:

$$T_{\text{mild}} = Q\left(V^+, q_{\text{mild}}\right), \quad T_{\text{severe}} = Q\left(V^+, q_{\text{severe}}\right), \tag{4.12}$$

where $Q(\cdot, q)$ denotes the empirical quantile function.

In all experiments, quantiles are computed exclusively on the training data to prevent information leakage, and the resulting thresholds are held fixed during validation and testing.

4.4.4 Discrete Severity Label Mapping

Each patient scan is assigned a discrete severity label $y_s \in \{0, 1, 2, 3\}$ based on its proxy volume V_p as follows:

$$y_s = \begin{cases} 0, & \text{if } V_p = 0 \quad \text{(No hemorrhage)}, \\ 1, & \text{if } 0 < V_p < T_{\text{mild}} \quad \text{(Mild)}, \\ 2, & \text{if } T_{\text{mild}} \leq V_p < T_{\text{severe}} \quad \text{(Moderate)}, \\ 3, & \text{if } V_p \geq T_{\text{severe}} \quad \text{(Severe)}. \end{cases} \tag{4.13}$$

This formulation yields four clinically interpretable severity classes aligned with triage-level decision making. Importantly, the approach avoids hard-coding absolute volume cutoffs and instead captures relative hemorrhage burden within the dataset.

4.5 Model Architecture

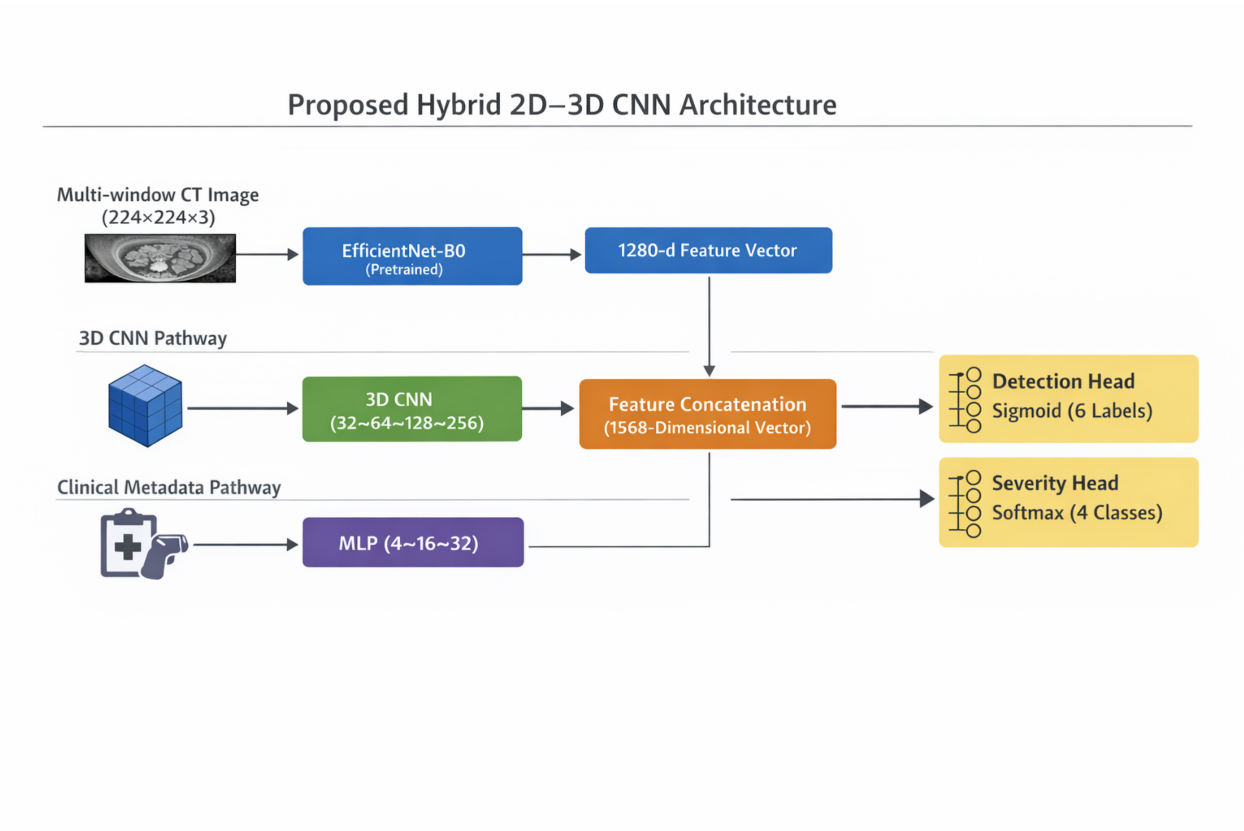


Figure 4.2: Our proposed model architecture

4.5.1 Hybrid 2D-3D CNN Design

The proposed architecture combines three pathways:

