# A multi-window convolutional neural network (CNN) study for automatic cerebral hemorrhage staging

Yixiao Zhang Youlong Jiao

April 16, 2025

**Abstract**

*Intracranial hemorrhage (ICH) is a life-threatening condition that requires quick and accurate diagnosis to improve patient care. In this project, we explore a convolutional neural network (CNN)-based approach to automatically classify different types of cerebral hemorrhages using head CT scans. By combining multi-window CT views and validating the model through a step-by-step process, we assess its training feasibility, ability to avoid overfitting, and overall performance. Our best model, trained on 20% of the dataset with tuned dropout, reached approximately 50% accuracy in a six-class classification task. These results highlight the potential of CNNs to support clinical decision-making and offer a foundation for future improvements.*

## 1 Introduction

Intracranial hemorrhage (ICH) refers to bleeding events within the skull and is classified into several major types:

- **Subdural**
- **Subarachnoid**
- **Intraventricular**
- **Intraparenchymal**
- **Epidural**

Each hemorrhage type has distinct clinical significance and requires a specific intervention strategy. Misclassifications could delay appropriate treatment and increase patient risk. Early and accurate identification of the hemorrhage type not only guides immediate medical decisions but also reduces the likelihood of long-term neurological complications.

Computed tomography (CT) imaging is the most widely used diagnostic tool for detecting ICH due to its speed and sensitivity in visualizing acute bleeding. Despite its strengths, CT interpretation remains a manual process that is time-consuming, requires years of radiological training, and can introduce inter-observer variability or diagnostic errors, especially in high-pressure emergency settings.

## 2 Related work

**Convolutional Neural Networks** (CNNs) are widely used in medical image analysis tasks, ranging from brain tumor detection to pulmonary and dermatological diagnostics. Prior research on intracranial hemorrhage (ICH) detection typically focuses on binary classification (e.g., hemorrhage vs. non-hemorrhage) and often uses only a single CT window as input.

In contrast, our study incorporates *multi-window* CT slices—such as brain window, bone window, and contrast-enhanced views—which provide complementary anatomical and radiological information. This allows for more robust feature learning and enables the model to address a multi-class classification problem, distinguishing between several ICH subtypes.

A standard CNN model architecture consists of the following key components:

- **Convolution Layers:** Extract local spatial patterns such as edges, gradients, and textures.
- **Pooling Layers:** Downsample feature maps to reduce dimensionality and computation while preserving key structures.
- **Flattening and Fully Connected Layers:** Transform spatial features into class predictions through dense connections.

The convolutional neural network (CNN) processes the input CT image through a sequence of convolutional layers, each designed to extract increasingly abstract and spatially meaningful features such as edges, textures, and shapes. These convolutional layers are interleaved with pooling layers, which downsample the feature maps to reduce computational complexity and enhance translational invariance, thereby enabling the network to generalize better across varying patient scans and anatomical differences.

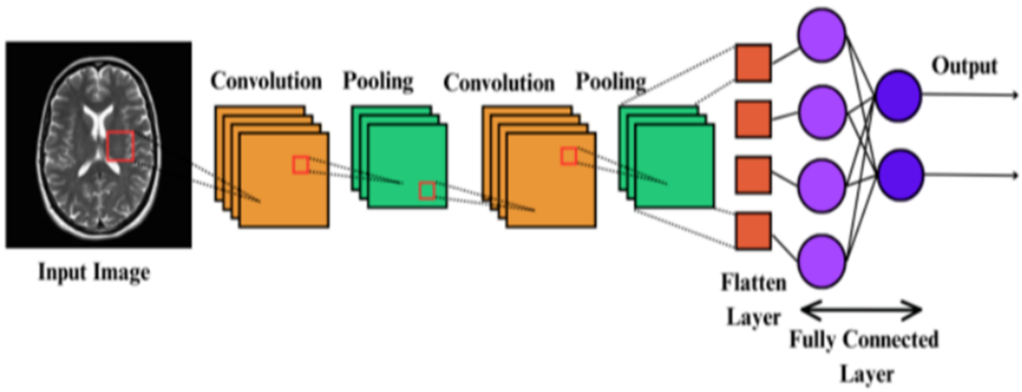


Figure 1: CNN architecture used in our cerebral hemorrhage classification pipeline.

