Transfer learning based diagnosis of brain haemorrhage

Mr. Yuk Fai Cheung 1

1 Research Associate,
Department of Electrical and Electronic Engineering
The University of Hong Kong, Hong Kong

Abstract: In this study, we deal with the question of diagnosing brain bleeding that radiologists consider a laborious procedure, particularly in the early stages of bleeding. The challenge is resolved with a deep learning strategy in which a convolutions neural network and well-known neural network (CNN) are trained to categorise brain computer tomography (CT) images into haemorrhage or nonhemorrhage images. Alexnet also has a modified new version (Alex Net-SVM). In medical image analytics and classification the objective of using the model of deep learning is to answer the key problem: can the requirement to develop CNN be removed by a suitable smoothing of a pre-trained model (transfer learning)? In addition, this study will also explore the benefits of employing SVM as a classifier rather than a three-layer neural network. We utilise the identical classification task in three deep networks; one is developed from scratch and the other is a pretrained model that was fine tuned to the brain CT haemorrhage classification task. The three networks have been trained with the same number of CT brain pictures. The investigations show that it is possible to transfer information from natural photos to the classification of medical images. Moreover, our findings have shown that the proposed modified "AlexNet-SVM" model may be used to identify the brain bleeding through an overall neural network developed from scratch, and by the original AlexNet.

1. Introduction

ICH is a blooming in the intra-cranial vault[1]. Intra-cranial hemorrhogen (ICH). Such a medical disease is usually triggered by weak blood arteries, hypertension, trauma, and drug addiction. It can have numerous subtypes, such as basal ganglia, caudate nuclide or pons in the neurological emergency of ICH. Hemorrhaging kinds are usually based on anatomic bleeding location[2].

The early and expedient diagnosis of ICH, which is frequently associated with deteriorating patients in the first hours after occurrence, is important according to the American Heart Association and the American Stroke Association[3]. The imaging mode employed in the detection of bleeding due to its high availability and speed of non-contrast head computer tomography (CT). In diagnosing acute bleeding, this model has demonstrated great sensitivity and specificity[2].

Deep learning has recently increased quickly and efficiently. In order to resolve tough medical problems, deep learning networks have demonstrated great ability to generalise [4, 5], analysis of medical images [6], detection of medical bodies [7] and detection of diseases [8]. The most efficient networks in deep networks are the evolutionary neural networks that have the principles of architectures more inspired by biology than other traditional networks[9].

Eventually, several convolutionary networks were created such as AlexNet[10], VGG-NET[11], and ResNet[12]; these deep networks were all extensively trained and regarded as state-of-the-art in image classification[11—13] in a large database called the ImageNet, Large-Scale Visual Recognition Challenge. These networks are seen as machine learning practises which can hierarchically learn features from lower to upper level through the construction of a deep input architecture.

Due to their abstractions, the growth of profoundly convoluted neural networks has motivated numerous researchers to transfer their knowledge acquired by those networks to new tasks such as the classification of medical images through their millions of
images and to benefit, in particular, from the weights they have learned.
The neural network models use entirely linked layers, which represent a neural feedback network that has been trained with the usual backbone approach. This suggests that the usual basic neural network could have the same disadvantages.

An excellent model of a neural grid is the best model for both exercises and testing data sets; a fair balance must be established between variance mistake and bias error[14]. For basic models, a high bias and a low variance show a lack of adaptation while training these models. The development of formation may allow the models to enter an area where there is a little variance or bias in more complicated neural network models; this can be seen as a good fit. However, the model might undergo a large variance and low bias, which is called overfitting, as training continues further (more complex models). This is seen in the training of a complex neural network model as a severe difficulty.

This difficulty can be alleviated by numerous approaches[15]. The approaches include early halt, penalization of weights, weight pre-training and disappearance of hidden neurons. But we need to circumvent these issues in our study by substituting the neural network of SoftMax with an SVM multiclass that is a classifier for both pretrained models. Many studies have been carried out[16–18] that seek to identify a SoftMax classification function substitute. All these research have shown that the Support Vector Machine (SVM) can be a good alternative since it can increase neural network performance marginally in comparison to the standard SoftMax.

In this article, we intend to transfer AlexNet’s knowledge to a new goal: categorising the CT brain haemorrhage in images of the blood or non-haemorrhage. In addition, the CNN and the modified AlexNet are developed from scratch in conjunction with the SVM to complete the same classification task. A further purpose of using a pre-trained model generated from scratch for a CNN for the same classification problem is to demonstrate that transfer-based learning network can better perform with not much data. It also aims to show the necessity for the deep CNN training that generally takes a long while for a big number of images to learn may be eliminated by sufficient fine tuning of a pretrained model.