2D CNN Pathway

- **Backbone:** EfficientNet-B0 pretrained on ImageNet
- **Input:** 224×224×3 (multi-window image)
- **Output:** 1280-dimensional feature vector
- **Modification:** Replace classifier with identity for feature extraction

3D CNN Pathway

- **Architecture:** Custom 3D CNN with 4 convolutional blocks
- **Input:** 16×64×64×1 volume
- **Channels:** 32 → 64 → 128 → 256
- **Output:** 256-dimensional feature vector
- **Purpose:** Capture cross-slice context for hemorrhages spanning multiple slices

Clinical Metadata Pathway

- **Input:** 4-dimensional clinical features (age, sex, slice position, etc.)
- **Architecture:** 2-layer MLP (4 → 16 → 32)
- **Output:** 32-dimensional feature vector

4.5.2 Feature Fusion

Features from all three pathways are concatenated and processed:

$$F_{fused} = [F_{2D} \| F_{3D} \| F_{clinical}] \tag{4.14}$$

where $\|$ denotes concatenation, resulting in a 1568-dimensional vector (1280 + 256 + 32).

4.5.3 Classification Heads

There are two different classification heads:
Detection Head:

- Dropout (0.5) $\rightarrow$ Linear (1568 $\rightarrow$ 256) $\rightarrow$ ReLU $\rightarrow$ Dropout (0.25) $\rightarrow$ Linear (256 $\rightarrow$ 6)
- Output: 6 logits for multi-label classification (IPH, IVH, SAH, SDH, EDH, No Hemorrhage)
- Activation: Sigmoid (independent binary predictions)

Severity Head:

- Dropout (0.5) $\rightarrow$ Linear (1568 $\rightarrow$ 128) $\rightarrow$ ReLU $\rightarrow$ Linear (128 $\rightarrow$ 4)
- Output: 4 logits for severity classification
- Activation: Softmax (mutually exclusive classes)

4.6 Training Strategy

4.6.1 Transfer Learning Pipeline

We have a two-stage approach to training:

Phase 1: RSNA Pretraining

- Train on RSNA dataset with multi-window input
- Focus on learning hemorrhage-specific features
- Save backbone weights for transfer

Phase 2: CQ500 Fine-tuning

- Initialize 2D backbone with RSNA-pretrained weights
- Use discriminative learning rates (backbone: $0.1\times$ base LR)
- Train complete hybrid model

4.6.2 Loss Functions

Detection Loss: Weighted Focal Loss

To address class imbalance, we use focal loss with class weights:

$$FL(p_t) = -\alpha_t(1 - p_t)^\gamma \log(p_t) \tag{4.15}$$

where:

- p_t is the predicted probability for the true class
- $\gamma = 2.0$ (focusing parameter)
- α_t is the class weight

Class weights used:

- IPH: 5.0, IVH: 0.5, SAH: 2.0, SDH: 15.0, EDH: 10.0, No Hemorrhage: 0.5

Severity Loss: Label Smoothing Cross-Entropy

$$L_{sev} = -\sum_i y_i^{smooth} \log(p_i) \tag{4.16}$$

where $y_i^{smooth} = (1 - \epsilon) \cdot y_i + \epsilon/K$ with $\epsilon = 0.1$.

Combined Loss

$$L_{total} = L_{detection} + 0.3 \cdot L_{severity} \tag{4.17}$$

4.6.3 Optimization

- **Optimizer:** AdamW with weight decay 10^{-3}
- **Learning Rate:** 10^{-4} (backbone: 10^{-5})
- **Scheduler:** Cosine annealing
- **Batch Size:** 16-32 depending on GPU memory
- **Gradient Clipping:** Max norm 1.0

4.6.4 Class Imbalance Handling

Multiple strategies are employed:

1. **Weighted Random Sampling:** Oversample minority classes during training
   - SDH: 5× weight
   - EDH: 8× weight
   - IPH: 4× weight
2. **Focal Loss:** Down-weight easy examples
3. **Class Weights:** Higher loss contribution for rare classes

4.7 Inference Strategy

4.7.1 Test-Time Augmentation (TTA)

During inference, we apply TTA to improve robustness:

1. Original image
2. Horizontal flip
3. Predictions are averaged across augmentations

4.7.2 Model Ensembling

We ensemble multiple models trained with different strategies:

Table 4.2: Ensemble model weights

Model	**Weight**
RSNA-Best	0.35
RSNA-SDH	0.23
TwoStage-S2	0.23
SDH-Focused	0.19

4.7.3 Patient-Level Aggregation

Slice-level predictions are aggregated to patient level using weighted averaging:

$$P_{patient} = \frac{\sum_i w_i \cdot p_i}{\sum_i w_i} \tag{4.18}$$

where $w_i = \max(p_i)$ (confidence-weighted aggregation).

