A Deep Learning Model for Classifying Spontaneous Intracranial Hypotension on Brain MRI

Purpose

Spontaneous intracranial hypotension (SIH) is an important, but underrecognized cause of positional headache. SIH is a debilitating but treatable condition. Characteristic brain MRI findings of smooth dural thickening/hyperenhancement, venous distention, and brain sag are key to this diagnosis. Thus, we sought to develop a deep learning model to assist radiologists in identifying potential cases of SIH on brain MRI examinations.

Methods & Materials

In total 428 brain MRI examinations from unique subjects were included (155 positive for SIH and 273 normals). Positive cases were selected from an institutional database of confirmed SIH cases. Normal controls were identified by screening consecutive, contemporary brain MRI reports with a natural language processing tool. Reports flagged as likely normal were manually reviewed and confirmed. Data were split at the subject level into training (45%), validation (30%), and test (25%) sets. Axial and coronal post-contrast, T1-weighted sequences were selected for further analysis. An empirically derived heuristic was applied to select 1 axial image and 2 coronal images from each study most likely to exemplify the characteristic findings of SIH. Transfer learning was employed with ResNet-50 architecture pre-trained on ImageNet, adapted to train 3 separate models for the axial and 2 coronal images. Models were fine-tuned for 50 epochs each. Predictions from the 3 models were ensembled with a logistic regression model to generate a final prediction. Performance was evaluated with receiver operating characteristic (ROC) and precision-recall (PR) curves, including area under the ROC (AUROC).

Results

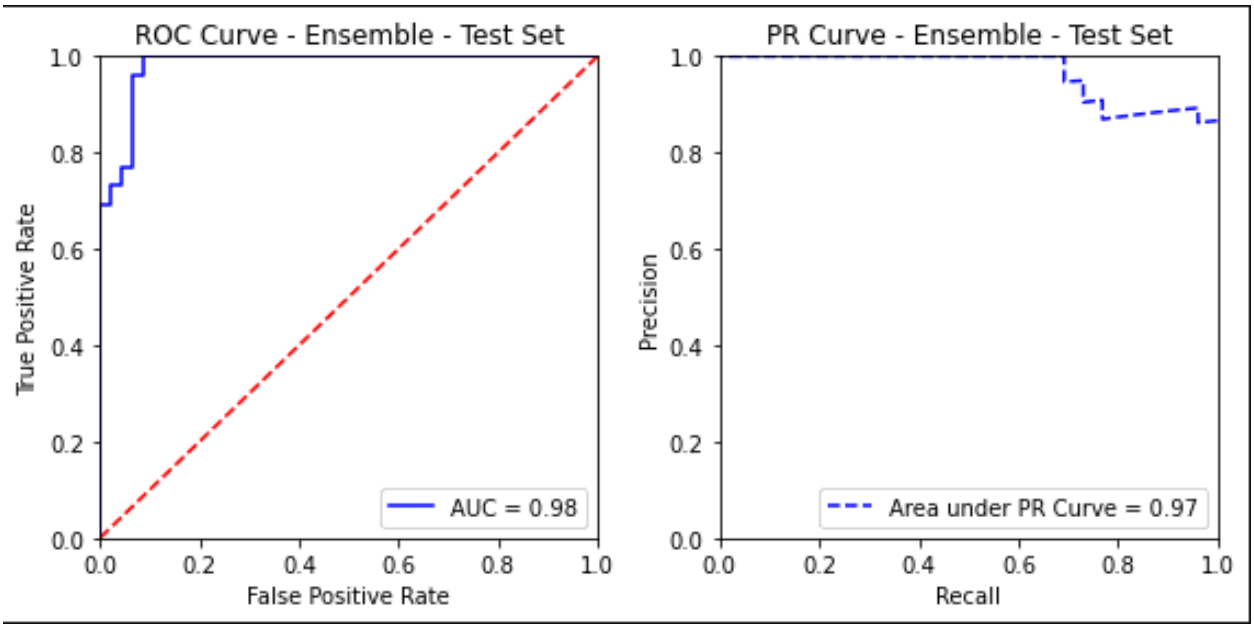
AUROC for axial and coronal models were 0.98, 0.98, and 0.97. Ensemble AUROC was 0.98. Given the surprisingly high accuracy of our results, subsequent testing and experiments were performed for validation. There was no evidence of data leakage. Running model inference on the test data set without fine-tuning the pre-trained model resulted in AUROC 0.55. Fine-tuning of the model on a small subset of the data - 10 positive and 20 negative cases from the training set, 5 positive and 10 negative from the validation set - resulted in a slight improvement with AUROC 0.61. Repeating training with the full dataset reproduced the result of AUROC 0.98.

