# A Convolutional-Sequential Model for Intracranial Hemorrhage Detection

Richard Qiu, Jorma Görns, Philippe Noël, Jonathan Berman

December 16, 2019

**Abstract**

Brain hemorrhages are extremely hard for trained medical professionals to detect from brain scan images, which are slices of a patient's brain imaged through an fMRI. We adapt the long-term recurrent convolutional network (LRCN), a model architecture conventionally used for action recognition in video data, to brain CT scans used by radiologists to screen for intracranial hemorrhage. Specifically, we represent 3D CT scans as series of 2D images and train a convolutional neural network, VGG19, on these 2D images. Then, we take the sequential convolutional output of this network train a bi-directional LSTM to create a binary prediction for every 2D image, for each sequence. This allows us to better utilize the sequential properties between brain scans that is lost in traditional 2D CNN models. We also show how this convolutional-sequential model is more powerful than a CNN alone for this task, obtaining a recall of 73% for the LRCN, versus a recall of 94% for the CNN alone. This work demonstrates how combining image recognition and sequential models can help better solve complicated real-life problems with non-uniform information, and shows how a model like this can have applications in many other healthcare problems.

## 1 Introduction

Intracranial hemorrhages cause about 10% of strokes in the U.S., and strokes are the fifth-leading overall cause of death. Determining whether or not a hemorrhage is present is the first step towards treating the condition and can be a very challenging task for radiologists who have to sort through many dozens, or even hundreds, of images under time constraint.

To detect intracranial hemorrhages, radiologists typically examine 3D brain CT scans, looking for asymmetries caused by swelling [4]. They interact with these CT scans as sequences of 2D images, relying on context from many subsets of images to make diagnoses. Traditional machine learning models for this task often disregard this human approach, typically modeling CT scans as simple 2D images, without the context of adjacent slices. This is typically because 3D

CT scans are often variable in length, and therefore difficult to model using conventional approaches.

We integrate 3D CT scan context into the problem of detecting acute intracranial hemorrhage by adapting a model architecture which can account for regional and long-term dependencies within a given CT scan, and that is more robust to in-practice conditions as it can account for variable length 3D CT scan inputs.

We note that the goal of this work is not to replace radiologists, but rather to complement their work and reduce the strain of filtering through vast quantities of CT images, the majority of which are not particularly relevant. Because of this and the nature of medical trauma, we find it much more important to detect the presence of hemorrhage rather than predicting the type of hemorrhage present. As such, we focus only on predicting whether an hemorrhage is present or not (accuracy), and focus on ensuring there are few false negatives (high recall), a critical case to avoid with medical trauma.

# 2 Background

## 2.1 Recent Work in Medical Image Recognition

Predicting conditions based on medical images is a complex problem that even state-of-the-art machine learning techniques struggle to solve. These images are generally very noisy and difficult to interpret, datasets are generally very imbalanced or incomplete, and the implications of a model's predictions very scrutinized, making it no simple image recognition task. As such, many different feature engineering and other pre-processing techniques have been developed and used over the years to further augment standard image recognition models for those tasks.

### 2.1.1 Windowing

One such technique is "windowing", which consists in re-scaling pixels in a specific area to emphasize a intensity-region of an image. All pixels outside of a given range of intensities are set to 0. Various biological structures appear under CT scan in different intensity ranges. By windowing to a limited range of intensities, radiologists and medical professionals can better understand what is going on in cranial structures associated with that region and better recognize whether or not there is acute trauma [8].

### 2.1.2 Data Imbalance

Medical image datasets are often fairly unbalanced, as medical conditions arise infrequently, which makes it hard to train machine learning models on those datasets. Supervised learning historically performs very poorly in severely unbalanced datasets, and so techniques have been developed to re-balance datasets or weigh more heavily classes with a lower representation. One such example that has been used heavily in medical imaging is, SMOTE an algorithm that creates new, fake, examples of the minority class while also under-sampling the majority class, which has been shown to drastically help learning models generalize in many context, including medical imagery [2].

## 2.2 Convolutional Neural Networks

Unsurprisingly, convolutional neural networks, or CNNs, have been applied frequently to medical image recognition tasks in the past [10]. They are well suited for this type of task due to their ability to somewhat mimic the human visual system and they generalize well on complex image recognition tasks [1][9]. They are especially suited to this type of problem as the kernel filtering and downsampling pooling processes help identify small but highly relevant components in an image, often how medical conditions appear under imaging. Since medical machine learning models are mainly intended to support doctors, it is especially important that they pick up on the smallest, hardest-to-see edge cases which a doctor would have difficulty identifying or could easily overlook, making CNNs an ideal base model.

### 2.2.1 Augmented Convolutional Neural Networks

Many different reworkings of CNNs exist to help them tackle better certain image recognition tasks. One example is the mask R-CNN, or regional CNN, used for object detection [6]. Another example is the Long-Term Recurrent Convolutional Network (LRCN) [5], where the convolutional output of a CNN is adapted as input for an RNN or LSTM. This approach was originally developed to handle action recognition in video data, but in this paper, we adapt this approach to brain CT scans and hemorrhage prediction, as the recurrent nature of the model allows information relevant for prediction to be shared across image slices within a given CT scan.

# 3 Our Approach: A Convolutional-Sequential Model

## 3.1 Data

The data consists of DICOM image slices of patient brain CT scans, each with physician-created labels for the presence of hemorrhage, and if hemorrhage is present, which specific type(s) of hemorrhage are present. Notably, each image also contains metadata which allows us to reconstruct a specific brain image stack (i.e. the 3D CT scan corresponding to a specific patient and exam), which inspires our convolutional-recurrent approach.This dataset is the largest collection of its kind to-date, consisting of over 25,000 CT exams. It is provided by the Radiological Society of North America (RSNA) in collaboration with members of the American Society of Neuroradiology (ASNR) and MD.ai as part of the "RSNA Intracranial Hemorrhage Detection" Kaggle challenge.

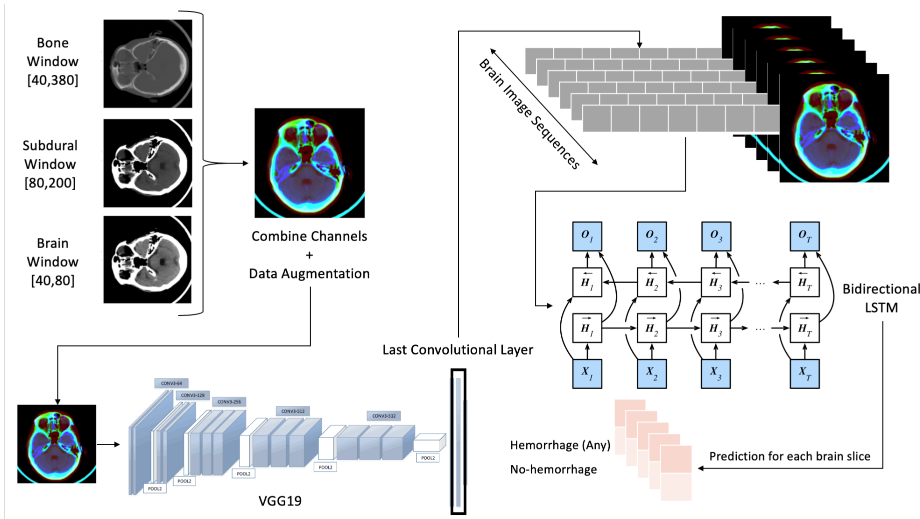


Figure 1: Convolutional-Recurrent Neural Network model we developed to tackle the RSNA brain hemorrhage dataset. The brain scan images are taken from the RSNA Kaggle blog post.

## 3.2 Data Processing & Features Engineering

We pre-processed the DICOM images by down-sampling them to 128x128 PNGs images due to memory limitations. We then randomly sampled 2,500 patients from the datasets to use, which gave us a dataset of about 87,000 images. For each patient brain image, we windowed the gray-scale image with three different standard windowing regions (bone, subdural, and brain; see Figure 1), which gave us three different important intensity-regions for detecting hemorrhages. We then combined these images into a single three-channel image as input to a 3-channel CNN. We then performed standard data augmentation by shifting images around, rotating and mirroring them. We also explored SMOTE as a way to balance the dataset, which had about 85% no-hemorrhage labels and 15% hemorrhage labels, but decided against it as it would prevent us from keeping the patient ID association to the data, which is necessary for the sequence part of our model later on [3].