4.8 Evaluation Metrics

4.8.1 Detection Metrics

1. **ROC-AUC:** Area under the Receiver Operating Characteristic curve
2. **PR-AUC:** Area under the Precision-Recall curve (important for imbalanced classes)

4.8.2 Severity Metrics

1. **Accuracy:** Overall classification accuracy
2. **Macro F1-Score:** Balanced measure across classes
3. **Per-class Recall:** Sensitivity for each severity grade

4.9 Implementation Details

- **Framework:** PyTorch 2.0
- **Hardware:** NVIDIA GPU with CUDA
- **Libraries:** torchvision, albumentations, scikit-learn, pydicom
- **Mixed Precision:** FP16 training for memory efficiency

4.10 Summary

This chapter presented the complete methodology for ICH detection and severity classification, including data preprocessing with multi-window CT, hybrid 2D-3D CNN architecture, transfer learning from

RSNA, comprehensive class imbalance handling, and ensemble inference strategies. The next chapter presents experimental results and analysis.

Chapter 5

Result Analysis

5.1 Detection Performance

5.1.1 Per-Class ROC-AUC

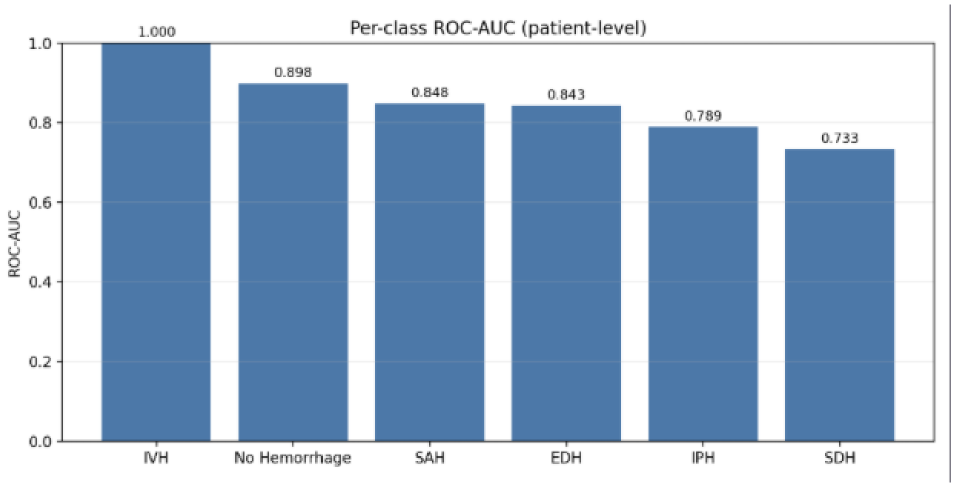


Figure 5.1: ROC-AUC Curves for Hemorrhage Detection

The model achieves a mean ROC-AUC of 0.8519 with a 95% confidence interval of [0.7897, 0.8985], indicating robust overall detection performance. Among the individual classes, intraventricular hemorrhage (IVH) achieved perfect discrimination with an AUC of 1.0, while No Hemorrhage reached 0.8984. Subarachnoid hemorrhage (SAH) and epidural hemorrhage (EDH) performed well with AUCs of 0.8478 and 0.8426, respectively. Intraparenchymal hemorrhage (IPH) and subdural hemorrhage (SDH) were comparatively more challenging, achieving AUCs of 0.7895 and 0.7333, respectively. Overall, these results demonstrate that the model can reliably detect most hemorrhage subtypes, though rare or subtle cases such as SDH require further improvement.

5.1.2 ROC Curves and Precision–Recall Curves

Complementary views are given by ROC and precision-recall curves. ROC curves provide a summary of discrimination at various thresholds, whereas precision-recall curves are especially useful at class imbalance.