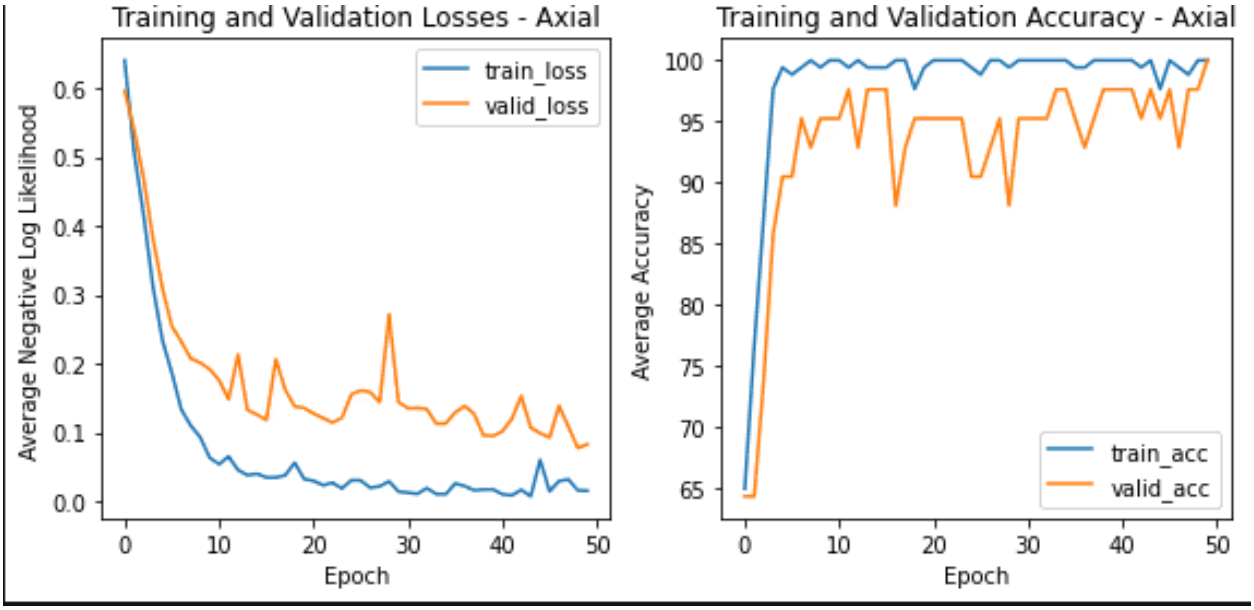
Conclusions

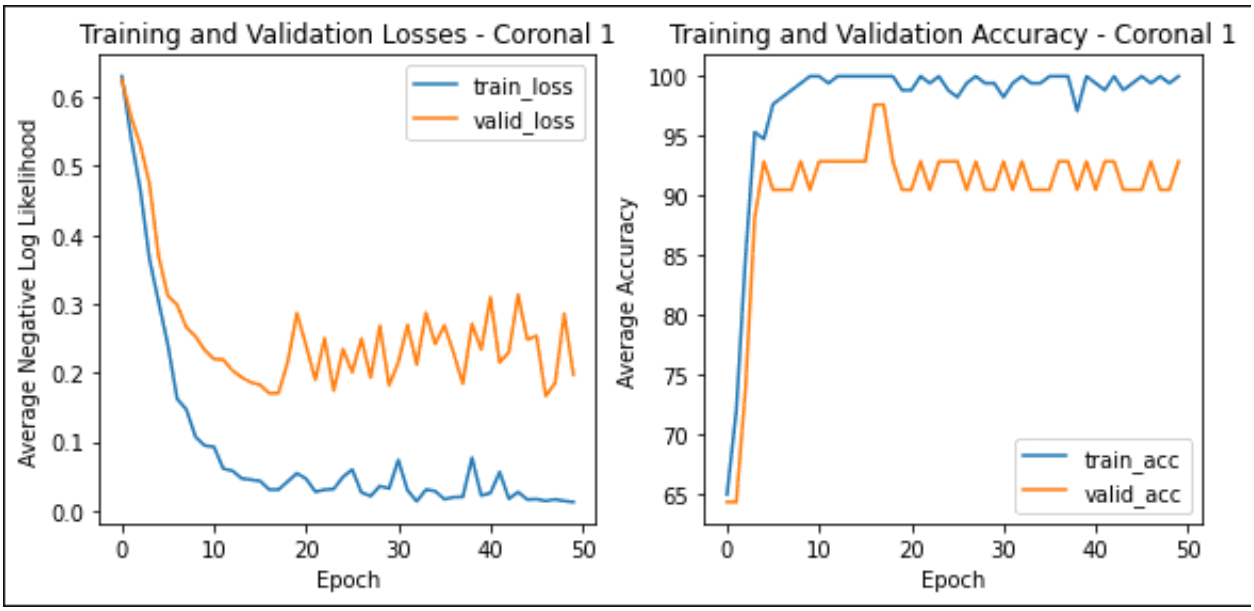
We successfully trained a highly accurate deep learning model capable of classifying SIH on brain MRI examinations from our institutional database.

