Identification of Abnormalities in Head Computerized Tomography Scans

M. DelRocini, C. Angelini and G. Rasool

Department of Electrical and Computer Engineering, Rowan University, Glassboro, USA
{delroc72, angelinic0}@students.rowan.edu, rasool@rowan.edu

Stroke and traumatic brain injuries are both leading causes of death and long-term disability globally [1]. Early detection of abnormalities in head computerized tomography (CT) scans reduces patient risk of serious medical complications from brain injuries such as hemorrhagic stroke and cranial fractures [2]. A machine learning algorithm trained to autonomously identify and classify head CT scan irregularities has the potential to decrease detection time of such anomalies and allow for quicker, more effective treatment.

There has been a growing number of machine learning algorithms developed for CT scan classification using deep neural networks [3-7]. The individual slices of a brain CT scan can be considered frames (or images) of a movie and modern machine learning techniques of image and time-series processing can be combined for the joint analysis of the whole CT scan. A single end-to-end neural network consisting of a 2D convolutional neural network (CNN) and a recurrent neural network (RNN) is proposed in this study. The former network processes individual slices while the latter combines information from multiple slices to identify and classify 14 different types of brain injuries. Two types of RNNs were considered: long-short term memory (LSTM) and gated recurrent unit (GRU). LSTMs and GRUs use multiple gates in their architecture to avoid gradient vanishing and exploding problems faced by simple RNNs.

The *CQ500* dataset was used for training all networks [6]. The *CQ500* dataset is a set of CT scans collected by the Centre for Advanced Research in Imaging, Neurosciences and Genomics, New Delhi, India. Each CT scan in the *CQ500* dataset was independently labeled by three senior radiologists for the presence of any of the following 14 types of brain trauma: intracranial hemorrhage (ICH), any of the ICH types (intraparenchymal [IPH], intraventricular [IVH], subdural [SDH], extradural [EDH], and subarachnoid [SAH]), midline shift, mass effect, bleed location (right and left), chronic hemorrhage, fracture, calvarial fracture, and other fracture. Table 1 shows the percentage of scans in the *CQ500* dataset with each type of brain trauma. This dataset was also used in the classification efforts of Qure.ai.

Table 1. Percent of scans containing each type of brain trauma

Brain Trauma Type	% of Scans
ICH	41.8%
IPH	27.3%
IVH	5.7%
SDH	10.8%
EDH	2.7%
SAH	12.2%
Bleed Location Left	26.7%
Bleed Location Right	26.3%
Chronic Bleed	3.9%
Fracture	8.0%
Calvarial Fracture	6.9%
Other Fracture	2.4%
Mass Effect	26.9%
Midline Shift	13.3%

CT scans were minimally processed for training. A non-contrast axial scan was selected for each patient and resampled to a slice thickness of 5 mm. In the *CQ500* dataset, the number of slices in each CT scan varies. In order to standardize each input to the model, scans were either truncated or padded with matrices of zeros so that each scan had the same number of slices. A majority of scans contained around 38 slices so 38 was chosen as the standard number of slices for each scan. Hounsfield unit (HU) windowing was used to separate the brain (WL:40, WW:80), bone (WL:500, WW:3000), and subdural (WL:175, WW:50) windows. Each slice was saved as a three-channel image with a different HU window correlating to each channel. A sample processed slice is provided in Figure 1.

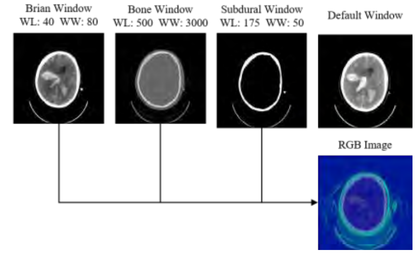


Figure 1. CT scan preprocessing methodology

Four different CNN architectures were compared: SmallerVGGNet [8], VGG16 [9], DenseNet-121 [10], and MobileNet [11]. For each of the four different networks a CNN baseline was created. The

CNNs were trained on the *CQ500* dataset where the three channel input images, created from the HU windows shown in Figure 1, were converted to a single channel grayscale image and placed in a stack correlating to the slice position in the CT scan. The final layer for each network contained a sigmoid activation function, which allowed for multi-class multi-label classification for all 14 types of brain trauma evaluated in the *CQ500* dataset. 10-fold cross validation was used in each of the networks which resulted in ten training runs for each network with a different test set for each run.