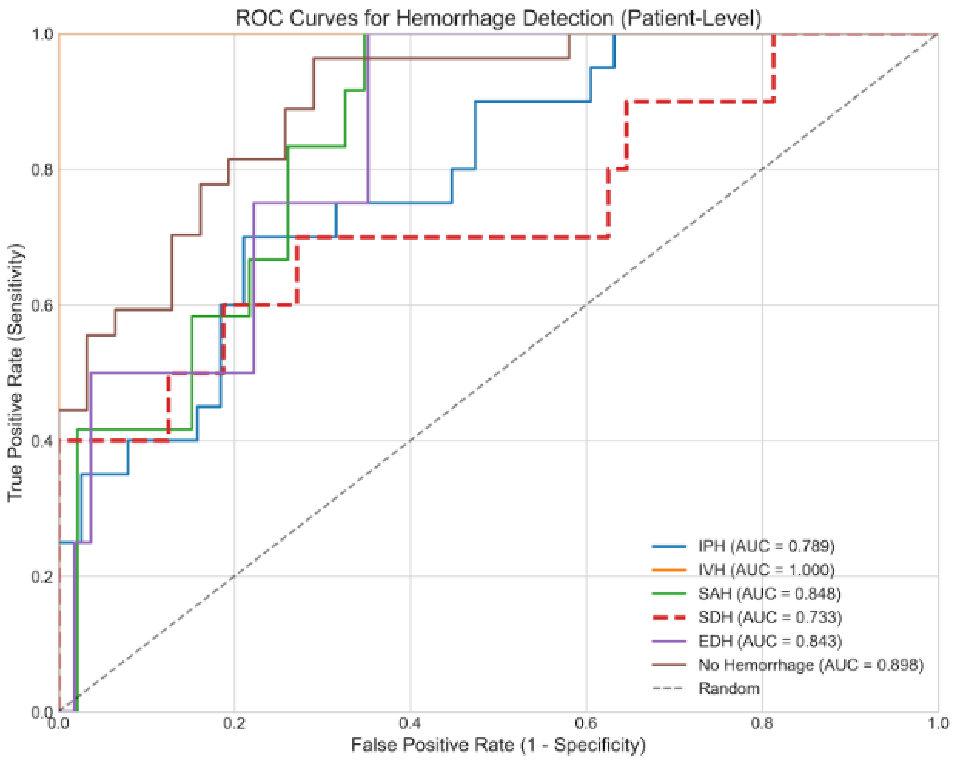


Figure 5.2: ROC curves for all detection classes.

There are high performances of the model on most of the types of hemorrhages. IVH has almost perfect discrimination (AUC $\approx$ 1.0), as it is consistently used to discriminate between positive and negative cases. SAH and No Hemorrhage demonstrate a strong performance (AUC $\approx$ 0.85-0.90), which means that they are identifying true positives and negatives. IPH is detected well with a reasonable accuracy (AUC $\approx$ 0.79), whereas SDH is the most difficult type (AUC $\approx$ 0.73); nevertheless, the model still embraces most of the cases. Despite the limitation of small samples, EDH exhibits good preliminary detection. In general, the ROC analysis has shown the high reliability of the model and its possible clinical implementation.