In this research, the scratch-based CNN is denoted as CNN, and AlexNet is the pre-training model that uses its original AlexNet architecture, and AlexNet-SVM is the modified model.

The paper has the following structure: In Section 1 the work is shown. Section 3 provides a short overview of the core neural networks, while Section 4 introduces the concept of transfer learning, with AlexNet included.

The training of the two profound networks in which the data utilised for training are presented is discussed in section 5.3. Section 6 examines the performance of the network and compares the findings of both models. Finally, paragraph 8 is the end of the paper.

2. Literature review

In order to meet major medical issues such as segmented images[19] and disability control, revolutionary neural networks were applied[20]. A convolutionary neural system designed for the segmentation of the most frequent brain tumours, that is, glioma, has been developed by Hussain et al.[19]. A system of two networks, layered together to form a novel ILinear architecture, was proposed by the authors. This new architecture was possible for all of the proposed and related architectures to achieve the greatest outcomes. A further Abiyev and Arslan study[20] found that neural networks can also support people with disabilities. In addition, they are also employed as supporting elements.

The authors proposed an interface between the human and machine based on two neural networks intended to control mouse-by-eye motions for handicapped individuals. Their approach has been evaluated and tested by a hand-crafted data set, and results demonstrate that many previous related works have been overcome by network performance.

In addition, Helwan et al.[21] used deep learning approaches for classifying images in haemorrhage or healthy brain computer tomography [CT]. To accomplish this job writers used autoencoders and deep neural networks. The authors claimed that when trained and tested on 2527 photos, the utilised model was performed differently. It has been discovered that the autoencoder that is stacked in their study is composed of three hidden levels and other utilised networks, in which the highest classification and lowest MSE have been attained. The scientists found that it was because of the short amount of data used for
training that the likely result of this overview in the stacked auto encoder over the convolutionary neural network was that the CNN needs a large number of training examples to converge. In another study by Mahajan and Mahajan, the refined use of the watershed algorithm together with an Artificial Neural Network (ANN) for CT diagnosis in brain haemorrhage type was investigated[22] in brain haemorrhage studies. The authors of the work utilise extraction features before feeding the neural classifier images where various features have been removed using the co-occurrence gray-level matrix (GLCM). Features then were categorised by a traditional neural backback network for the identification of bleeding type. They observed that suitable approaches of image processing such as noise removal and highly segmented algorithms are required for reliable bleeding identification.

In addition, Gong et al. [23] worked on splitting brain CT images into sections where either normal or hemorrhagic region might occur. The regions without haemorrhage were considered as normal, resulting in a severely skewed dataset for photos with haemorrhage. The researchers have employed an ellipse fitting, background removal and wavelet decomposition technique for the image segmentation scheme. For this approach, weighted precision and retroactive values were about 83.6 and 88.5%.

3. Method

The method combines convolutional and recurrent neural networks (RNN) to process a high-resolution 3D input image in its entirety. A schematic overview of the network architecture is shown in Figure 2.

![Figure 1: Schematic overview of update mechanism of LSTM cell state.](image)

3.1 Long Short-Term Memory

CNNs have proven to be powerful tools for analysis of spatial image information, but are limited to processing of independent input data points or samples. Furthermore, CNNs are not originally designed for learning dependencies or connections across sequential information such as a time series. Contrarily, RNNs are capable of selectively processing and retaining sequential information in an element-wise fashion. Although, these models are in no manner limited to applications in a specific domain. However, RNNs are unable to learn long-term dependencies because the gradient diminishes when the interval between relevant input data points becomes too large [24]. Long Short-Term Memory (LSTM) was designed to overcome this problem by learning to select which information is relevant to remember [25]. Since its inception it has been modified to forget irrelevant information to prevent indefinite growth and release internal resources [26]. A schematic overview of the internal mechanism of the LSTM is shown in Figure 1.

3.2 Network Architecture

The first part of the network consists of a CNN that takes 2D axial slices as input. The CNN is comprised of five units of two layers of 3x3 convolutions and rectified linear unit (ReLU) activation functions followed by maximum pooling operations with strides of two in each direction. The number of filters is doubled before each pooling operation. A sixth unit with a single layer of a 3x3 convolutions and ReLU followed by maximum pooling reduces the feature map size to 1x1. This is followed by two fully connected layers, resulting in a feature vector of 512
elements that represents each 2D axial input slice. Each 3D input image is processed by sequentially taking 2D axial slices as input for the CNN.