In order to utilize information in the spatiotemporal properties of CT scans, networks that combined 2D CNN results with one-dimensional CNN (1DConv), LSTM, and GRU layers were also trained. The original goal for this paper was to explore CNN-LSTM networks with different standard CNNs as inputs to the LSTM, but preliminary results showed insignificant to no improvements and poor generalization in test accuracy using larger standard CNNs such as VGG16, DenseNet, and MobileNet. As a result, those architectures were not used with 1DConv, LSTM or GRU layers due to time constraints with training. Instead, this study further explores the use of the SmallerVGGNet CNN architecture in conjunction with 1DConv, LSTM and GRU networks.

Figure 2 outlines the CNN-LSTM architecture. The CNN-1DConv and CNN-GRU networks are like the CNN-LSTM network, but a 1DConv or GRU layer is used instead of an LSTM layer. Code and models for all trained networks are available at https://github.com/marbd11/ctscan-stroke.

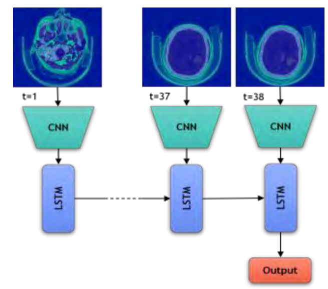


Figure 2. CNN-LSTM Network architecture

Out of the seven networks trained, SmallerVGGNet-LSTM performed significantly better than the other networks. The average test accuracy for the SmallerVGGNet-LSTM network was 99.2 ± 0.2% while the average test accuracies for the base CNNs, SmallerVGGNet-1DConv, and SmallerVGGNet-GRU ranged between 74% and 83%. Out of the ten runs executed for each network, an average run was selected to represent each network type in Figure 3. Figure 3 provides a visual representation of average change in test accuracy per epoch.

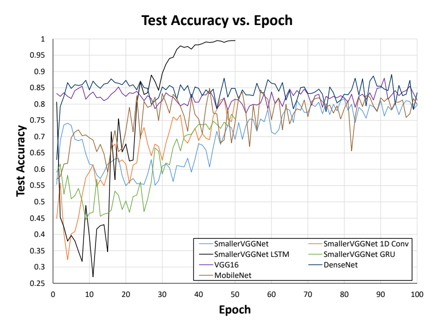


Figure 3. Change in test accuracy per epoch for each network. SmallerVGGNet-LSTM test accuracy is represented in black

Overall average accuracies and standard deviations for each network can be found in Table 2. A box and whisker plot was generated to assess the statistical significance of the results and can be viewed in Figure 4. Out of all of the trained networks, the SmallerVGGNet-LSTM network results had the lowest variance and highest mean. Student's t-tests performed between each network indicate that the SmallerVGGNet-LSTM network performed significantly better than all other networks. T-test values calculated between 2D CNNs indicate that there is not a significant difference in the performance of the different base 2D CNNs.

Table 2. Average test accuracy ± standard deviation

	CNN	CNN-1DConv	CNN-LSTM	CNN-GRU
SmallerVGGNet	82.9 ± 2.6%	78.9 ± 4.0%	99.2 ± 0.2%	74.5 ± 4.0%
VGG16	82.5 ± 3.7%	-	-	-
DenseNet	80.8 ± 3.6%	-	-	-
MobileNet	81.0 ± 3.1%	-	-	-

As hypothesized, the SmallerVGGNet-LSTM network outperformed all 2D CNNs that were trained, including the base SmallerVGGNet CNN. Higher accuracy was achieved by leveraging the generally ignored slice position information in the CT scan. This makes intuitive sense as hemorrhaging is present across different CT scan slices. Radiologists look across different slices to make their decision and machine learning assisted radiology should as well. It was unexpected, however, that the accuracy of the network was inversely correlated to the number of parameters in the CNN. Based on preliminary results, it was found that the addition of an LSTM to any CNN did not achieve significantly higher accuracies than the CNN itself. Higher accuracies were only achieved using a smaller CNN with fewer parameters paired with an LSTM.