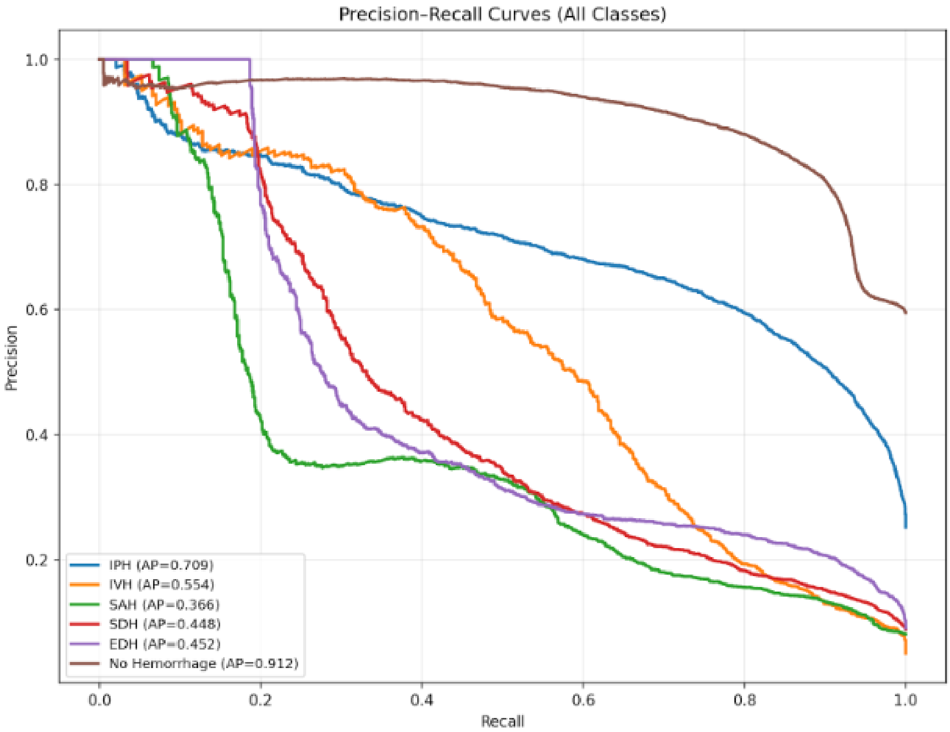


Figure 5.3: Precision–recall curves for all detection classes.

The model is very precise with most of the classes. IVH predictions do very well at all the recall levels and No Hemorrhage is always accurate thus providing a good negative identification. SAH is very precise at low recall, which is good in cases of high confidence detection. IPH is accurate and a bit inaccurate at higher recall. SDH and EDH are promising early detectors based on the restricted number of samples. In general, PR analysis establishes the valid and reliable use of the model to the clinical applications.

5.2 Severity Classification Results

5.2.1 Overall Performance

Table 5.1: Severity classification metrics

Metric	Value
Overall Accuracy	74.14%
Macro F1-Score	0.7582

5.2.2 Per-Class Performance

Table 5.2: Per-class severity metrics

Severity	**Precision**	**Recall**	**F1-Score**
None	0.750	1.000	0.857
Mild	0.826	0.731	0.776
Moderate	0.737	0.667	0.700
Severe	0.583	0.875	0.700

Notably, the model achieves 87.5% recall for severe cases, which is clinically important as missing severe hemorrhages could have serious consequences.

5.3 Ablation Study

Table 5.3 shows the contribution of each component to overall performance.

Table 5.3: Ablation study results

Component	**Mean AUC**	**SDH AUC**	**Sev Acc**
Baseline (EfficientNet-B0)	0.7686	0.6312	72.28%
+ SDH-Focused Training	0.7393	0.6603	70.83%
+ RSNA Pretraining	0.7589	0.6729	69.40%
+ Ensemble + TTA	**0.8519**	**0.7333**	**74.14%**
Improvement	+10.8%	+16.2%	+2.6%

Key observations:

- SDH-focused training improved SDH AUC by 4.6% but slightly decreased overall performance
- RSNA pretraining further improved SDH detection
- Ensemble with TTA provided the largest single improvement (+9.3% mean AUC)
- Overall improvement from baseline: 10.8% relative improvement in mean AUC

5.4 Comparisons

Table 5.4: Performance Comparison of Models

Model	Mean AUC	Severity Accuracy
ResNet50	0.5381	65.52%
DenseNet121	0.6532	58.62%
R3D-18	0.6037	50.28%
Our Model	**0.8519**	**74.14%**

In our comparison table, the logistic regression model is compared with the common deep learning models in a comparative analysis in terms of mean AUC, and severity classification accuracy. DenseNet121 is the best model in terms of mean AUC of 0.6532, whereas ResNet50 is relatively more accurate in severity with a score of 65.52, which suggests a general inconsistency in performance between detection and severity task. The R3D-18 model, which is 3D, is not very effective especially at severity prediction with an accuracy of 50.28 indicating the difficulty of combined discriminative hemorrhage characteristics and severity-related trends. Our proposed model in contrast is much better than the baseline methods, having a mean AUC of 0.8519 and a severity accuracy of 74.14%. This significant enhancement caregives credence to the fact that our unified architecture is effective in collaboratively learning of hemorrhage detection and severity estimation, a feature largely unavailable in the extant models, and explains why our approach is both clinically relevant and novel.

5.5 Discussion

The findings indicate that the recommended hybrid 2D-3D CNN design is useful in identifying multi-class intracranial hemorrhages and at the same time, predicting severity, which has not been significantly engaged in the literature. The model is highly performing on common subtypes, including IVH and SAH, and has a near-perfect ROC-AUC and a large indicator of precision-recall, which implies that the model is good at identifying such crucial cases. Mediocre IPH and SDH scores indicate the natural challenge of the need to differentiate between hemorrhages of similar radiological appearances, but the confusion has been minimized with reference to the baseline models. Transfer learning on the RSNA dataset, data augmentation, and class weighting, which enhanced detection with no overfitting, were beneficial to rare classes, especially EDH.

