School of Informatics

Informatics Research Review
Intracranial Haemorrhaging Detection and Segmentation With Fully Convolutional Networks

[redacted]

January 2021

Abstract

The recent success of deep learning models in identifying and segmenting features from images has drawn academic interest in utilizing those models in medical applications. An area of high mortality rate that uses medical imaging to perform diagnosis is Intracranial haemorrhaging (ICH). The identification of ICH is done visually by radiologists to identify areas on CT scan slice of characteristic white lesions. Fully convolutional networks are a subset of deep learning that can highlight areas of interest. This paper will investigate the current literature that uses fully convolutional networks to segment ICH pathologies.

Date: Friday 22nd January, 2021

Supervisor: [redacted]

Contents

1 Introduction **1**

2 Background **2**

2.1 Deep Learning . . . 2

2.1.1 Fully Connected Network . . . 2

2.1.2 Convolutional Neural Networks . . . 3

2.1.3 Fully Convolutional Neural Networks . . . 4

2.2 Computer Tomography and Imaging . . . 4

3 Literature Review **5**

3.1 Patch Fully Convolutional Networks . . . 5

3.2 Ensemble Model with Windowing and Image Interpolation . . . 7

3.3 Cascaded Deep Learning . . . 8

3.4 U-Net and Even Smaller Datasets . . . 8

3.5 Condensed Results . . . 9

4 Summary & Conclusion **10**

1 Introduction

The consequences of brain hemorrhaging (a.k.a intracranial hemorrhaging - ICH) serve to be a morbid reminder of the sensitivity of the brain to any affliction. They occur as a result of internal bleeding, putting pressure onto the brain and depriving it of oxygen. What follows is cellular death.

Worldwide it affects roughly 25 out of 100,000 individuals with a mortality rate up to 52% occuring within the first month [1]. Around half of the mortality occurs within the first 24h. Only 20% diagnosed are expected to make a full recovery. Due to the time sensitive nature of the problem an early and accurate diagnosis is necessary to expediate the process of administering the correct treatment. The current process for providing a diagnosis, requires an initial analysis followed by a CT scan, which has to be followed up by a diagnosis from a radiologist. The presence of ICH on a CT scans can range from very obvious to virtually imperceptible and is heavily dependent on the ability and experience of a radiologists to detect them. Thus this area is heavily suited for an automated solution that can both cut down on time needed for a diagnosis but also mitigate the effects of human error. Such a solution would not only have the capacity to save additional lives but also free up hospital resources.

In recent years there have been developments in research to develop AI models that can accurately classify CT scan images for the presence of hemorrhaging [2]. The most significant development has been achieved by utilising a subset of Deep Learning, known as convolutional neural networks [3]. It comes at little surprise that this technique has found success within medical imaging, considering that CNNs were originally successfully applied in image classification. Furthermore, the National Institute for Health and Care Excellence, (NICE) a public body in the UK that publishes health related guidelines, has released a document recommending 6 AI software packages from various companies, where every single one uses deep learning as a basis for creating their models [4].

The aim of this paper is to survey the current publicly available research for classifying CT scans of ICH with the help of deep learning. In doing so it aims to answer the following questions:

1. What is the current state of the art in ICH detection and segmentation?
2. How do the various technologies differ and how does that affect classification?
3. What does the future hold for this technology?

Before continuing to the background section it is important to qualify the specific choice and scope of the topic. Head scans can be executed by either an MRI machine or a CT scanner. Although MRI scans tend to produce more accurate images of the target area, CT scans are more ubiquitous and comparatively cheaper, which lends itself to providing more data [5]. It is also faster in producing an image than an MRI, which is crucial in time-sensitive problems like ICH.

2 Background

The following section will delineate the necessary techniques in understanding deep learning and its various subsets that are necessary to understand current ICH model solutions.

2.1 Deep Learning

The origins of machine learning theory can be dated back to the 1950s along with Alan Turing's proposal of a "learning machine" [6]. Since then the field has seen significant academic interest with various theoretical breakthroughs. Only recently however, as a result of the development and ubiquity of computational power and the access to a vast amount of training data, did machine learning see practical and commercial success.