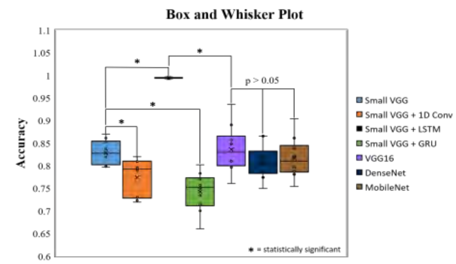


Figure 4. Test accuracy box and whisker plot

Compared to existing brain trauma classification algorithms, the SmallerVGGNet-LSTM network matches high accuracies while identifying more types of brain trauma in a single algorithm. This solution to CT scan analysis provides a means for identifying the presence of five types of hemorrhaging, fractures, mass effect, midline shift, and bleed locations. Existing algorithms typically only provide classification for up to five types of hemorrhages. Other solutions require more than one algorithm to determine the presence or absence of these 14 types of brain trauma. Ultimately, this work provides promising results for a simple end-to-end trainable solution for identifying many types of head trauma.

Acknowledgments

Christopher Angelini is supported by the US Department of Education GAANN award P200A180055. The project was partly supported by NSF Award OAC-2008690 and the Camden Health Initiative of the Rowan University.

References

[1] M. Katan, M.D., A. Luft, M.D., “Global burden of stroke,” Semin Neurol, vol. 38, pp.208-211, 2018.

[2] J. Wardlaw, “Overview of Cochrane thrombolysis meta-analysis,” American Academy of Neurology, vol. 57, Sept 2001.

[3] T. Gong, et al., “Classification of CT brain images of head trauma,” In IAPR International Workshop on Patter Recognition in Bioinformatics, 2007.

[4] Liu, Ruizhe, et al. "Hemorrhage slices detection in brain CT images," In 19th International Conference on Pattern Recognition, Dec 2008

[5] M. Al-Ayyoub, D. Alawad, K. Al-Darabsah, I. Aljarrah, “Automatic detection and classification of brain hemorrhages,” WSEAS Transactions on Computers, vol. 12, pp.395-405, Oct 2013.

[6] S. Chilamkurthy, et al., “Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study,” The Lancelet, vol. 392, Dec 2018.

[7] J. Ker, S. Singh, Y. Bai, J. Rao, T. Lim, L. Wang, “Image thresholding improves 3-dimensional convolutional neural network diagnosis of different acute brain hemorrhages on computed tomography scans,” Sensors, vol. 19, May 2019.

[8] A. Rosebrock, “Keras and Convolutional Neural Networks (CNNs),” pyimagesearch.com, Apr. 2018. [Online].

[9] K. Simonyan, A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” arXiv preprint arXiv:1409.1556, Apr 2015.

[10] G. Huang, Z. Liu, K. Weinberger, L. van der Maaten, “Densely Connected Convolutional Networks,” In Proceedings of the IEEE conference on computer vision and pattern recognition, Dec 2016.

[11] Howard, et. al., “MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications,” arXiv preprint arXiv:1704.04861, Apr 2017.

HENRY M. ROWAN COLLEGE OF ENGINEERING

Identification of Abnormalities in Head Computerized Tomography Scans

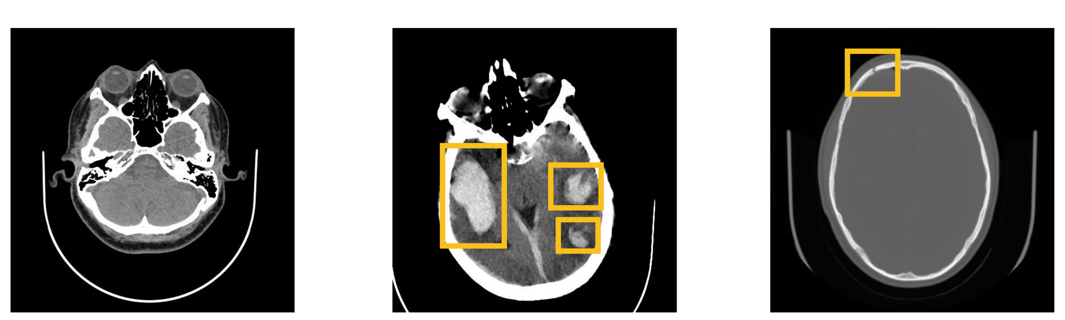
M. DelRocini, C. Angelini and G. Rasool
Department of Electrical and Computer Engineering, Rowan University, Glassboro, USA