# 3 Dataset Preprocessing

## 3.1 Data Composition

The dataset used in this study comprises over 100,000 labeled cranial CT images. These images are annotated across six types of intracranial hemorrhage—namely, subdural, subarachnoid, intraventricular, intraparenchymal, epidural—as well as a control class labeled as "normal." Each instance corresponds to a single CT slice derived from real-world radiological examinations.

To enrich diagnostic information, each CT slice was rendered into four distinct window settings:

- **Brain window:** Enhances soft tissue contrast, making brain parenchyma structures more visible.
- **Bone window:** Emphasizes high-density structures such as the skull, aiding in trauma assessment.
- **Max contrast window:** Highlights regions of contrast uptake, useful for detecting blood pooling or vascular leakage.
- **Hemorrhage-specific window:** Optimized to accentuate acute bleeding characteristics.

To effectively utilize these diverse imaging perspectives, we developed a custom CNN architecture that accepts multi-window input as a 4-channel tensor. By stacking the four CT windows along the channel dimension, the network is exposed to complementary radiological features, allowing it to learn richer spatial representations. This preprocessing pipeline ensures consistent input formatting, enhances the diversity of visual features, and helps mitigate overfitting by leveraging anatomical variance across window types.

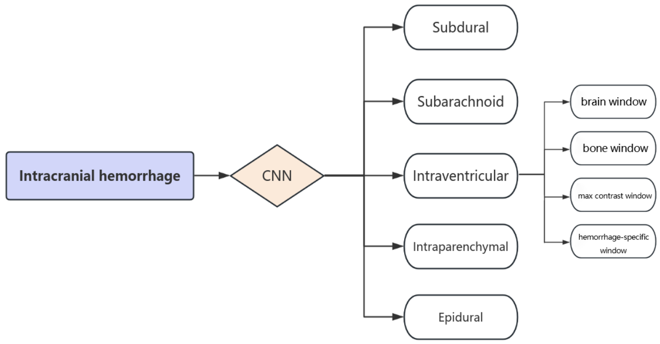


Figure 2: Illustration of the CNN-based classification workflow

### 3.2 Preprocessing Steps

To prepare the data for deep learning, the following preprocessing pipeline was applied to all CT images:

- **Resize:** Each image was resized to a fixed spatial resolution of $224 \times 224$ pixels to ensure compatibility with standard CNN input layers and reduce computational load.
- **Normalization:** Pixel intensity values were scaled to the $[0, 1]$ range to facilitate faster convergence during training and to prevent numerical instability.
- **Data Augmentation:** A random horizontal flip was applied with a 50% probability to simulate anatomical variance and improve model robustness.
- **Multi-channel stacking:** The four CT window images corresponding to each sample were stacked along the channel axis, creating a 4-channel tensor for each patient slice. This structure mimics the RGB input format commonly used in CNNs but with modality-specific radiological data.
- **Train-test split:** The processed dataset was partitioned into an 80% training set and a 20% testing set to enable model evaluation on unseen data.

## 4 Results

### 4.1 Phase 1: Feasibility Test (Subdural vs. Normal)

**Objective:** To validate the overall training pipeline and model capacity using a simplified binary classification task.

We initially trained SimpleCNN on a binary classification problem using only subdural hemorrhage and normal samples. The training curve (see Figure 3) shows that training loss steadily decreased over epochs, while test loss fluctuated noticeably. The accuracy plot highlights a gap between training and test performance, with training accuracy rising consistently while test accuracy remained relatively flat.

These trends suggest that while the model was able to learn from training data, it failed to generalize well—an indication of early overfitting. The results prompted the introduction of regularization strategies.