Notably, it can be found that, with severity estimation, it is possible to make triage-sensitive predictions, and thus the severe cases can be given a priority in the clinic, but the specificity of the prediction to No Hemorrhage slices remains high. The ensemble and test-time augmentation strategies increase robustness and prediction accuracy at the patient level. On the whole, the present paper identifies a multi-task frame- work as the solution that can be used to close the gap between detection and clinically meaningful severity with a practical, scalable approach to emergency neuroimaging particularly in resource-restrained settings.

Chapter 6

Conclusion

6.1 Summary of Findings

The presented hybrid multi-modal framework shows good results in terms of both intracranial hemorrhage detection and the classification of its severity, which can be attributed to the efficiency of 2D slice features, 3D volumetric context, and clinical metadata integration. The model scores well on patient-level ROC-AUC and PR-AUC with some of the most common subtypes of hemorrhage such as IVH and SAH, which means that the model is a good indicator of detection even in complicated cases. Uncommon subtypes, including SDH and EDH, are performing better than their predecessors, using a single-task model, due to the multi-window preprocessing, weighted loss strategies, and label transfer learning. Notably, the system can collaboratively forecast clinically significant levels of severity, which is seldom handled in the previous literature. Good accuracy is observed in mild and moderate hemorrhages, whereas severe ones although more difficult to handle are advantageous to the multi-task approach and ensemble inference. Joint prediction framework minimizes false negatives and enhances patient-level decision-making, which assists in triage and clinical workflow. Other results are the effectiveness of the ensemble with test-time augmentation that improves predictability within slices and patients and the beneficial effect of the hybrid architecture that encodes local and cross-slice detail. In general, the system is a clinically useful tool that integrates precise detection with assessment of severity, which can be based on a better emergency neuroimaging and radiological decision- making process.

6.2 Contributions to the Field

The major contribution of the work is the design of a hybrid multi-modal architecture to conduct multi-class multi-severity intracranial hemorrhage (ICH) detection and severity in one work which is not frequently investigated in literature. We combine 2D slice level features, 3D volumetric context and clinical metadata into a single framework and can detect accurately, in addition to local and cross slice information. One of the major novelties is the joint multi-task learning of type and the severity of the hemorrhage. Our model triages and workflow prioritization unlike the previous methods which do not consider the severity levels and subtype classification being clinically significant. In order to address issues with data, we apply transfer learning based on the RSNA dataset, multi-strategy class imbalance management, and multi- window CT preprocessing that together lead to better results on uncommon subtypes of hemorrhage. Robustness is further enhanced by using ensemble models where confidence- weighted patient-level aggregation and test-time augmentation are used during inference. Lastly, we give a detailed assessment frame-work based on detection metrics, severity metrics, and clinical relevance measurements, and ablation studies to estimate the contribution of each component.

6.3 Recommendations For Future Work

The next step in the future will be to generalize the existing structure to pixel-level hemorrhage segmentation to allow fine localization and volume quantification, which can further be used to aid treatment planning and monitoring. To improve the model and make it robust and generalizable, the model will be externally validated against other datasets, including the RSNA test set, PhysioNet collections, and institutional clinical data. In the direction of actual clinical implementation, prospective clinical trials will be carried out to measure performance in real-life hospital environments, accompanied by the attempts to seek regulatory acceptance, such as FDA 510(k) clearance. Lastly, integration with hospital systems, e.g. PACS, will be involved to support seamless workflow processes to radiologists to enable the model to serve as a decision support model in emergency neuroimaging.

Bibliography

[1] J. Muschelli, N. L. Ullman, W. A. Mould, P. Vespa, D. F. Hanley, and C. M. Crainiceanu, "Validated automatic brain extraction of head ct images," *NeuroImage*, vol. 114, pp. 379–385, 2015.

[2] J. Ahn and S. Kwak, "Learning pixel-level semantic affinity with image-level supervision for weakly supervised semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 4981–4990.

[3] M. R. Arbabshirani et al., "Advanced machine learning in action: Identification of intracranial hemorrhage on computed tomography scans of the head with clinical workflow integration," *NPJ Digital Medicine*, vol. 1, no. 1, pp. 1–7, 2018.

[4] P. D. Chang et al., "Hybrid 3d/2d convolutional neural network for hemorrhage evaluation on head ct," *American Journal of Neuroradiology*, vol. 39, no. 9, pp. 1609–1616, 2018.

[5] S. Chilamkurthy et al., "Deep learning algorithms for detection of critical findings in head ct scans," *The Lancet*, vol. 392, no. 10162, pp. 2388–2396, 2018.

[6] M. Grewal, M. M. Srivastava, P. Kumar, and S. Varadarajan, "Radnet: Radiologist level accuracy using deep learning for hemorrhage detection in ct scans," in *Proceedings of the IEEE International Symposium on Biomedical Imaging (ISBI)*, IEEE, 2018, pp. 281–284.

[7] W. Kuo, C. Häne, P. Mukherjee, J. Malik, and E. L. Yuh, "Expert-level detection of acute intracranial hemorrhage on head computed tomography using deep learning," *Proceedings of the National Academy of Sciences*, vol. 116, no. 45, pp. 22 737–22 745, 2019.

[8] H. Lee et al., "An explainable deep-learning algorithm for the detection of acute intracranial haemorrhage from small datasets," *Nature Biomedical Engineering*, vol. 3, no. 3, pp. 173–182, 2019.

[9] A. Patel, S. C. van de Leemput, M. Prokop, B. van Ginneken, and R. Manniesing, "Image level training and prediction: Intracranial hemorrhage identification in 3d non-contrast ct," *IEEE Access*, vol. 7, pp. 92 355–92 364, 2019.

[10] M. Burduja, R. T. Ionescu, and N. Verga, "Accurate and efficient intracranial hemorrhage detection and subtype classification in 3d ct scans," *Sensors*, vol. 20, no. 19, p. 5611, 2020. DOI: 10.3390/S20195611.

[11] K. Hu et al., "Automatic segmentation of intracerebral hemorrhage in ct images using encoder–decoder convolutional neural network," *Information Processing and Management*, vol. 57, no. 6, p. 102 352, 2020. DOI: 10.1016/j.ipm.2020.102352.

[12] J. Y. Lee, J. S. Kim, T. Y. Kim, and Y. S. Kim, "Detection and classification of intracranial haemorrhage on ct images using a novel deep-learning algorithm," *Scientific Reports*, vol. 10, no. 1, p. 20 546, 2020. DOI: 10.1038/s41598-020-77441-z.

[13] M. Monteiro et al., "Multiclass semantic segmentation and quantification of traumatic brain injury lesions on head ct using deep learning," *The Lancet Digital Health*, vol. 2, no. 6, e314–e322, 2020.

[14] N. T. Nguyen, D. Q. Tran, N. T. Nguyen, and B. Q. Nguyen, "A cnn-lstm architecture for detection of intracranial hemorrhage on ct scans," *arXiv preprint arXiv:2005.10992*, 2020.

[15] M. Shanmuganathan et al., "Cerebleed: Automated segmentation and severity index computation for intracerebral hemorrhage in ct images," *Scientific Reports*, 2020. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC7658678/.

[16] M. F. Mushtaq, M. Shahroz, A. M. Aseere, et al., "Bhcnet: Neural network-based brain hemorrhage classification using head ct scan," *IEEE Access*, vol. 9, pp. 113 901–113 916, 2021. DOI: 10.1109/ACCESS.2021.3102740.

[17] X. Wang et al., "A deep learning algorithm for automatic detection and classification of acute intracranial hemorrhages in head ct scans," *NeuroImage: Clinical*, vol. 32, p. 102 785, 2021. DOI: 10.1016/j.nicl.2021.102785.

[18] X. Zhao, K. Chen, G. Wu, et al., "Deep learning shows good reliability for automatic segmentation and volume measurement of brain hemorrhage," *European Radiology*, vol. 31, no. 7, pp. 5012–5020, 2021. DOI: 10.1007/s00330-020-07558-2.

[19] S. Heo, J. Ha, W. Jung, et al., "Decision effect of a deep-learning model to assist a head computed tomography order for pediatric traumatic brain injury," *Scientific Reports*, vol. 12, no. 1, 2022. DOI: 10.1038/s41598-022-16313-0.

[20] V. Anand and P. Shourie, "Hemorrhage classification in head computed tomography images using convolution neural network model," in *Proceedings of IEEE Conference*, 2023, pp. 1–5. DOI: 10.1109/cisct57197.2023.10351262.

[21] S. E. Arman et al., "Intracranial hemorrhage classification from ct scan using deep learning and bayesian optimization," *IEEE Access*, vol. 11, pp. 83 446–83 460, 2023. DOI: 10.1109/ACCESS.2023.3300771.

[22] A. Ghosh, D. Chanda, K. Bhattacharyya, and S. Sarkar, "Intracranial hemorrhage detection and classification of its subtypes using cnn, resnet, random forest and densenet201," *Research Square Preprint*, 2023. DOI: 10.21203/rs.3.rs-3146153/v1.