Though many different techniques fall under the umbrella term of machine learning, the common denominator in most approaches is access to representational data pertinent to a given use case, as well as a predefined metric, which will use said data to define our model. In the case of deep learning, the topology of the model takes inspiration from biological underpinnings in the structure of neural pathways in brains, insofar that it is defined by layers of interconnected nodes that propagate incoming data to produce a desired output. The following will be a description of a standard fully connected network. It is important to preface that the current state of the art methods no longer utilize a fully connected model due to limitations that shall be described in section 2.1.2. However, concepts introduced by this architecture are pertinent to models that are currently in use.

2.1.1 Fully Connected Network

When designing a model using a fully connected network architecture for a CT scan classification problem, one could imagine the input to the network being a vector representation of a CT scan image. Each pixel is a separate feature in the input vector corresponding to a specific greyscale value. An additional step of normalization is required to limit the scope of the value in each feature, for example by changing the range from [0, 255] to [0, 1]. Without this normalization, pixels with higher greyscale values will be perceived as more significant simply due to the intensity of the color.

The input vector $\hat{x} = [x_1, x_2, ...x_n]$ represents the input layer of the network with each feature consisting of a weight that connects it with every node in the subsequent layer $\hat{w}_i = [w_i 1, w_i 2, ...w_i n]$. The values of these weights will be adjusted during the learning process in order to identify the salient features that will allow for accurate classification. Each node in the subsequent layer takes the weighted sum of each node in the preceding layer, as well as applying a bias and an activation function. The application of an activation function ensures that our model can represent the data non-linearly. There is a substantial amount of literature dedicated to exploring various types of activation functions such as RELU, Tanh and sigmoid [7]. The bias is simply an additional parameter that allows the activation function to be shifted. The formula for the output of a given node would look as follows:

$$\hat{y} = g(\hat{x} \cdot \hat{w} + b) \tag{1}$$

Where g is the activation function and b is the bias.

Once the inputs have been propagated through the hidden layer, they reach the output layer where the decision for classification is made. In the particular use-case of CT scan analysis, it can be a simple binary choice regarding the presence or lack of haemorrhaging, represented by a single output node in the output layer. A sigmoid activation function normalizes the value between 0 and 1, allowing it to be treated as a confidence metric. Values greater than 0.5 will be classified as positive (presence of haemorrhaging), otherwise it will be classified as negative (no presence of haemorrhaging). It also possible to create an output of multi-class classification, in this case the number of output layers will be determined by the sub-regions of the brain where haemorrhaging may be present. For a multi-class classification problem, softmax is applied as an activation function and the highest value will be set as the classification for a given input.

Random initialisation of the parameters of the model virtually guarantees that all initial classification will be inaccurate. Training of the weights and biases of the model can be performed through stochastic gradient descent [8]. For classification tasks the training data is a pairwise combination of ICH scans and their respective correct classification, which have been assigned by radiologists. By comparing the values of the correct classification and the predicted value an error function can be generated on the basis of which, the parameters can be iteratively adjusted to improve overall classification accuracy. By taking the gradient of the error function, the algorithm backpropagates through the model to minimize the overall error.

2.1.2 Convolutional Neural Networks

CNNs were practically applied for the first time in 1989, through the identification of handwritten zip code digits [9]. However its true impact on image classification arose after the success of the AlexNet on the ImageNet dataset [10]. The relative success of CNNs in these tasks has seen to their application in ICH detection.

The comparative improvement in performance of CNNs against fully connected network models can be attributed to the architecturally embedded, learnt feature extraction and regularization. Overfitting is more prominent on fully connected networks due to over-parameterization. A single image may contain 100,000s of features that are densely connected with the nodes of the subsequent layer. The significant number of trained parameters will negatively impact the overall generalizability of the model.

Instead of a web of densely connected layers, CNNs use kernels that identify the presence or lack of specific patterns within feature maps. In a typical CT scan image, radiologist pinpoint haemorrhaging by the presence of lighter patches. The model may train the kernels to be sensitive horizontal or vertical arrangement of data, where there is a sudden jump in the pixel value of a group of features. Abstracting this concept further, once the model identifies these borders it can then check if their arrangement warrants a positive classification in the latter layers. In many CNN architectures it is common to see the placement of a few fully connected layers at the tail-end of the network, to potentially learn non-linear combinations of extracted low-level features.