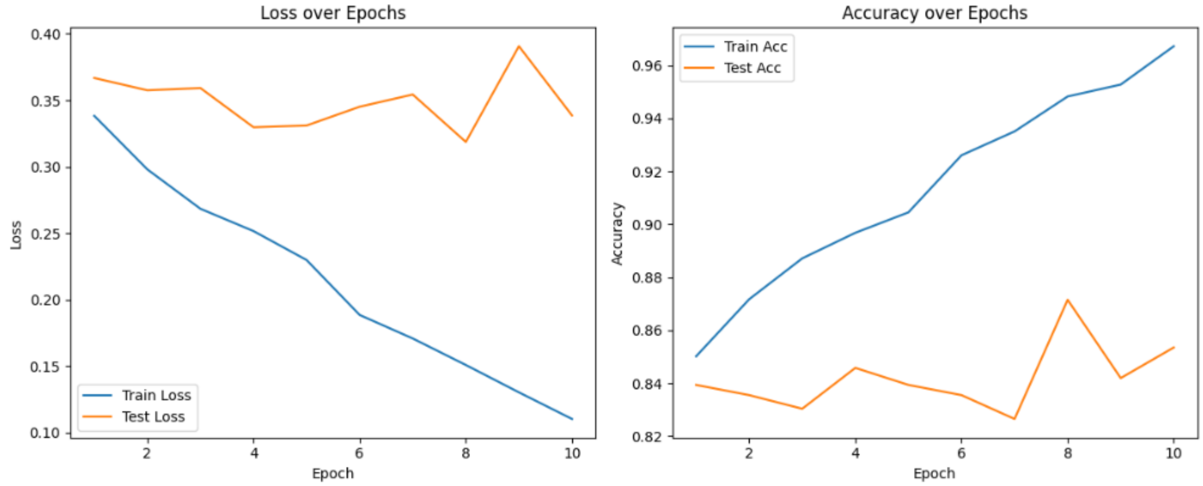


Figure 3: Phase 1: Training and validation loss/accuracy for binary classification (Subdural vs. Normal). Overfitting is evident.

### 4.2 Phase 2: Dropout Adjustment

**Objective:** To mitigate overfitting using dropout regularization.

A dropout layer with probability $p = 0.5$ was added before the final fully connected layer. As shown in Figure 4, the gap between training and test loss decreased, and the accuracy curves aligned more closely than in Phase 1. Test accuracy also showed less variance, indicating improved generalization.

This confirmed the effectiveness of dropout in regularizing the model and reducing overfitting, even on small datasets.

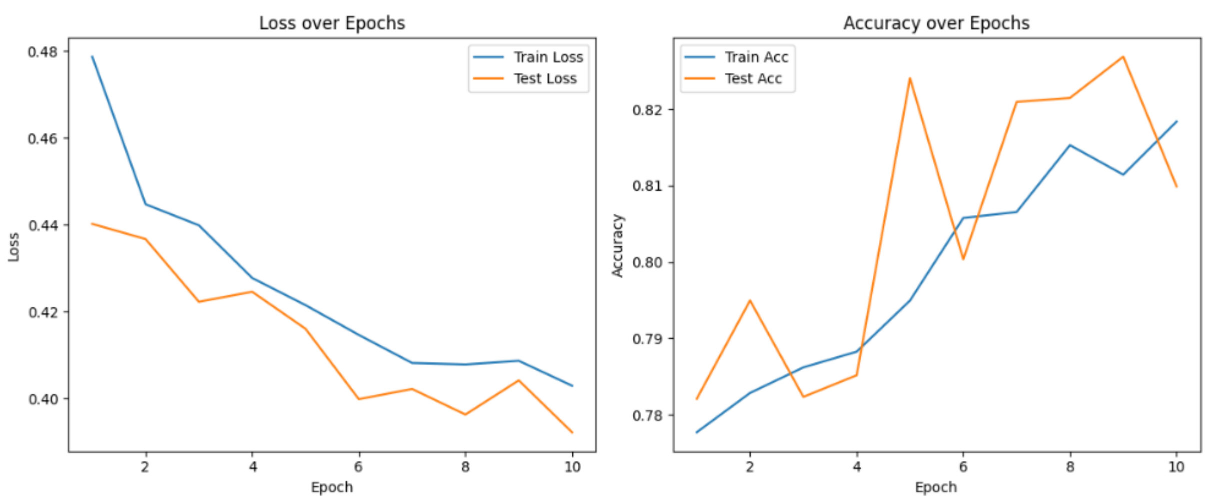


Figure 4: Phase 2: Dropout applied ($p = 0.5$) to reduce overfitting. Accuracy stabilized and test loss declined more consistently.

### 4.3 Phase 3: Multi-Type Classification (10% Sample)

**Objective:** To evaluate CNN performance on a six-class classification task using a small sample size.

The dataset was expanded to include all six hemorrhage types and the normal class. Inputs were formatted as 4-channel tensors (brain, bone, max contrast, and hemorrhage-specific windows). Training was conducted on a 10% sample of the full dataset.

As illustrated in Figure 5, both training and test losses declined gradually, and the accuracy curves remained closely aligned. However, overall accuracy plateaued around 40–41%, indicating that the model was underfitting—possibly due to insufficient training data or model depth.