Abstract

- Stroke and traumatic brain injuries are leading causes of death and long-term disability globally.
- Early detection of abnormalities in head computerized tomography (CT) scans reduces patient risk of serious medical complications from brain injuries.
- A machine learning algorithm trained to identify and classify CT head scan irregularities autonomously has the potential to decrease anomaly detection time and allow for quicker, more effective treatment.
- A single end-to-end neural network consisting of a convolutional neural network (CNN) and a recurrent neural network (RNN) is proposed in this study. The former network processes individual slices while the latter combines information from multiple slices to identify and classify 14 different types of brain injuries.
- Two types of RNNs were considered: long-short term memory (LSTM) and gated recurrent unit (GRU).
- LSTMs and GRUs use multiple gates in their architecture to avoid gradient vanishing and exploding problems faced by simple RNNs.

Dataset

- The *CQ500* dataset is a set of CT scans collected by the Centre for Advanced Research in Imaging, Neurosciences and Genomics, New Delhi, India.



Healthy Patient | ICH, IPH, SAH, Bleed Location Left, Bleed Location Right | Fracture, Calvarial Fracture

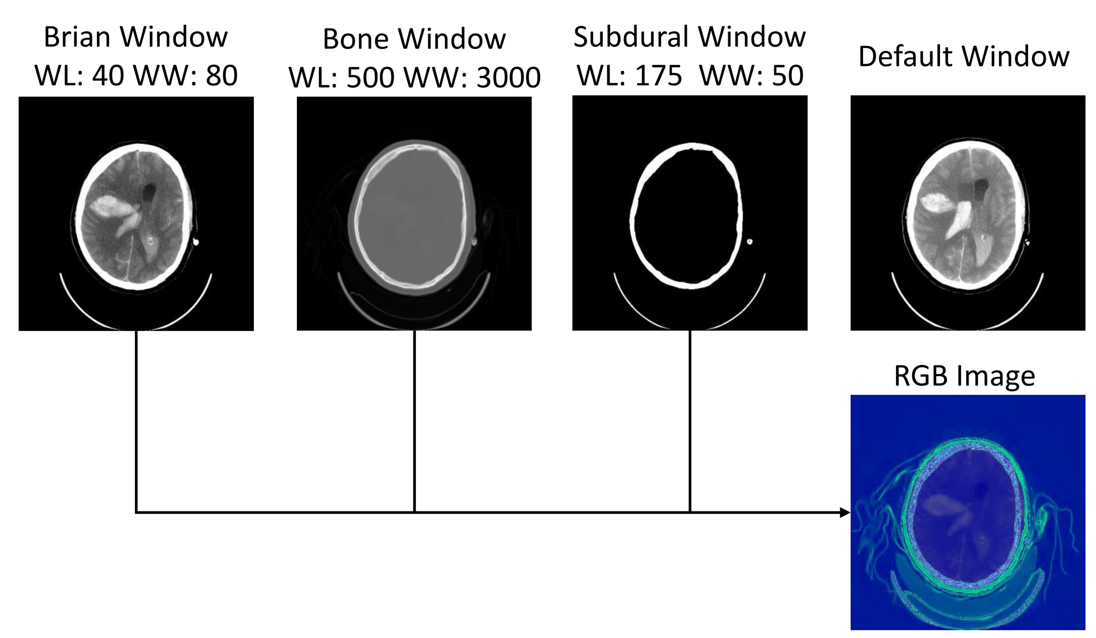
- CT scans were independently labeled by three senior radiologists for the presence of 14 types of brain trauma.