One final important element of CNNs in feature extraction is downsampling, where the height and the width of the feature maps are halved whenever they are passed through pooling layers. This step is necessary as it makes identification of haemorrhaging more invariant to positional changes.

2.1.3 Fully Convolutional Neural Networks

From a practical point of view, it can be reasonable to expect that regardless of how well deep learning models classify an image, a radiologist will still be needed to provide a final diagnosis. A confirmation of the presence of ICH by a model, would still require identifying the suspected area. In the worst case scenario, erroneous prediction could lead to wasted time or wrongful diagnosis.

Semantic segmentation offers a solution to the above problem. Instead of assigning a classification label to an entire image, each feature or pixel of the original image is assigned a class. The final output returns the original image with highlighted areas where the model suspects the presence of haemorrhaging. The elimination of ambiguity regarding where the model detected the pathology will accommodate the verification process for a radiologist.

This process is achieved through a fully convolutional architecture [11]. In CNNs the height and width of the features gets downsampled as it passes through convolutional blocks and pooling layers, ending with a few fully connected layers before the final output layer. Since the final output needs to be the same dimensionality as the input, the fully connected layers are replaced with upsampling convolutional layers, making the entire network "fully convolutional". Downsampling is necessary in the architecture, for both computational efficiency as well as for feature positional invariance. The detriment of that is the loss of information as dimensionality is reduced. Fully convolutional layers circumvent this problem by summing the upsampled feature maps with previous layers at various stages of downsampling.

A slight disadvantage of fully convolutional networks when compared to CNNs, is the requirement to provide explicitly drawn ground truths for each image.

2.2 Computer Tomography and Imaging

The field of research regarding image segmentation ICH scans has seen significant interest with a considerable number of papers being published yearly from various institutions beginning in 2009 with automatic segmentation technique [13]. Prominence of deep learning began in 2017, with the successful application of LeNet, googLeNet and Inception-ResNet in ICH detection tasks [14]. Though extensive research was performed in many subsets of artificial intelligence,

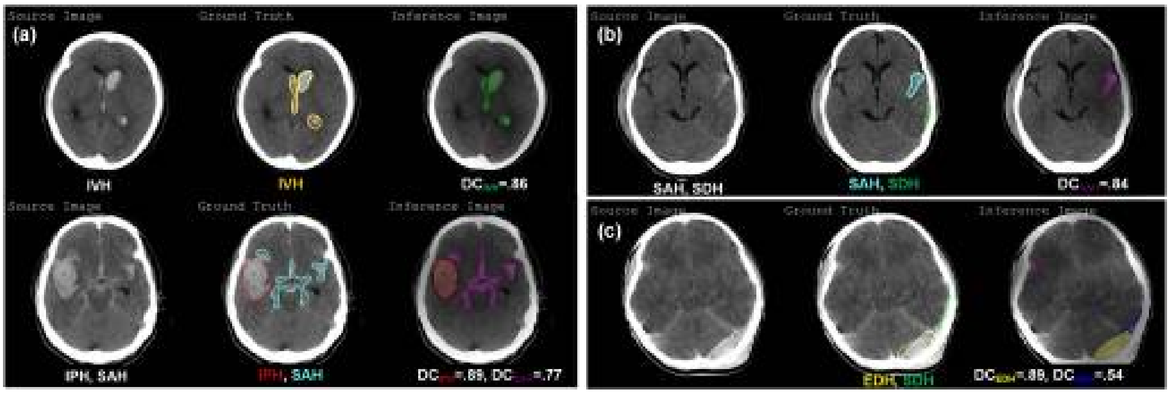

Figure 1: Image taken from the cascaded deep learning for ICH detection/segmentation paper [12]. The leftmost images are the inputs, the middle images are the ground truths annotated by radiologists and the rightmost images are the predicted segmentation masks, with color-coded class predictions.

such as fuzzy C-means and superpixels. The vast majority of recent research was primarily orientated around convolutional and fully convolutional neural networks [15].