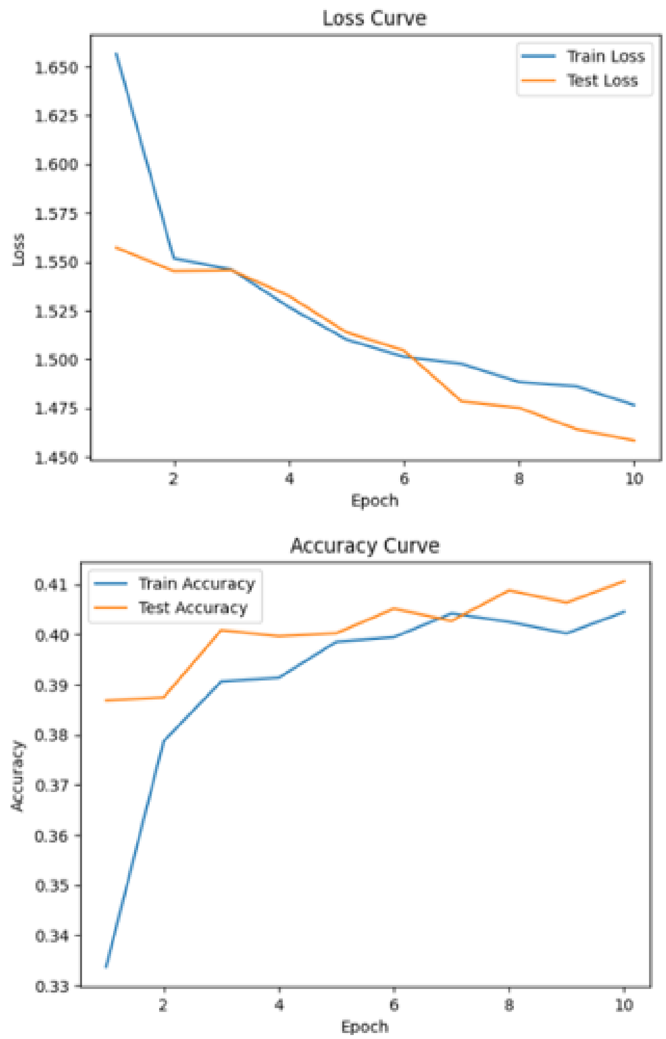


Figure 5: Phase 3: Multi-class classification on a 10% dataset sample. Accuracy stabilized but remained low, suggesting underfitting.

### 4.4 Phase 4: Optimization (20% Sample + Dropout = 0.25)

**Objective:** To improve accuracy and generalization by increasing data volume and tuning regularization strength.

In this phase, training data volume was doubled (20% sample), and dropout was reduced to $p = 0.25$ to balance learning flexibility and regularization. The training and test loss curves in Figure 6 show consistent convergence to around 1.28, and test accuracy rose to approximately 0.50.

The results demonstrate improved generalization and reduced underfitting, confirming the effectiveness of data scaling and dropout optimization.

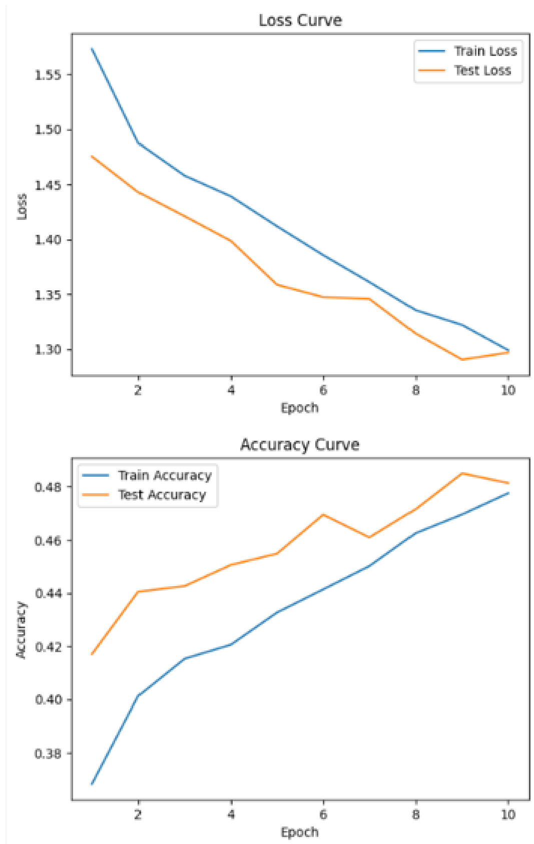


Figure 6: Phase 4: Model trained with 20% dataset and dropout $p = 0.25$. Accuracy improved to ∼50%, and loss stabilized.