The second part of the network consists of two bidirectional LSTM layers of 512 units. The full sequence of vectors representing the input image serves as input for the LSTM. The LSTM layers are followed by a softmax function for prediction of the class probabilities for the full 3D input image.

3.3 Training

To reduce the complexity of the optimization problem and achieve suitable initialization, the individual parts of the model were pre-trained, as described in Sections III-C.1 and III-C.2, before training the model end-to-end as described in Section III-C.3.

3.3.1 CNN

The CNN part of the network was pre-trained using the data described in Table 3, with the reference standard consisting of binary labels for each 2D axial slice. For the purpose of this training, the CNN part of the network shown in Figure 2 was extended with an additional dense layer with softmax function.

Positive and negative slices were equally sampled from the training data where the selection was limited to slices depicting part of the cranial cavity. Negative samples were only obtained from negatively labeled cases, to avoid erroneously presenting positive samples to the model. A maximum filter of $2 \times 2$ was applied to each slice, reducing the input dimensions to $256 \times 256$ voxels. Randomly selected training samples were processed with a batch size of 25 using Adam optimization with a learning rate of 0.0001 to minimize the binary cross-entropy loss function [28]. Data augmentation in the form of random rotations between $-25$ and $25$ degrees, mirroring over the vertical axis and random shifting in all directions between $-15$ and $15$ voxels were used to enrich the training dataset. Selected training samples had equal probability of being used in their original form or as a rotated, mirrored or shifted variant. A maximum filter with a kernel size of $2 \times 2 \times 2$ was applied to each sample to reduce the dimensionality whilst retaining high intensity hemorrhagic regions. The binary cross-entropy loss function was minimized using RMSprop optimization with an initial learning rate of 0.001 [31]. Performance during training was evaluated by calculation of classification accuracy of 500 randomly selected patches obtained from the validation dataset.

3.3.2 LSTM

The best performing model achieved during pre-training of the CNN was used as initialization for the corresponding part of the full model. The weights of the CNN were fixed and were blocked from being updated during subsequent training. The weights $W_i$, $W_f$, $W_o$ and $W_c$ were initialized using Glorot uniform initialization [29]. The recurrent weights $U_i$, $U_f$, $U_o$ and $U_c$ were initialized using random orthogonal matrices. The bias for the forget gate was initialized with a value of one, all other biases were set to zero as recommended in [30].

An equal number of positive and negative samples were randomly sampled from the training dataset described in Table 2, with the reference standard consisting of binary labels for each 3D input image. A batch of a single image was used as input during training. Random rotations between $-25$ and $25$ degrees over a randomly selected axis, mirroring over the vertical axis and random shifting in all directions between $-15$ and $15$ voxels were used to enrich the training dataset. Selected training samples had equal probability of being used in their original form or as a rotated, mirrored or shifted variant. A maximum filter with a kernel size of $2 \times 2 \times 2$ was applied to each sample to reduce the dimensionality whilst retaining high intensity hemorrhagic regions. The binary cross-entropy loss function was minimized using RMSprop optimization with an initial learning rate of 0.001 [31]. Performance during training was evaluated by calculation of the area under operating characteristic curve (ROC) on the separate validation dataset. Approximately 13,000 samples were used to achieve the best performing model.

3.3.3 End-to-End

Once the weights for the LSTMs were initialized by training, the weights of the CNN part of the model were released to facilitate end-to-end training. With the exception of a lower initial learning rate of 0.0001, the same training scheme was used as discussed in Section III-C.2. Approximately 14,000 samples were used to achieve the best performing model.
3.4 Implementation

The method was developed using an NVIDIA GeForce GTX Titan X GPU and the Keras library with Theano backend [32], [33].

4. Experimental Results

Multiple experiments were performed to assess both the impact of the training steps described in Section III-C and the addition of LSTMs on the performance of the model. Furthermore, several approaches that directly combine the output of the CNN part of the method for classification were investigated. For all experiments the best performing model was determined by evaluation on the separate validation dataset. The final models were applied to the test dataset for comparison of ROC curves.

4.1 Alternative Approaches

Following the training scheme described in Section III-C.1, the CNN produces binary classification probabilities for each 2D axial input slice. Combining the output of each slice within a 3D volume without the use of LSTMs can also provide image level classification.