3 Literature Review

3.1 Patch Fully Convolutional Networks

A paper published by the University of Berkeley offers a patched based approach in analysing CT scan data called PatchFCN [16]. In most cases, an entire image would be provided as input to a convoluntional network. A patch based approach subdivides the CT scan images of 512x512 pixels into seperate patches of size 240x240 pixels, with the possibility of overlap. The formula provided by the paper for the number of patches is $N = \left[\frac{\beta H}{C}\right]^2$, where β is the overlap value (set at 3 during test time), H is the input image size and C is the patch size. One problem with overlapping is that the same pixel will receive multiple predictions. This is circumvented by taking an average of each score they receive.

The paper provides several justification as to why a PatchFCN approach will net better results than a vanilla FCN. First of all, they argue that the full context of a CT scan is not necessary to make a valid prediction, as most pathologies if present are concentrated within subregions. Furthermore, the architectural implicit restriction of only obtaining local information, could have a regularization effect. Another justification is that the smaller sized patches allow for greater batch-sizes, leading to increased stabilization during training. Finally they argue that a standard FCN would be more prone to capturing long-range dependencies, which can be prone to overfitting due to the limited amount of available training data. The paper also provides empirical evidence of the better performance of PatchFCN, demonstrating better Dice, Jaccard, Pixel AP and Frame AP scores. It is worth mentioning that PatchFCN does have a significant advantage of higher batch size and the number of epochs, which could explain its relative performance. One final important element of the architecture is that utilizes two distinct models for detection and segmentation. Detection for classifying it as positive for haemorrhaging and segmentation for highlighting suspected pixels.

The issue of access to a significant amount of data is unfortunately a common thread in most papers done in this domain and most likely heavily informed the decision of this particular architecture. Such data is primarily obtained from a few institutions and require specialists for labelling and constructing ground truths. This paper had access to 591 head CT scans with a trainval/test split of 443/148 obtained from 4 separate CT scanners. Labelling and ground truths were done by certified neuroradiologists with a minimum of 10 years of experience.

The paper conducted restrospective and prospective validations of the model. Retrospective validation was done on the test data and achieved an AUC score 0.976. Prospective validation was done upon additionally collected 200 head CT scans once the training of the model was completed and achieved an AUC score of 0.966. Although the paper does state that the model was trained on all the various subtypes of acute intracranial haemorrhaging, they do not provide a breakdown of the final score on each subtype. The omission of those results may be resultant from the small amount of data available.

Another lacking element of the paper is the omission of quantitative distinction with regards to different sizes of intracranial haemorrhaging. It can be expected that a model trained for a segmentation task would be more effective at identifying pathologies that are more present within a scan compared to near-imperceptible ones. Providing separate AUC scores could have better informed the reader as to the merits of this model. Unfortunately near-imperceptible traumas are far less common to be detected and included in labelling of a CT scan, compounded with an already small dataset, may have resulted in the analytical exclusion.

The AUC score achieved by this model is comparable but does not quite beat the score of the state-of-the-art detection model [17] (at that point in time) as well as the score of the radiologist who was brought in, as a comparative benchmark. However, it is important to note that the model still achieved an impressive score considering the limits of accessible data. Nearly 11,000 CT scans for the state-of-the-art compared to 591 scans for this paper. Although this model does not beat the current best scores it does provide a framework for other research where access to data is a significant impediment.

The differences in available data could have been the deciding factor in the final score of each model and it does not necessarily reflect the true quality PatchFCN. This was later demonstrated to be the case by the followup paper done by the same group of researchers [18].

A few notable changes were done to the original PatchFCN study. The architecture was enhanced by adding two additional inputs to the network of patches relatively superior and inferior to the main patch. This approach is also done by radiologists who look and antecedent and subsequent slices of the main slice of interest. Although it may seem a bit unintuitive to add more information to the network, the decision to do so becomes clearer when taking into account the limitations of CT scans, which can be prone to producing a fuzzy image or generate artefacts that can be misinterpreted as ICH. Haemorrhaging is fluid and can be present on multiple slices, so providing a more 3d oriented picture of a given patch allows the model to better discriminate between artefacting and actual ICH.

Another notable change is the increase in available CT scan data, increasing from 591 to 4,396. The paper also provides exact positive-negative split of training and test data in tables as well as the performance of the model in detecting haemorrhaging in the various subtypes of the brain. Both of these things were missing in the earlier paper. Additionally the paper seems to be more thorough in both delineating the parameters and structure of their model as well as providing more pertinent figures.