# 5 Discussion

## 5.1 Observations

Our experimental results provide several key insights into the model's learning behavior and generalization performance:

- **Increased training data improves model learning capacity.** As shown in Phase 4, expanding the dataset from 10% to 20% led to a noticeable improvement in accuracy and smoother convergence in both loss and accuracy curves. This suggests that the model benefits significantly from data diversity, which enhances its ability to capture relevant features across hemorrhage subtypes.
- **Dropout is effective in mitigating overfitting.** Introducing dropout layers—particularly with $p = 0.5$ in early experiments—reduced the gap between training and test performance, leading to improved stability in test accuracy. This regularization strategy successfully prevented the network from memorizing training patterns too early.
- **Training and test curves show consistent behavior, indicating model stability.** Unlike typical signs of overfitting (where test loss diverges from training loss), our learning curves remained closely aligned across all phases. This consistency implies that the model was not overfitting, but rather underfitting due to insufficient capacity or data.
- **Overall accuracy remains limited, suggesting mild underfitting.** Despite improved convergence, the best model accuracy plateaued around 50%, which—while better than random (16.7%)—is still below clinical usability thresholds. The consistent performance gap suggests a need for deeper architectures or stronger representations.

### 5.2 Limitations

While the project establishes the feasibility of CNN-based hemorrhage classification, several constraints limit the current model's performance:

- **Limited training epochs.** All experiments were conducted using a fixed 10-epoch training schedule. Given that training loss was still decreasing at epoch 10, extending training to 30 or more epochs may lead to better convergence and higher accuracy.
- **Absence of pretrained CNN backbones.** The use of custom CNNs was intentional for transparency and interpretability. However, pretrained models such as ResNet or DenseNet have demonstrated superior feature extraction capabilities in medical image analysis and may yield better performance if integrated via transfer learning.
- **Class imbalance in the dataset.** Some hemorrhage types had significantly fewer examples compared to others, which may have biased the model toward more frequent classes. This imbalance can skew learning and reduce sensitivity for rare subtypes.
- **Lack of advanced augmentation.** The only augmentation technique used was random horizontal flipping. Incorporating additional transformations—such as rotation, brightness/contrast shifting, elastic deformation, or intensity scaling—could help the model generalize to variations found in real-world clinical imaging.

## 6 Conclusion

This study demonstrates the feasibility of using convolutional neural networks (CNNs) for automatic classification of cerebral hemorrhage subtypes from head CT scans. By leveraging a multi-window input strategy, we enabled the model to capture complementary radiological features that are often critical in clinical interpretation.

Through phased experimentation, we observed that increasing training data volume and introducing regularization (e.g., dropout) significantly improved model generalization. Our optimized model, trained on a 20% sample with reduced dropout, achieved a test accuracy of approximately 50%, a notable improvement over the earlier configurations.

Despite these gains, the model remains limited by underfitting and modest accuracy, highlighting the need for future enhancements. These include longer training durations, use of pretrained backbones like ResNet, class balancing strategies, and advanced augmentation techniques. With these improvements, the proposed pipeline has the potential to support real-time, high-throughput clinical decision-making in emergency radiology settings.

## References

[1] M. R. Arbabshirani, B. K. Fornwalt, G. J. Mongelluzzo, *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, p. 9, 2018. [Online]. Available: https://www.nature.com/articles/s41746-017-0015-z

[2] Q. Dou, H. Chen, L. Yu, J. Qin, and P.-A. Heng, "Automatic detection of cerebral microbleeds from MR images via 3D convolutional neural networks," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1182–1195, May 2016. [Online]. Available: https://doi.org/10.1109/TMI.2016.2528129

[3] S. Chilamkurthy, R. Ghosh, S. Tanamala, *et al.*, "Development and validation of deep learning algorithms for detection of critical findings in head CT scans," *arXiv preprint arXiv:1803.05854*, 2018. [Online]. Available: https://arxiv.org/abs/1803.05854

[4] "Enhancing brain tumor diagnosis using CNN models: A comparative analysis," *Scientific Figure on ResearchGate*, [Accessed: Apr. 16, 2025]. [Online]. Available: https://www.researchgate.net/figure/General-Convolutional-Neural-Network-CNN-architecture-for-classifying-brain-MRI-CT_fig3_377312701