## 3.3 The Convolutional-Sequential Architecture

We introduce a *Convolutional-Sequential* model (LRCN) for the brain hemorrhage prediction problem, with the goal of further improving upon conventional CNN models by integrating sequential context over over a given patient's CT brain slices (see Figure 1). We design our model as a combination of a CNN and a bi-directional LSTM. The CNN learns on 2D CT images and predicts a label, hemorrhage or no hemorrhage, without sequential context. We then scrape the classification layers off of the CNN and freeze the convolutional layers, attaching them to a standard bi-directional LSTM, with each image representing a time-step, going through the convolutional portion of the architecture before entering the LSTM. Finally, we feed the output of each LSTM time-step into two fully connected layers (with shared weights across time steps) to generate a final binary prediction. This new convolutional-sequential model is then trained on sequential data formed by grouping the data on individual patients, ordered by slice depth. Essentially, we train the convolutional-sequential model on the full 3D images, applying the convolutional portion of the network to each individual 2D slice. This allows us to capture local (if slice $s$ has hemorrhage, it's likely slices $s+1$ and $s-1$ have hemorrhage too) as well as potential long-range dependencies across a given CT scan [7].

We trained our models on a Windows Server 2016 virtual machine with 48 vCPUs, 192 GB of RAM and a Tesla T4 GPU with CUDA 10.0 and cuDNN 7.6.4.

### 3.4 Baselines

We established rapid baselines by performing PCA analysis on the dataset and keeping the 200 highest explained-variance PCs, which were fed to a logistic regression and a random forest model. We obtained 85.4% accuracy and a recall of 1.7% for the logistic regression and 87.2% accuracy and 11.7% recall for the random forest, which is what we were expecting for such simple models, ill-suited to the unique features of the dataset. This nonetheless help us guide our direction in regards to class-imbalance and sets recall as a very important metric to ensure the model is not always predicting the most frequent class, no hemorrhage, with 85% importance.

### 3.5 CNN Model

We used VGG19 as the CNN model on which we trained the windowed brain images. We froze the first 3 convolutional layers to mimic transfer learning, as VGG19 is a strong model trained on a wide array of images and the first few layers pick up very abstract shapes, which were good enough for our task. We kept the last two layers unfrozen for learning and added two 1,000 unit dense layers at the end with a 0.5 dropout layer between them to prevent overfitting. We used Adam optimizer with a 0.0001 learning rate and a binary cross-entropy loss function, which has been proven to work well on binary classification. We trained the model for 50 epochs using a 80/20 train-test split on our 87,000 images, resulting in 2,150 mini-batches of 32 brain images per epoch.

### 3.6 Convolutional-Sequential Model

We construct the convolutional-sequential model as described above. We train on using the same dataset and split, only grouping images into sequences based on CT scan, and using the same optimization parameters. We start training by mimic the structure of the dense layers of the VGG19 architecture, using 1,000 unit dense layers with 0.5 dropout layers in between. Then, to offset the added complexity of the LSTM layer, we reduced the complexity of the dense layers and introduced batch normalization between each dense layer such that the relative complexity between the LRCN and CNN architectures was more comparable.

## 4 Results

Our main results compare the VGG19 CNN's performance on the classification that and that of the LRCN on the same task. Our simple baselines were meant as mere guidance in our approach and are thus not reported in the table below.

Table 1: Test Metrics of CNN & LRCN

| Model | Accuracy | Recall | Precision | AUC |
|---|---|---|---|---|
| VGG19 | 86.83 | 73.32 | 84.37 | 95.23 |
| LRCN | 87.71 | 93.78 | 81.97 | 90.50 |

As we can observe, we obtain higher accuracy and much higher recall for the LCRN than for the VGG19 CNN. This can be explained by the fact that we predict much more aggressively for hemorrhages with the LCRN, as compared to the VGG19 CNN model which mostly predicts relatively fewer hemorrhages. This is the behavior we are looking for in our model, as false positives are much more tolerable than false negatives, making the LCRN a much better model in this way. This difference could be explained by the improved learning efficiency of the LCRN provided by the additional CT scan context the model has available.

The precision can also be explained the same way. The precision for the LCRN is slightly lower, at 82%, than the VGG19 CNN model, at 84%, because we predict more aggressively and get more false positives. However, so long as we get very few false negatives, we are willing to tolerate a this moderate increase in the number of false positives.