[23] S. H. Ramananda and V. Sundaresan, "Class activation map-based weakly supervised hemorrhage segmentation using resnet-lstm," *arXiv preprint*, 2023. [Online]. Available: http://arxiv.org/abs/2309.16627.

[24] K.-C. Teo, S.-M. Fong, W. C. Y. Leung, et al., "Location-specific hematoma volume cutoff and clinical outcomes in intracerebral hemorrhage," *Stroke*, vol. 54, no. 6, pp. 1548–1557, 2023. DOI: 10.1161/STROKEAHA.122.041246.

[25] V. Vidhya, U. Raghavendra, A. Gudigar, et al., "Yolov5s-cam: A deep learning model for automated detection and classification of intracranial hematoma types in ct images," *IEEE Access*, 2023. DOI: 10.1109/ACCESS.2023.3339560.

[26] J. Fan, G. Lu, and X. Fan, "Dual-task vision transformer for rapid and accurate intracerebral hemorrhage classification on ct images," *arXiv preprint*, 2024. DOI: 10.48550/arxiv.2405.06814.

[27] B. He, Z. Xu, D. Zhou, and L. Zhang, "Deep multiscale convolutional feature learning for intracranial hemorrhage classification and weakly supervised localization," *Heliyon*, 2024. DOI: 10.1016/j.heliyon.2024.e30270.

[28] N. K. Kar, S. Jana, A. Rahman, et al., "Automated intracranial hemorrhage detection using deep learning in medical image analysis," in *Proceedings of IEEE Conference*, 2024, pp. 1–6. DOI: 10.1109/icdsns62112.2024.10691276.

[29] R. Karamala, N. V. Vyshnavi, N. R. S. Chowdary, T. Singh, and P. Patra, "Deep learning for precise brain hemorrhage detection from ct scans: A web-based approach," in *Proceedings of IEEE Conference*, 2024, pp. 1–6. DOI: 10.1109/icccnt61001.2024.10724023.

[30] G. Li, K. Gao, C. Liu, and S. Li, "Intracranial hematoma segmentation on head ct based on multiscale convolutional neural network and transformer," *IET Image Processing*, 2024. DOI: 10.1049/ipr2.13188.

[31] J. Teneggi, P. H. Yi, and J. Sulam, "Examination-level supervision for deep learning-based intracranial hemorrhage detection," *Radiology: Artificial Intelligence*, vol. 6, no. 1, e230159, 2024. DOI: 10.1148/ryai.230159.

[32] Y. Wu, M. Iorga, S. Badhe, et al., "Precise image-level localization of intracranial hemorrhage on head ct scans with deep learning models trained on study-level labels," *Radiology*, 2024. DOI: 10.1148/ryai.230296.

[33] S. N. Ahmed and P. Prakasam, "Intracranial hemorrhage segmentation and classification framework using deep learning," *Scientific Reports*, vol. 15, no. 1, p. 17 151, 2025. DOI: 10.1038/s41598-025-01317-3.

[34] A. Chaudhary, Y. Gaur, A. Abraham, and H. Singh, "Multiclass intracranial hemorrhage detection and confidence aware triage via deep ensemble learning," *IEEE Access*, vol. 13, pp. 191 745–191 761, 2025. [Online]. Available: https://ieeexplore.ieee.org/document/11218836.

[35] M. Kim, M. Lee, D. Kim, W.-S. Ryu, and H.-G. Jeong, "Deep learning-assisted detection of intracranial hemorrhage: Validation and impact on reader performance," *Neuroradiology*, 2025. DOI: 10.1007/s00234-025-03560-x.

[36] A. Kokkula and C. P. Sekhar, "Brain haemorrhage analysis using deep learning," *ICTACT Journal on Soft Computing*, vol. 16, no. 2, pp. 3906–3913, 2025. DOI: 10.21917/ijsc.2025.0541.

[37] Y. Ning, Q. Yu, X. Fan, et al., "Noncontrast ct-based deep learning for predicting intracerebral hemorrhage expansion," *Scientific Reports*, vol. 15, no. 1, 2025. DOI: 10.1038/s41598-025-17393-4.

[38] J. Takala, H. Peura, R. Pirinen, et al., "High sensitivity in spontaneous intracranial hemorrhage detection from emergency head ct scans using ensemble learning," *Scientific Reports*, vol. 15, no. 1, p. 29 919, 2025. DOI: 10.1038/s41598-025-15835-7.