Brain Trauma Type	% of Scans	Brain Trauma Type	% of Scans
Intracranial Hemorrhage (ICH)	41.8%	Intraparenchymal Hemorrhage (IPH)	27.3%
Intraventricular Hemorrhage (IVH)	5.7%	Subdural Hemorrhage (SDH)	10.8%
Extradural Hemorrhage (EDH)	2.7%	Subarachnoid Hemorrhage (SAH)	12.2%
Bleed Location Left	26.7%	Bleed Location Right	26.3%
Chronic Bleed	3.9%	Fracture	8.0%
Calvarial Fracture	6.9%	Other Fracture	2.4%
Mass Effect	26.9%	Midline Shift	13.3%

- The *CQ500* dataset was used for training all networks.

Data Pre-Processing

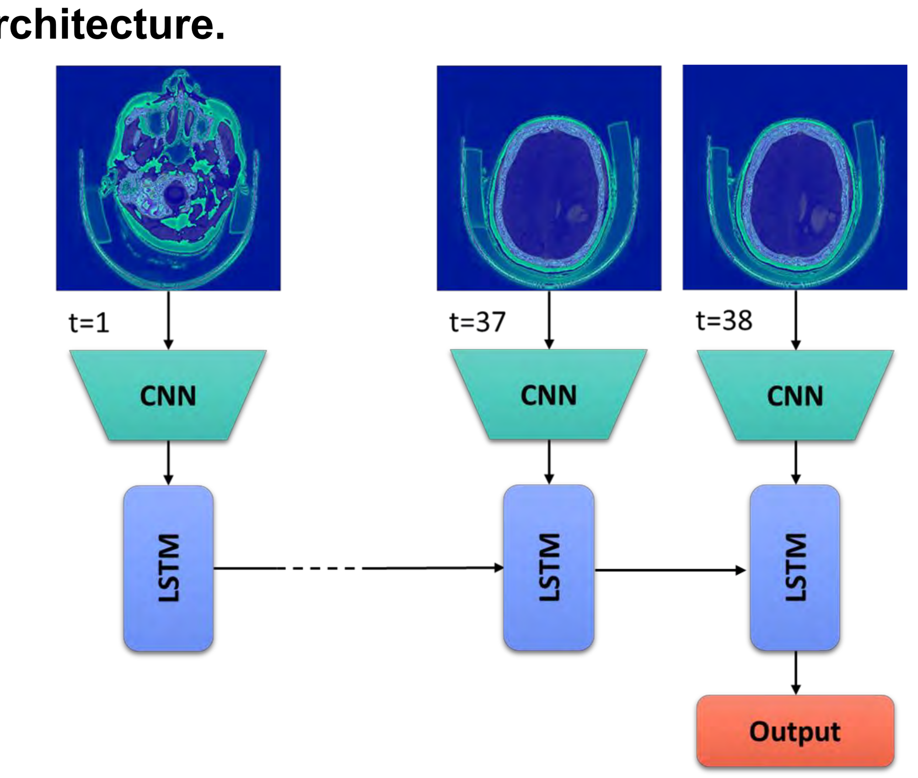
- A non-contrast axial scan was selected for each patient and resampled to a slice thickness of 5 mm.
- Scans were either truncated or padded with matrices of zeros so that each scan had the same number of slices.
- Hounsfield unit (HU) windowing was used to separate the brain (WL:40, WW:80), bone (WL:500, WW:3000), and subdural (WL:175, WW:50) windows.



- The individual slices of a brain CT scan can be considered frames (or images) of a movie and machine learning techniques of image and time-series processing can be combined for the analysis of the whole CT scan.

Networks

- Four different CNN architectures were compared; SmallerVGGNet, VGG16, DenseNet-121, and MobileNet.
- Multi-class multi-label classification was performed for all 14 types of brain trauma evaluated in the CQ500 dataset.
- 10-fold cross validation was used in each of the networks.
- In order to utilize information in the spatiotemporal properties of CT scans, networks that combined CNN results with one-dimensional (1D) CNN, LSTM, GRU layers were trained.
- The figure below outlines the CNN-LSTM architecture.



- The CNN-1DConv and CNN-GRU networks are like the CNN-LSTM network, but a 1D Convolutional or GRU layer is used instead of an LSTM layer.

