Sep 09, 2024

Deep-learning tool for early identification of non-traumatic intracranial hemorrhage etiology based on CT scan

DOI
dx.doi.org/10.17504/protocols.io.yxmvmejoog3p/v1

Meng Zhao[1]

[1]Beijing Tiantan Hospital

Meng Zhao
Beijing Tiantan Hospital

OPEN ACCESS

DOI: **dx.doi.org/10.17504/protocols.io.yxmvmejoog3p/v1**

Protocol Citation: Meng Zhao 2024. Deep-learning tool for early identification of non-traumatic intracranial hemorrhage etiology based on CT scan. **protocols.io** **https://dx.doi.org/10.17504/protocols.io.yxmvmejoog3p/v1**

Protocol status: Working
We use this protocol and it's working

Created: September 08, 2024

Last Modified: September 09, 2024

Protocol Integer ID: 107136

Abstract

Objective

To develop an
artificial intelligence system that can accurately identify acute non-traumatic intracranial hemorrhage (ICH) etiology based on non-contrast CT (NCCT) scans and investigate whether clinicians can benefit from it in a diagnostic setting.

Methods

The deep learning model
was developed with 1868 eligible NCCT scans with non-traumatic ICH collected between January 2011 and April 2018. We tested the model on two independent datasets (TT200 and SD 98) collected after April 2018. The model's diagnostic performance was compared with clinicians' performance. We further designed a simulated study to compare the clinicians' performance with and without the deep learning system augmentation.

Results

The proposed deep
learning system achieved area under the receiver operating curve of
0.986 (95% CI 0.967–1.000) on aneurysms, 0.952 (0.917–0.987) on hypertensive hemorrhage, 0.950 (0.860–1.000) on arteriovenous malformation (AVM), 0.749 (0.586–0.912) on Moyamoya disease (MMD), 0.837 (0.704–0.969) on cavernous malformation (CM), and 0.839 (0.722–0.959) on other causes in TT200 dataset. Given a 90% specificity level, the sensitivities of our model were 97.1% and 90.9% for aneurysm and AVM diagnosis, respectively. The model also shows an impressive generalizability in an independent dataset SD98. The clinicians achieve significant improvements in the sensitivity, specificity, and accuracy of diagnoses of certain hemorrhage etiologies with proposed system augmentation.

Conclusions

The proposed deep learning algorithms can be
an effective tool for early identification of hemorrhage etiologies based on NCCT scans. It may also provide more information for
clinicians for triage and further imaging examination selection.

1 Dataset and Clinical Taxonomy Retrospectively review NCCT scans from 4019 patients at Beijing Tiantan Hospital (2011-2018). Include only the first NCCT scan post-ICH event; exclude traumatic, post-surgical, or delayed scans.

2 Inclusion and Exclusion Criteria Exclude scans with poor quality, lacking a complete series, or from patients under four years. Review medical records for patient demographics and clinical history.

3 Independent Test Datasets Collection Collect additional datasets from Beijing Tiantan Hospital (TT200) and Shandong Jining Medical College (SD98) from 2018 to 2019.

4 Ethical Compliance Obtain IRB approval from both institutions. Anonymize all clinical data and images. Waive informed consent based on non-harm justification.

5 Data Labeling and Etiologic Classification Use Meretoja et al.'s method for etiologic classification. Label data independently by two radiologists, with discrepancies resolved by a third.

6 Model Development: Image Preprocessing Rotate images every 20 degrees using a bilinear algorithm. Resize volumes to a consistent voxel size. Apply unsupervised skull-stripping and crop the images.

7 Model Architecture Develop the ICHNet using the "SlowFast Networks" architecture. Use separate pathways for different resolution features, combining them before the output layer.

8 Training Protocol Use 4 NVIDIA Tesla P40 GPUs and Pytorch-0.4.1. Implement min-max normalization and oversampling to address class imbalance. Perform five-fold cross-validation to select the best model.

9 Test Procedure Test the model using an ensemble strategy with data from both test datasets. Decrease rotation frequency during testing to reduce processing time.

10 Statistical Analysis Generate confusion matrices and calculate AUCs. Use the Clopper-Pearson and Hanley and McNeil methods for confidence intervals. Evaluate rater concordance and assess model impact using Cohen and Fleiss κ coefficients. Apply bootstrapping for performance metrics and Wilcoxon rank-sum tests for statistical significance.