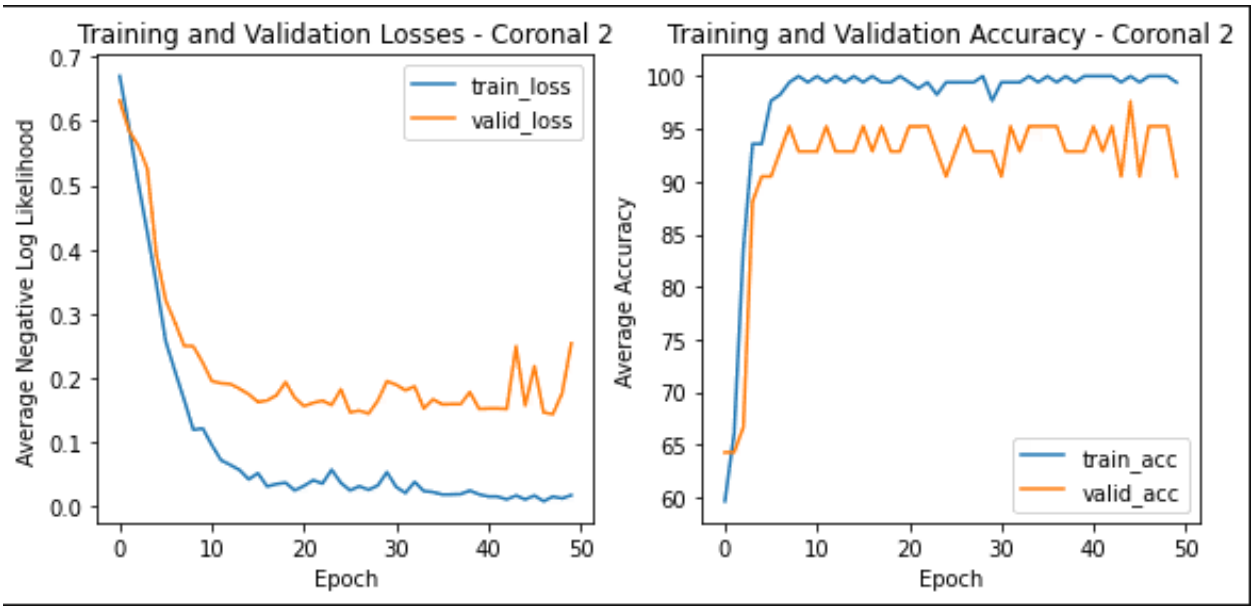
Clinical Relevance/Application

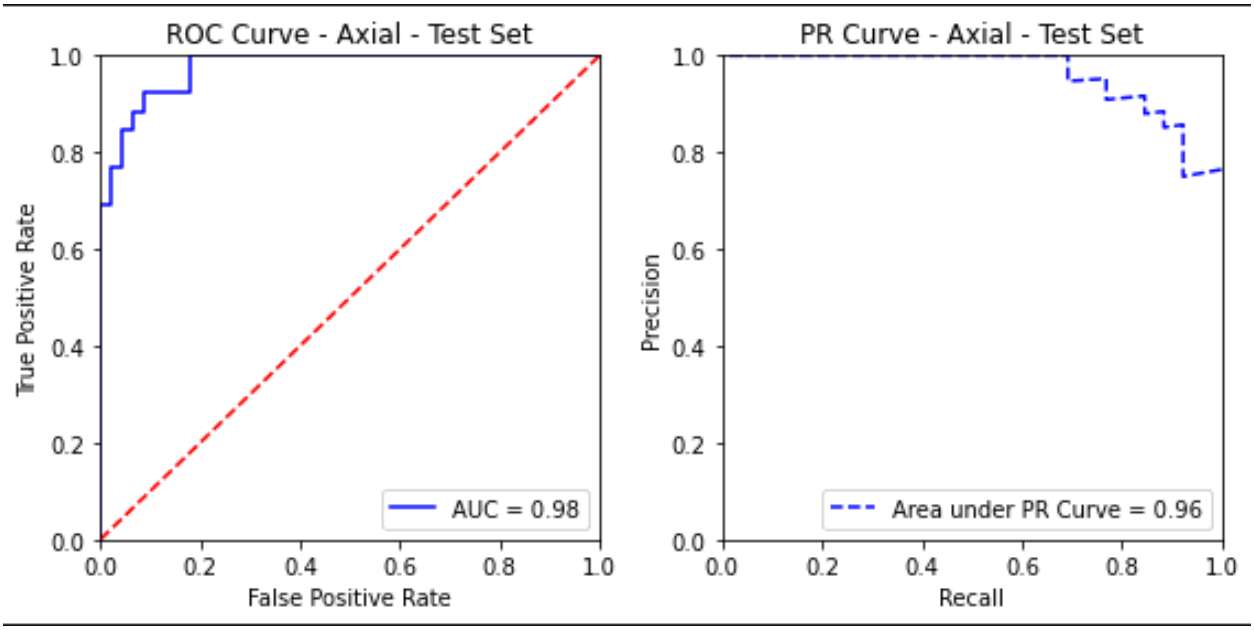
If shown to be generalizable to more challenging cases and/or cases from other institutions, this model could help radiologists identify possible cases of SIH on brain MRI examinations.