Results

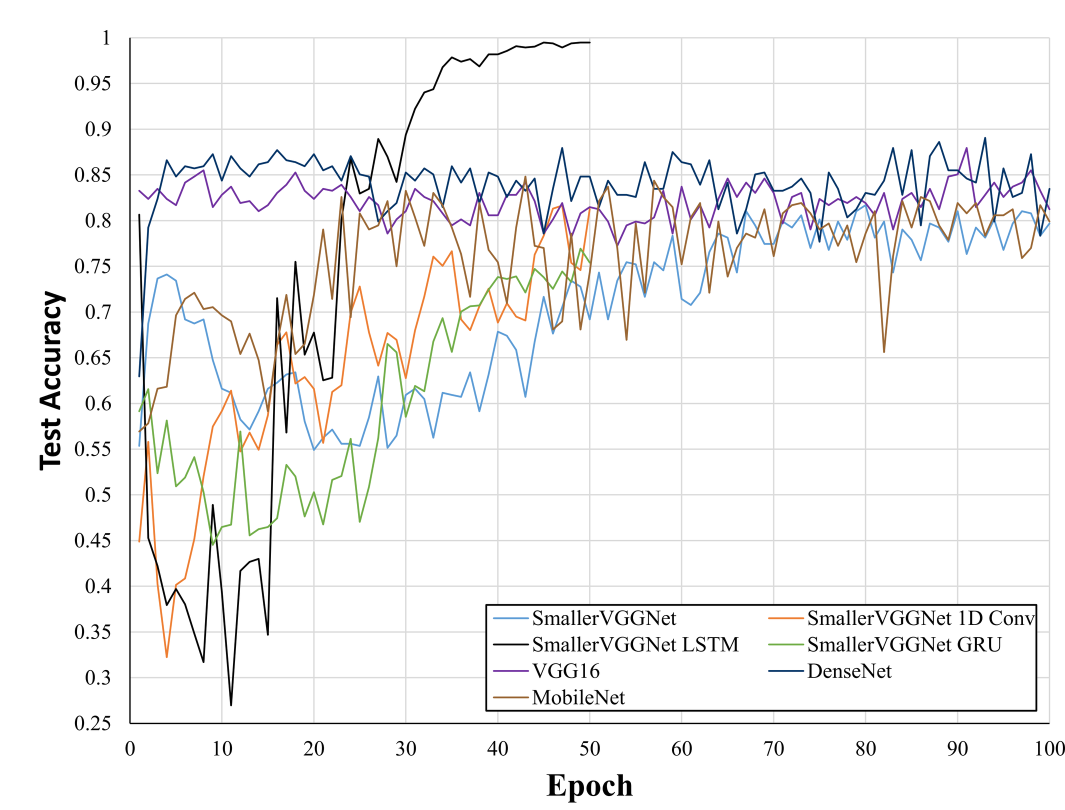
- SmallerVGGNet-LSTM produced higher test accuracies than the other networks.
- A statistical analysis showed that the SmallerVGGNet-LSTM network results had the lowest variance and highest mean.
- Student's t-tests performed between each network indicate that the SmallerVGGNet-LSTM network performed significantly better than all others.
- Preliminary results showed insignificant to no improvements and poor generalization in test accuracy using larger standard CNNs such as VGG16, DenseNet, and MobileNet.