One final important change is increasing the number of specialists against, which the model was benchmarked from 1 to 4. The model managed to beat the result of two specialists and achieved a receiver operating characteristic (ROC) area under the curve (AUC) of 0.991. Beating the previous result of 0.976 and establishing the best result of any detection model to that point.

The researches did not mention whether they ran the older version of the PatchFCN on the new dataset, providing difficulty in ascertaining whether the improvement in the final score came from the increase in available training data, the addition of superior or inferior patches to the network input or the various more subtle structural changes to the architecture itself.

3.2 Ensemble Model with Windowing and Image Interpolation

Another proposed method for ICH detection and segmentation also attempts to address the problem with small datasets like in PatchFCN [19]. The dataset size in this particular study is 904 CT Head scans. The paper discussed testing several model pre-trained on ImageNet for the detection part of their architecture and provided the results for each. These models were VGG16, ResNet-508, Inception-v39 and Inception-ResNet-v2. Interestingly they achieved the best result by combining all the models into one ensemble and by taking an average of each models final decision score, similar in nature to a Random Forest model that takes the highest voted decision from an amalgamation of different decision trees.

In directly addressing the problems related to the small dataset, the paper provides two essential solutions, both of them related to data augmentation. The first relates to creating a feature map containing three channels of the same slice with different Window Widths (WW) and Window Levels (WL). A CT slice gets generated at various levels of radiation, this is necessary as different elements of head require various levels of radiation to penetrate, i.e. the skull needs relatively high levels of radiation to penetrate followed by moderate levels for blood (haemorraghing) and low for brain tissue. The intensity of radiation is given by Hounsfield Units (HU) and Window Width specifies the range of HU in a slice and the Window Level specifies where that range is located. Providing a feature maps with three separate channels at varying levels of radiation provides the model with more context mimicking a radiologists evaluation process.

The second solution makes the same observation as PatchFCN regarding using superior and inferior slices to ascertain greater context, however they both differ in execution. Whereas PatchFCN provided the inferior and superior slices as additional inputs to the network, this paper applied an image interpolation technique to fuse all three slices into one image.

One final pertinent detail of the architecture is the addition of a 6th output node in the detection model. This node is responsible for general detection of ICH as opposed to the other five that detect ICH subtypes.

Interestingly when comparing the ROC AUC scores for both papers, the Ensemble Model paper performed better than PatchFCN on the retrospective set with a score of 0.993 but worse on the prospective set 0.961. The most likely explanation for the discrepancy in scores is due to PatchFCN using cross-validation in its retrospective score, which comparatively is a more conservative estimate of generalizability. Considering that the prospective test set is more comparable it could be stated that PatchFCN has slightly better performance. When looking at the performance of the the subsets ICH for both papers the results are fairly comparable. An interesting future experiment would be to see if the performance improves in Ensemble model by additionally applying the patched-based approach of PatchFCN.

3.3 Cascaded Deep Learning

Continuing with the theme of using slices with varying window settings, the cascaded deep learning paper offers a unique solution to ICH segmentation problem [12]. Using the default window 50/100 (WL/WW) as the primary input, the image is passed to a CNN for detection. Upon a negative result the secondary input stroke image 40/40 (WL/WW) is passed to a second CNN network as an additional verification step. If both models with their respective images are unable to detect haemorrhaging, the slice is classified as negative for ICH presence. If either network detects the presence of ICH then the input is passed to the FCN networks, for segmentation and detection of ICH subtypes. The cascaded nature of the architecture is likewise applied to the FCNs, with each image being segmented separately before being combined at the very end into one image. The paper demonstrated that this cascaded approach has improved the sensitivity of the entire model compared to a simpler model with one CNN and FCN.

Relative to the previously cited papers, this study uses the largest dataset of 5702 CT head scans. However even with the larger dataset the final specificity and sensitivity is comparable or slightly worse than the aforementioned papers. Furthermore, the results for the various subtypes ICH fared much worse than in PatchFCN and the ensemble (prospective results were not done by this research group). One potential reason for this is, unintuitively, the greater dataset, as there might be a greater make-up of the dataset that consists of rarer cases of ICH that are more difficult to detect. However, without access to the original dataset, this point remains fairly speculative. A less speculative reason could be the erroneous decision to have the FCN models to both do segmentation and detection. In PatchFCN and Ensmeble model the CNN models would be responsible for both identifying ICH as well as the subtype, whilst the FCNs would only do segmentation. The cascaded model opts only for a binary detection of ICH in the CNN and doing segmentation and subtype detection in the FCN. The architecture of FCN of downsampling feature maps then upsampling them to the size of the original image, may have lead to information loss that has contributed lower specificity and sensitivity for ICH subtypes.

One final reason for the slightly worse performance is that this paper did not include additional contextual information of superior and inferior slices to the target image, specifically only providing two different window settings for one slice. This could be an important factor as the previous papers attribute this addition as significant for the final performance of their respective models.

3.4 U-Net and Even Smaller Datasets

As a natural evolution of FCN, U-Net expands on the idea of upsampling layers by increasing the amount of feature channels, improving the context propagation process. In essence, U-Net mirrors the number of downsampling layers with upsampling layers, creating U shape of the architecture. The original U-Net paper was created with medical imaging segmentation in mind and has seen success on small datasets [20]. It was to be expected that U-Net would see application in other image segmentation tasks.

A 2019 paper published by natureresearch, applied a single U-Net model just for ICH segmentation [21]. A worrying aspect of this paper is the small datasets used for training, 51 and 150 head CT scans from different institutions. The 51 head dataset was unconventionally split into 21 for training, 25 for testing and 5 for observer variability. The limits of the paper are clearly stated related to the small dataset, which has lead to the omission of 2 ICH subtypes, those contained within the subarachnoid and intraventricular compartment.

The omissions and small dataset provides a certain challenge in comparing the results with previous papers. The main metric of evaluation in this case is the Dice score, which achieved a DSC score of around 0.9 on both datasets, without applying cross-validation compared to PatchFCN's 0.75 score. Neither the cascaded nor ensemble models can be compared as they used an overlap percentage metric instead, where they evaluated how many of the predicted pixels overlapped with the ground truth.

The paper did provide two outlier scans that struggled with accurate segmentation. One of which was a small ICH pathology, which grants some credence that the this model may be unreliable for generalizability, especially since no prospective testing was done after the model was completed.

Another paper, that also utilized the U-Net, model implemented both segmentation and detection [15]. Furthermore, it also had a relatively small dataset of 82 scans. This paper however, did not aim to create state-of-the-art performance model but rather provide a thorough literature review in automated ICH diagnosis, as well as providing a publicly available dataset and a baseline model for future academic cooperation. A significant contribution as the majority of papers in this field provide no to minimal access to the training data.

Interestingly, the model failed to generate any segmentation maps when full CT slices were provided for training. The researchers hypothesized since only a small subset of each image actually belonged to the positive class the model was heavily biased towards negative classification. By dividing each slice into 49 overlapping windows each 1/16 the size of the original slice, allowed for undersampling of the negative regions, similar to the approach taken in PatchFCN of context localization. The results of the model, however are the weakest out of all the cited papers, providing a Jaccard Index of 0.21 and Dice Coefficient of 0.31. The detection results didn't fare much better, since no AUC was provided and only sensitivity and specificity and different thresholds, the best overall accuracy of 87% was achieved with a sensitivity of 63.1% and specificity of 88.6%.

An interesting suggestion provided by the paper in the future work section, was creating a model that used LSTMs in incorporating superior and inferior slices to the model.

3.5 Condensed Results

Table 1: Full Results

Paper	No. of CT Scans	ICH Detection Score	ICH Segmentation Score
PatchFCN (original)	591	Retrospective - AUC 0.976 Prospective - AUC 0.966	Retrospective - Dice: 0.766, Jaccard: 0.620, AP: 0.785
PatchFCN (updated)	4,396	Retrospective - AUC 0.978 Prospective - AUC 0.991	Retrospective - Dice: 0.75
Ensemble Model	904	Retrospective - AUC 0.993 Prospective - AUC 0.961	Retrospective - AP: 0.781
Cascaded Deep Learning	2647	Retrospective Sensitivity 0.9791 Specificity 0.9876	Retrospective Precision 0.8019 Recall 0.8215
U-Net segmentation	25/50	-	Retrospective - Dice: 0.90/0.91
U-Net segmentation + detection	82	Retrospective (Best Acc.) Sensitivity 0.631 Specificity 0.886	Retrospective Dice: 0.31 Jaccard: 0.21

4 Summary & Conclusion

In this written report a comprehensive review of literature related to ICH segmentation and detection that utilised fully convolutional networks was provided. The first segment provided a brief introduction into deep learning techniques that underpin the architecture of ICH models. This was followed by delineating and comparing various studies in this field that were available at the time of writing, concluding with a table for succinct quantitative analysis.

From the cited literature only 2 papers stand out in terms of thoroughness and model performance, the updated PatchFCN model and the Ensemble model, with results rivalling that of experienced radiologists. On the basis of segmentation it is hard to compare the two as they provided different evaluation metrics, however from a practical point of view a radiologist doesn't need pixel perfect classification of image maps. A radiologists needs to identify the region on a CT slice that contains ICH and increasing the overlap percentage score has severe diminishing returns. A better metric for comparison is the detection score as a higher sensitivity and specificity provides immediate benefit to the radiologist. In those metrics both papers are fairly comparable, with a slight edge given to PatchFCN for the prospective score, which could be argued is a better indicator of generalizability. Unfortunately, there is a potential problem with direct comparison of the various architectures namely the discrepancies in the type and availability of data. The initiative done by U-Net paper (segmentation + detection) to start amassing a dataset of CT scans with labelled ground truths is a good approach to alleviate that concern. Once the dataset grows to significant size in the future (around 1000 CT scans), true comparative analysis can be done of the various architectures, much in the same way that ImageNet has spurred innovation in Computer Vision tasks.

There are a few common themes between the various architectures. Better performance could be noted when architectures seperated their models for detection (CNN) and segmentation (FCN) instead of trying to do both in a single model like in the U-Net model. Another key performance booster was incorporating superior and inferior slices to the network either as separate inputs or through interpolation, giving the model more information to discern ICH from image artefacting. The final common element in the cascaded network and the ensemble network was the utilization of slices at various window widths and levels.

Considering the comparative performance of the top models against radiologists it can assumed currently that these models are good enough to at least be an assistive tool for medical specialists for diagnosis. As already stated in the introduction there are already available commercial solutions of AI ICH detectors, however what technology underpins those models is hard to tell, as these companies have not openly shared their studies.

If given the opportunity to form my own study, I would primarily focus on obtaining a diverse set of CT scan data from various institutions, CT scanner types and age groups. The target model would implement the patch based approach of PatchFCN for context localization and employ an ensemble based approach for ICH detection. Each slice would be a feature map contain slices at various window widths and levels and the network, each slice would likewise be interpolated with the superior and inferior sllices.

References

[1] J. Alfredo Caceres and Joshua N. Goldstein. Intracranial Hemorrhage. *Emergency medicine clinics of North America*, 30(3):771–794, August 2012.

[2] Muhammad Imran Razzak, Saeeda Naz, and Ahmad Zaib. Deep Learning for Medical Image Processing: Overview, Challenges and Future. page 30.

[3] S. Albawi, T. A. Mohammed, and S. Al-Zawi. Understanding of a convolutional neural network. In *2017 International Conference on Engineering and Technology (ICET)*, pages 1–6, August 2017.

[4] Overview | Artificial intelligence for analysing CT brain scans | Advice | NICE. Publisher: NICE.

[5] Stuart Currie, Nayyar Saleem, John A Straiton, Jeremy Macmullen-Price, Daniel J Warren, and Ian J Craven. Imaging assessment of traumatic brain injury. *Postgraduate Medical Journal*, 92(1083):41–50, 2016. Publisher: The Fellowship of Postgraduate Medicine _eprint: https://pmj.bmj.com/content/92/1083/41.full.pdf.

[6] A. M. TURING. I.—COMPUTING MACHINERY AND INTELLIGENCE. *Mind*, LIX(236):433–460, 10 1950.

[7] B. Ding, H. Qian, and J. Zhou. Activation functions and their characteristics in deep neural networks. In *2018 Chinese Control And Decision Conference (CCDC)*, pages 1836–1841, June 2018. ISSN: 1948-9447.