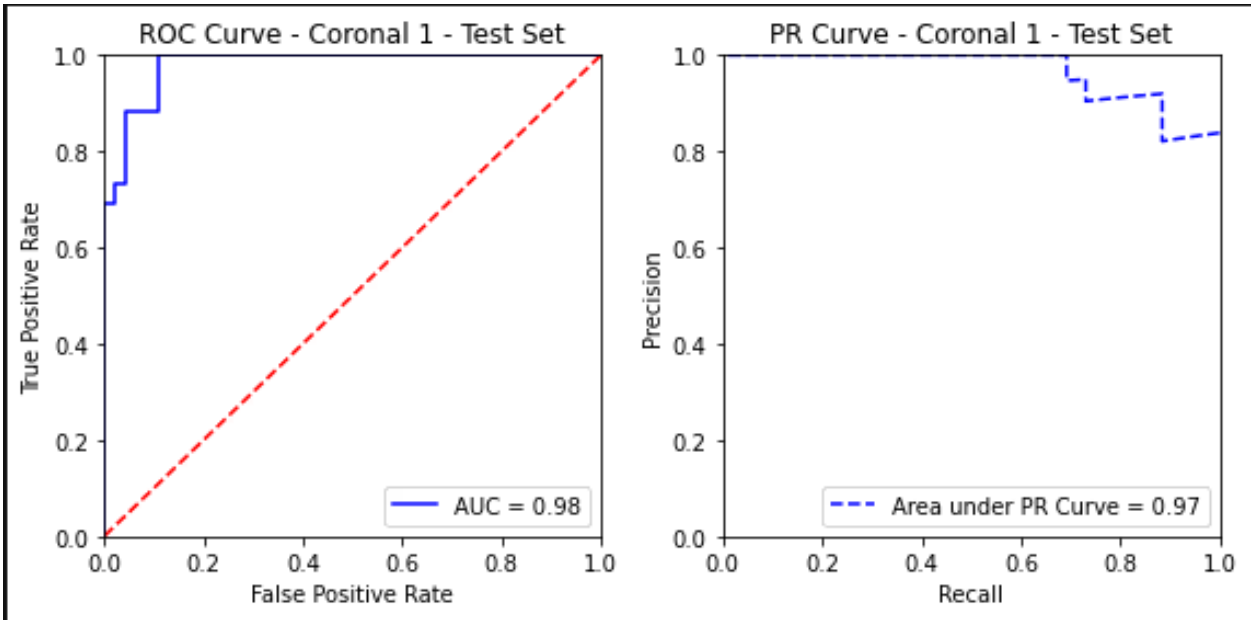


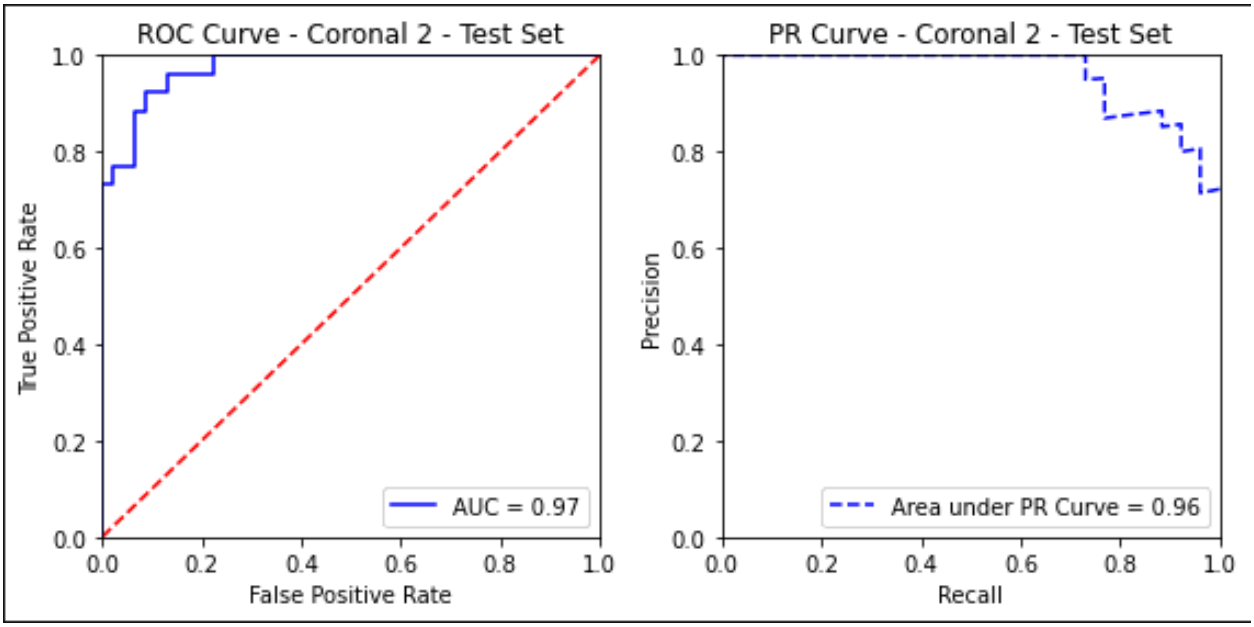

Training and Validation Losses - Axial
Average Negative Log Likelihood
Epoch
train_loss
valid_loss
Training and Validation Accuracy - Axial
Average Accuracy
Epoch
train_acc
valid_acc


Training and Validation Losses - Coronal 1
Average Negative Log Likelihood
Epoch
train_loss
valid_loss
Training and Validation Accuracy - Coronal 1
Average Accuracy
Epoch
train_acc
valid_acc


Training and Validation Losses - Coronal 2
Average Negative Log Likelihood
Epoch
train_loss
valid_loss
Training and Validation Accuracy - Coronal 2
Average Accuracy
Epoch
train_acc
valid_acc


ROC Curve - Axial - Test Set
True Positive Rate
False Positive Rate
AUC = 0.98
PR Curve - Axial - Test Set
Precision
Recall
Area under PR Curve = 0.96


ROC Curve - Coronal 1 - Test Set
True Positive Rate
False Positive Rate
AUC = 0.98
PR Curve - Coronal 1 - Test Set
Precision
Recall
Area under PR Curve = 0.97


ROC Curve - Coronal 2 - Test Set
True Positive Rate
False Positive Rate
AUC = 0.97
PR Curve - Coronal 2 - Test Set
Precision
Recall
Area under PR Curve = 0.96