Results Figures

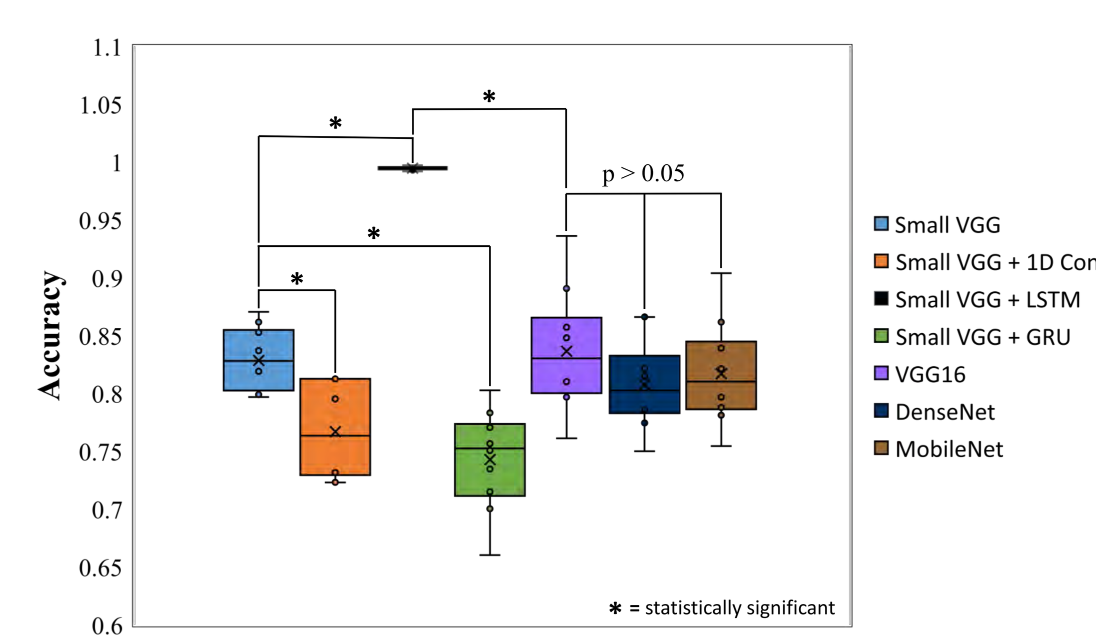
- Average Test Accuracies:

	CNN	1D-CNN	CNN-LSTM	CNN-GRU
SmallerVGGNet	82.9 ± 2.6%	78.9 ± 4.0%	99.2 ± 0.2%	74.5 ± 4.0%
VGG16	82.5 ± 3.7%	-	-	-
DenseNet	80.8 ± 3.6%	-	-	-
MobileNet	81.0 ± 3.1%	-	-	-

- Test Accuracy vs Epoch:



- The above figure uses an average run out of the 10 runs per network to represent each network type.
- Box and Whisker Plot:



- Above graphically depicts the statistical significance of the results and includes calculated statistic values for each network.

Conclusion

- The SmallerVGGNet-LSTM network outperformed all 2D CNNs that were trained.
- Higher accuracy was achieved by leveraging the generally ignored information of slice position in the CT scan.
- Based on preliminary results, it was found that the addition of an LSTM to any CNN did not achieve significantly higher accuracies than the CNN itself.
- Higher accuracies were only achieved using a smaller CNN with fewer parameters paired with an LSTM.
- Compared to existing brain trauma classification algorithms, the SmallerVGGNet-LSTM network can match high accuracies while identifying more types of brain trauma in a single algorithm.

Future Work

- Since head CT scans have subtle differences depending on where the scan is collected and the machine that the scan is collected on, it is vital that a study on the generalization ability of this network is performed using a larger dataset.
- Future work for this project includes the evaluation of algorithm performance on different sets of non-contrast head CT scans and evaluating different base CNNs for CNN-LSTM network.

Code

- Code and models for all trained networks are available at https://github.com/marbd11/ctscan-stroke.

References

1. M. Katan, M.D., A. Luft, M.D., "Global burden of stroke," Semin Neurol, vol. 38, pp.208-211, 2018.
2. J. Wardlaw, "Overview of Cochrane thrombolysis meta-analysis," American Academy of Neurology, vol. 57, Sept 2001.
3. S. Chilamkurthy, et al., "Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study," The Lancelet, vol. 392, Dec 2018.
4. A. Rosebrock, "Keras and Convolutional Neural Networks (CNNs)," pyimagesearch.com, Apr. 2018. [Online].
5. K. Simonyan, A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, Apr 2015.
6. G. Huang, Z. Liu, K. Weinberger, L. van der Maaten, "Densely Connected Convolutional Networks," In Proceedings of the IEEE conference on computer vision and pattern recognition, Dec 2016.
7. Howard, et. al., "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," arXiv preprint arXiv:1704.04861, Apr 2017.

Acknowledgements

- Christopher Angelini is supported by the US Department of Education GAANN award P200A180055. The project was partly supported by NSF Award OAC-2008690 and the Camden Health Initiative of the Rowan University.