The classification probabilities were produced for all slices in each case in the test dataset using the pre-trained CNN part of the method. A final classification for each case was determined by taking the maximum or 95th percentile classification probability predicted within that case. The AUCs for the maximum and 95th percentile approaches were 0.60 (95% CI: 0.54-0.65) and 0.65 (95% CI: 0.59-0.70) respectively.

Removing the final softmax function, as shown in Figure 2, enables the CNN to produce a feature vector of 512 features for each 2D axial input slice. Therefore, for each case a feature vector of $N \times 512$ is produced, where $N$ is the number of axial slices in that case. Feature vectors were created for all cases in the training and test datasets described in Table 2. All feature vectors were padded with zero values at both ends to ensure uniform shape across the datasets. A linear support vector machine (SVM) was fitted to the feature vectors of the training dataset. Applying the SVM to the feature vectors of the test dataset produced classification probabilities at an image level [34]. The AUC for the SVM was 0.90 (95% CI: 0.87-0.93). The ROC curves for the maximum probability, 95th percentile probability and SVM approaches are shown in Figure 3.

4.2 LSTM

The training schemes described in Sections III-C.2 and III-C.3 were applied to two variations of the architecture described in Section III-B. The first consisted of a single LSTM layer in addition to the CNN part of the model. The second was the full model as depicted in Figure 2 with two LSTM layers. For both architectures, the ROC curve for the test dataset was first determined after training of the LSTM layers and after end-to-end training, as shown in Figure 4. Furthermore, the full model was also trained end-to-end without use of the described training schemes for the individual parts of the model.
The area under the curve (AUC) after initialization of both parts of the model, as discussed in Section III-C.2, was 0.87 (95% CI: 0.83-0.90) for the prediction of ICH using a single LSTM layer. Increasing the number of LSTM layers did not produce a significantly different result and produced the same AUC (95% CI: 0.84-0.91). The addition of end-to-end training, as discussed in Section III-C.3, increased the AUC for the single LSTM architecture to 0.94 (95% CI: 0.91-0.96) (P<0.0001) and the full model to 0.96 (95% CI: 0.93-0.97) (P<0.0001). The difference between the two architectures after end-to-end training was not significant. Selection of an operating point aimed to maximize sensitivity, results in a sensitivity of 0.98, specificity of 0.78 and accuracy of 0.87. Statistical significance was determined using the method of DeLong et al. [35]. The choice in operating point results in identification of almost all hemorrhages with few false negative predictions, examples of which are shown in Figure 5. The average time to classification of a full 3D image was approximately 0.5 seconds.

5. Discussion

A method has been presented for the identification of ICH in 3D NCCT which combines convolutional neural networks and bidirectional LSTMs. The main contribution is that we have shown the feasibility of staged end-to-end training of a CNN based on image level annotations of high-resolution NCCTs, for accurate whole image level classification.

Several approaches using only the 2D output of the CNN were implemented. Not only the direct combination of slice-level prediction probabilities, but also image classification based on features using SVM was investigated. These approaches proved to be inferior to the proposed method that combines CNN and LSTM for contextual information integration. The method achieves a high performance, with an ROC AUC of 0.96. The use of multiple LSTM layers did not prove to significantly benefit the overall performance of the method. However, more extensive experimentation with various network architectures combining CNN and LSTM would be necessary to find an optimal configuration. An operating point was chosen for maximal sensitivity, which comes at the cost of a slightly lower specificity, because it is arguably preferable to have more false positive than false negative predictions. Although few false negative predictions were made at this operating point, the method has shown to be capable of detecting even small hemorrhages with subtle appearances that could easily be overlooked, as shown in Figure 5.

The architecture of the model allows for end-to-end training with reference standard labels given at image level. This approach incorporates maximum contextual information in the image for binary classification. This may highly simplify the manner in which data must be annotated for similar tasks in the future, resulting in an easier and more cost effective work-flow. Furthermore, such an approach can be utilized for a number of applications in neuro-imaging analysis, such as the automatic prediction of the Alberta Stroke Program Early CT (ASPECT) score for acute stroke patient triage [36]. However, such novel applications become subject to the availability of large datasets of example images with image level annotations for training the network.