[8] Léon Bottou, Frank E. Curtis, and Jorge Nocedal. Optimization methods for large-scale machine learning, 2018.

[9] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1(4):541–551, 1989.

[10] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, May 2017.

[11] E. Shelhamer, J. Long, and T. Darrell. Fully Convolutional Networks for Semantic Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(4):640–651, April 2017. Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence.

[12] Junghwan Cho, Ki-Su Park, Manohar Karki, Eunmi Lee, Seokhwan Ko, Jong Kun Kim, Dongeun Lee, Jaeyoung Choe, Jeongwoo Son, Myungsoo Kim, Sukhee Lee, Jeongho Lee, Changhyo Yoon, and Sinyoul Park. Improving Sensitivity on Identification and Delineation of Intracranial Hemorrhage Lesion Using Cascaded Deep Learning Models. *Journal of Digital Imaging*, 32(3):450–461, June 2019.

[13] A. Bardera, I. Boada, M. Feixas, S. Remollo, G. Blasco, Y. Silva, and S. Pedraza. Semi-automated method for brain hematoma and edema quantification using computed tomography. *Computerized Medical Imaging and Graphics*, 33(4):304–311, June 2009.

[14] Phong Duc Tong, Duong Hieu, Hien Nguyen, Trong Nguyen, Vu Nguyen Hong, Hoa Tran, and Vaclav Snasel. Brain Hemorrhage Diagnosis by Using Deep Learning. pages 34–39, January 2017.

[15] Murtadha D. Hssayeni, Muayad S. Croock, Aymen D. Salman, Hassan Falah Al-khafaji, Zakaria A. Yahya, and Behnaz Ghoraani. Intracranial Hemorrhage Segmentation Using a Deep Convolutional Model. *Data*, 5(1):14, March 2020. Number: 1 Publisher: Multidisciplinary Digital Publishing Institute.

[16] Weicheng Kuo, Christian Häne, Esther Yuh, Pratik Mukherjee, and Jitendra Malik. PatchFCN for Intracranial Hemorrhage Detection. *arXiv:1806.03265 [cs]*, April 2019. arXiv: 1806.03265.

[17] Peter D. Chang, Edward Kuoy, Jack Grinband, Brent D. Weinberg, Matthew Thompson, Richelle Homo, Jefferson Chen, Hermelinda Abcede, Mohammad Shafie, Leo Sugrue, Christopher G. Filippi, Min-Ying Su, Wengui Yu, Christopher Hess, and Daniel Chow. Hybrid 3D/2D Convolutional Neural Network for Hemorrhage Evaluation on Head CT. *AJNR. American journal of neuroradiology*, 39(9):1609–1616, September 2018.

[18] Weicheng Kuo, Christian Hne, Pratik Mukherjee, Jitendra Malik, and Esther L. Yuh. Expert-level detection of acute intracranial hemorrhage on head computed tomography using deep learning. *Proceedings of the National Academy of Sciences*, 116(45):22737–22745, November 2019. Publisher: National Academy of Sciences Section: Biological Sciences.

[19] Hyunkwang Lee, Sehyo Yune, Mohammad Mansouri, Myeongchan Kim, Shahein H. Tajmir, Claude E. Guerrier, Sarah A. Ebert, Stuart R. Pomerantz, Javier M. Romero, Shahmir Kamalian, Ramon G. Gonzalez, Michael H. Lev, and Synho Do. An explainable deep-learning algorithm for the detection of acute intracranial haemorrhage from small datasets. *Nature Biomedical Engineering*, 3(3):173–182, March 2019. Number: 3 Publisher: Nature Publishing Group.

[20] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv:1505.04597 [cs]*, May 2015. arXiv: 1505.04597.

[21] Ajay Patel, Floris H. B. M. Schreuder, Catharina J. M. Klijn, Mathias Prokop, Bram van Ginneken, Henk A. Marquering, Yvo B. W. E. M. Roos, M. Irem Baharoglu, Frederick J. A. Meijer, and Rashindra Manniesing. Intracerebral Haemorrhage Segmentation in Non-Contrast CT. *Scientific Reports*, 9, November 2019.