Likewise, the difference in AUC results are also a byproduct of the different prediction approach of the CNN vs LRCN. The VGG19 CNN obtains a higher AUC, with 95%, because it predicts very frequently no hemorrhage. Since 85% of the dataset is labelled as no hemorrhage, and the AUC is a metric which shows how good a classification model is at classifying 0s as 0s and 1s as 1s, it gets almost all the no hemorrhage as no hemorrhage and gets a very high

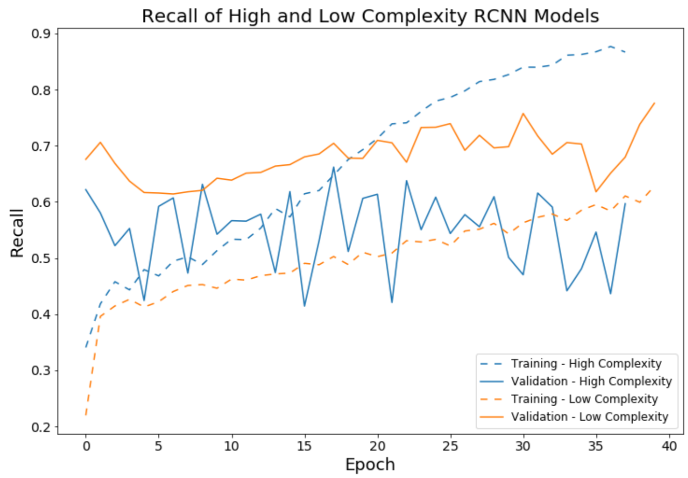


Figure 2: Evolution of the recall for the bi-directional LSTM model we trained on the CNN last convolutional layer weights as part of our convolutional-recurrent model.

AUC. This is, however, at the expense of often classifying hemorrhages as no hemorrhage.

The LCRN, on the contrary, obtains a lower AUC of 90%. We propose that this is because the LCRN better learns the data and is therefore less influenced by the data imbalance, leading it to more accurately predict case of hemorrhage, at the cost of AUC, given the data imbalance. This explains intuitively why the LCRN is much better suited for this task. It predicts many fewer false negatives, at the cost of slightly more false positives as compared to VGG19, because it is more robust to the inherent class imbalance of the dataset. Again, we propose that this is because of the additional sequential context the LRCN has available to it.

Alternatively we propose that this more aggressive prediction behavior is due to the locality of the LRCN (in that if a given brain slice is predicted to have hemorrhage, adjacent slices will become more biased towards predicting hemorrhage as well). Thus, because of the additional context, the LRCN model can better predict positive cases in images at the edge of the actual region of hemorrhage.

# 5 Conclusion

In this work we have shown how a convolutional-sequential model combining a CNN and an LSTM can be used to produce state-of-the-art result on the complex medical imaging task of predicting brain hemorrhage based on the RSNA Brain Hemorrhage dataset. We have successfully applied windowing to the dataset of DICOM images and used this resulting dataset to train a CNN of which the last dense layer for a specific patient was fed to a bi-directional LSTM to train over brain image sequences for a specific patient, further increasing accuracy and recall as compared to a CNN alone and performing significantly better than any off-the-shelf supervised learning model.

This work demonstrates how combining image recognition and sequence modeling can perform better than standard, state-of-the-art image recognition models like VGG19 alone could by harnessing sequences over complex image datasets like is often the case in medical data. It also demonstrates that strong results can be achieved on imbalanced datasets with the right model and serves as yet another illustration of the importance of data pre-processing. Many other applications of convolutional-sequential models exist, both in and out of the medical field, and we hope this can serve as a model to build upon in tackling those problems.

# 6 Future Work

We see many avenues for future work on this problem. First, we trained our model on predicting whether an hemorrhage was present or not, but did not train on predicting which type of hemorrhage was present. This decision was made on the prior that predicting an hemorrhage is much more important than predicting its type, as radiologists can then spend more time diagnosing the specific hemorrhage type once the scan has been flagged. Expanding the model to predict the specific hemorrhage type is a natural next step to the project.

In this work, we utilized a downsampled version of about 10% of the available CT data. Training on the full dataset of over 700,000 images in full-size posed barriers with respect to our available resources, but would likely help reach better performances, especially on predicting the hemorrhage type. More experimentation with respect to the model architecture is also needed. In particular, different base convolutional networks and recurrent architectures should be tested and benchmarked. Further exploring the differences in training between the two models explored in this work is important, and may help improve the architecture of the LRCN model. Other avenues we see for improving performance including ensembling the CNN and the LSTM components of the model to better improve performance, and considering attention mechanisms as an alternative to a bi-directional LSTM as we used. We hope the medical machine learning community will take up on those avenues and build upon our work.

**Acknowledgments**

We would like to thank professors Jean-Baptiste Tristan and Michael Wick from Harvard University for guiding us in this work, and Mari Tanako, medical student at Harvard Medical School, for providing a radiologist's perspective on our work.

# References


[1] Mohammad R. Arbabshirani et al. "Advanced machine learning in action: identification of intracranial hemorrhage on computed tomography scans of the head with clinical workflow integration". en. In: *npj Digital Medicine* 1.1 (Apr. 2018), pp. 1–7. ISSN: 2398-6352. DOI: 10.1038/s41746-017-0015-z. URL: http://www.nature.com/articles/s41746-017-0015-z (visited on 12/16/2019).

[2] N. V. Chawla et al. "SMOTE: Synthetic Minority Over-sampling Technique". en. In: *Journal of Artificial Intelligence Research* 16 (June 2002), pp. 321–357. ISSN: 1076-9757. DOI: 10.1613/jair.953. URL: https://www.jair.org/index.php/jair/article/view/10302 (visited on 12/16/2019).

[3] *Cleaning the data for rapid prototyping (fastai)*. en. URL: https://kaggle.com/jhoward/cleaning-the-data-for-rapid-prototyping-fastai (visited on 12/16/2019).

[4] *CT brain hemorrhage - Startradiology*. URL: https://www.startradiology.com/internships/neurology/brain/ct-brain-hemorrhage/ (visited on 12/16/2019).

[5] Jeff Donahue et al. "Long-term Recurrent Convolutional Networks for Visual Recognition and Description". In: *arXiv:1411.4389 [cs]* (May 2016). arXiv: 1411.4389. URL: http://arxiv.org/abs/1411.4389 (visited on 12/17/2019).

[6] Kaiming He et al. "Mask R-CNN". In: *arXiv:1703.06870 [cs]* (Jan. 2018). arXiv: 1703.06870. URL: http://arxiv.org/abs/1703.06870 (visited on 12/17/2019).

[7] Murtadha D Hssayeni et al. "Intracranial Hemorrhage Segmentation Using Deep Convolutional Model". en. In: (), p. 18.

[8] Hyunkwang Lee. *[1812.00572] Practical Window Setting Optimization for Medical Image Deep Learning*. URL: https://arxiv.org/abs/1812.00572 (visited on 12/16/2019).

[9] Hyunkwang Lee et al. "An explainable deep-learning algorithm for the detection of acute intracranial haemorrhage from small datasets". en. In: *Nature Biomedical Engineering* 3.3 (Mar. 2019), pp. 173–182. ISSN: 2157-846X. DOI: 10.1038/s41551-018-0324-9. URL: http://www.nature.com/articles/s41551-018-0324-9 (visited on 12/16/2019).

[10] Fabio Perez, Sandra Avila, and Eduardo Valle. "Solo or Ensemble? Choosing a CNN Architecture for Melanoma Classication". en. In: (), p. 9.

# 7 Appendix

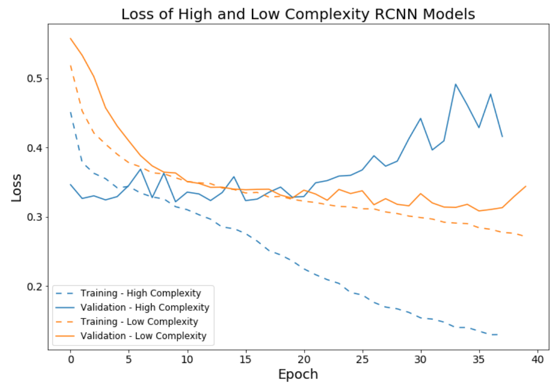


Figure 3: Evolution of the loss for the LRCN model we trained to further improve upon the CNN.

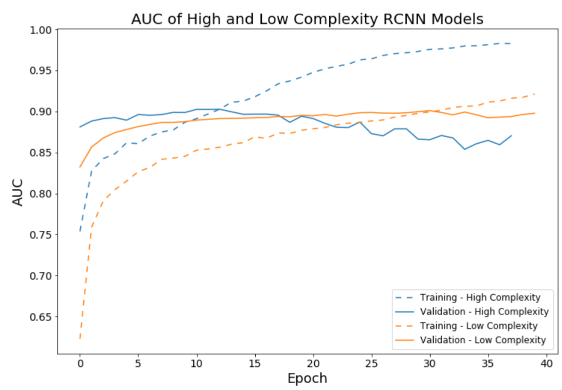


Figure 4: Evolution of the AUC for the LRCN model we trained to further improve upon the CNN.

**Code:** https://github.com/philippemnoel/cs281-rsna-hemorrhage-project