The presented approach has a number of limitations. First, a maximum filter is applied to the input images for dimensionality reduction. Therefore, there is to some extent a loss of contextual information. However, for the presented application the
maximum filter retains valuable information relevant for the identification of high density hemorrhagic regions in the images. For other applications this may be replaced with strided convolutions to learn task specific relevant features. Secondly, both the CNN and LSTMs require separate training for correct weight initialization. In particular, the CNN part of the model requires a dataset with axial slice level annotations for pre-training. However, end-to-end training without pre-training has shown to be possible, but achieves inferior results in comparison. Better results may be possible, but this most likely requires extensive optimization. As both parts of the method have shown to be sensitive to the choice of optimizer and its hyperparameters. Furthermore, the search space for finding the appropriate weights for all trainable parameters is vast. Training of the individual components of the model reduces complexity of the optimization problem and therefore also reduces overall training time. Superior results are attained in a shorter time using the proposed training scheme in comparison to direct end-to-end training. Thirdly, due to hardware restrictions a limited number of network configurations were investigated. With additional resources a more thorough analysis of the impact of the number of LSTMs could be performed. Finally, due to their absence in the clinical radiology reports, the volumes and sub-types of hemorrhages used in this study could not be included for sub-analyses. The use of such additional data may improve methods developed in future work by providing information pertaining to the relation between performance and hemorrhage sub-type and volume.

The combination of CNN and LSTMs has previously been used to process sequential 2D images such as in video recognition and classification tasks [37]–[38][39]. However, a similar approach can also effectively be employed for classification of other sequential data such as electrocardiographic signals [40], [41]. In the field of medical image analysis Liang et al. presented a combined convolutional and recurrent network for the classification of focal liver lesions in multi-phase CT images [42]. The method uses LSTM and combined global and local pathways to analyze four different contrast enhancement phases of 2D axial CT slices. Other related work also show the use of the combination of a CNN with a recurrent component to leverage the spatial correlation along the z-direction in 3D medical images. As the extent of certain pathology, such as a hemorrhage, is likely to be connected over a number of axial slices, the entire image can be regarded as a sequence of related images. Shahzadi et al. propose the use of a combination of a VGG-16 network with LSTM for brain tumour classification in MRI [43], [44]. The VGG-16 component of the method was pre-trained on the ImageNet natural image dataset for weight initialization prior to transfer learning. However, the cascaded method was not trained end-to-end for fine-tuning of all weights following transfer of the VGG-16 weights. Furthermore, the method was developed with a limited dataset and shows inferior results in comparison to multiple related studies employing SVM for the same task. Feng et al. developed a combined 3D CNN and LSTM for the classification of Alzheimer’s Disease in MRI and PET data [45]. The method employs a 3D CNN and LSTM pathway for each imaging modality which are fused to form an image level classification. The input data required substantial pre-processing in the form of segmentation and registration and all images were down-sampled by a factor four in all directions to reduce memory overhead. Several methods have employed convolutional-LSTM (C-LSTM) to incorporate spatial information for segmentation and classification tasks [46]–[47] [48] [49]. In C-LSTM, matrix multiplications within the LSTM unit are replaced with convolutions to integrate local neighborhood information instead of processing a 1D input vector at each point in a sequence, as with a traditional LSTM [50]. As a consequence, the number of model parameters, the size of the receptive field for contextual information integration and the generated output differs from LSTM. For segmentation, the use C-LSTM is a more logical choice as it retains spatial information relevant to the task. Using traditional LSTM in our proposed method allows for large context integration through identification of pathology on a slice level using a CNN, followed by subsequent bidirectional analysis of the entire sequence to form a patient level classification. Therefore, the recurrent component of the model is designed for the analysis of the spatial correlation of axial information within the 3D volume, much like how a human observer would view the image.

Identification of ICH is a fundamental step in emergent clinical management of trauma and stroke patients. It has been shown
that an estimated 9% of cerebrovascular events are missed at initial presentation [51]. Furthermore, approximately 5% of subarachnoid hemorrhage cases are misdiagnosed or missed [52]. A recent study showed that junior radiologists misdiagnosed ICH in approximately 4% of cases [22]. Future research may focus on analysis of network attention for pathology localization and inclusion of hemorrhage sub-types and volumes for improved performance. An accurate and automated method for patient level classification could assist physicians in the diagnostic process and prevent misdiagnoses.

This method forms the basis for an ICH identification and localization system that could be used as an additional reader in emergency radiology. Importantly, the presented approach can be employed for not only other cerebral pathologies, but also for different anatomy. It is a method for fast 3D image classification that aims to balance the trade-off between network architecture and input image size due to current hardware restrictions.

6. Conclusion

We have presented a fast and accurate method for 3D image classification. We have proposed an implementation of the method for the identification of intracranial hemorrhage in NCCT. This forms the basis for an automated system to aid diagnosis in an emergency clinical setting. The presented method can further be utilized for different anatomy and pathology.

References:


