IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 12, 2019, accepted July 8, 2019, date of publication July 9, 2019, date of current version July 26, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2927792

Image Level Training and Prediction: Intracranial Hemorrhage Identification in 3D Non-Contrast CT

AJAY PATEL, SIL. C. VAN DE LEEMPUT, MATHIAS PROKOP, BRAM VAN GINNEKEN, AND RASHINDRA MANNIESING

Department of Radiology and Nuclear Medicine, Radboud University Medical Center, 6525 GA Nijmegen, The Netherlands

Corresponding author: Ajay Patel (ajay.patel@radboudumc.nl)

This work was supported by research grants from the Netherlands Organization for Scientific Research (NWO), the Netherlands, and Canon Medical Systems Corporation, Japan.

ABSTRACT Current hardware restrictions pose limitations on the use of convolutional neural networks for medical image analysis. There is a large trade-off between network architecture and input image size. For this reason, identification and classification tasks are commonly approached with patch or region-based methods often utilizing only local contextual information during training and at inference. Here, a method is presented for the identification of intracranial hemorrhage (ICH) in three-dimensional (3D) non-contrast computed tomography (CT). The method combines a convolutional neural network and recurrent neural network in the form of bidirectional long short-term memory (LSTM) for ICH identification at image level. A convolutional neural network is trained for the identification of ICH in axial slices. LSTM is used to analyze the sequential information obtained from slice level classifications. The method is trained end-to-end using full high-resolution 3D non-contrast CTs. At inference, it produces a binary classification with respect to the presence of ICH. A total of 1554 cranial CTs were used to train and validate the method and a separate dataset of 386 images was used for testing. Quantitative analysis showed an area under receiver operating characteristic curve of 0.96. The average time to classification was approximately 0.5 s. Classification of whole 3D images is therefore possible without the need for pre-processing.

INDEX TERMS CAD, CNN, CT, identification, classification intracranial hemorrhage, LSTM, NCCT, stroke, trauma.

I. INTRODUCTION

The use of convolutional neural networks (CNN) for image classification tasks has gathered momentum in recent years, since a highly successful implementation won the ImageNet challenge in 2012 [1]. Since then, their extensive use has expanded further into, among others, the domains of object detection, segmentation and speech recognition [2]–[4]. Due to their proven success in computer vision, the application of CNNs in the various fields of medical image analysis has also gained a vast amount of interest [5]. For specific tasks, CNNs have shown to rival or surpass expert human performance [6]–[8]. Furthermore, state-of-the-art techniques may have important influences on future clinical practice [9].

The associate editor coordinating the review of this manuscript and approving it for publication was Khalid Aamir.

Although widespread use of CNNs in medical image analysis is now the norm, development of novel methodology is hampered by limitations of the hardware available in an average research setting, where single GPU systems are common. The amount of usable random access memory on the graphics processing unit (GPU) poses restrictions on the extent of the CNN architecture and size of input images that can be used during training. CNNs are also often critically dependent on the availability of high-quality reference standards. This is a time-consuming and costly procedure, particularly if the reference standard is required at voxel level. If whole images can be used to train a model end-to-end on a single GPU, then the reference standard can be provided at image level instead of at voxel level.

Therefore, detection and classification tasks in medical image analysis are restricted to methods that employ patch or region based approaches during training as well as at inference. Such approaches often require pre-processing of

TABLE 1. Overview of related work on identification of ICH in cranial non-contrast CT. Information that has not explicitly been specified in the cited publication is indicated by -. Network architecture or training procedure that has not clearly been explained in detail is indicated by †. The size of the test set is shown as the total number of 2D slices or 3D scans taken from a specified number of unique patients that was used for quantitative evaluation of the method. For example, this work included a single scan for each unique patient for evaluation. Numbers in parentheses indicate the number of cases containing pathology in a test set also containing healthy subjects. Chilamkurthy *et al.* and Pawlowski *et al.* used multiple test datasets. *n*D refers to the dimensionality of convolutional operations. ⋄ Indicates a study that reported a different performance measure. ∗ Indicates the average number of axial slices, range 280-450.

Year	Author	Method	*n*D	Input size (slice thickness)	Test set			ROC AUC [95% CI]	Sensitivity [95% CI]	Specificity [95% CI]
					Patients	*Scans*	*Slices*			
2017	Prevedello *et al.* [11]	GoogLeNet †	2D	-	-	226 (71)	-	0.91 [−]	0.9 [0.78 − 0.97]	0.85 [0.76 − 0.92]
2017	⋄ Merkow *et al.* [12]	Ensemble of GoogLeNet-like †	2D	-	-	29,925 (-)	4,800,000 (-)	-	-	-
2017	⋄ Gao *et al.* [13]	Fusion of straightforward CNNs †	3D/2D	200x200x20 (-)	-	105 (35)	-	-	-	-
2018	⋄ Helwan *et al.* [14]	Stacked autoencoder / straightforward CNN	2D	512x512 / 227x227 (-)	-	-	385 (249)	-	-	-
2018	⋄ Grewal *et al.* [15]	DenseNet + bidirectional LSTM †	2D	250x250mm (-)	-	77 (-)	-	-	-	-
2018	Chilamkurthy *et al.* [16]	ResNet18 + Random Forest †	2D	224x224 (5 mm)	- -	21,095 (2494) 491 (205)	- -	0.92 [0.92 − 0.93] 0.94 [0.92 − 0.97]	0.90 [0.89 − 0.91] 0.95 [0.91 − 0.97]	0.73 [0.72 − 0.74] 0.71 [0.65 − 0.76]
2018	Titano *et al.* [17]	ResNet50	3D	512x512x40 (1 mm)	-	180 (60)	-	0.73 [−]	0.79 [−]	0.48 [−]
2018	Arbabshirani *et al.* [18]	Straightforward CNN †	3D	256x256x24 (-)	6374 (-)	9499 (-)	331,092 (-)	0.85 [0.84 − 0.86]	0.73 [0.71 − 0.75]	0.80 [0.79 − 0.81]
2018	Chang *et al.* [19]	Mask R-CNN-like †	3D/2D	512x512x5 (-)	-	682 (82)	23,668 (-)	0.98 [−]	0.95 [−]	0.97 [−]
2018	Sato *et al.* [20]	Convolutional autoencoder †	3D	64x64x64 (1 mm)	-	38 (22)	-	0.87 [−]	0.68 [−]	0.88 [−]
2018	Pawlowski *et al.* [21]	Bayesian convolutional autoencoder †	2D	-	- -	107 98	107 98	0.93 [−] 0.88 [−]	- -	- -
2019	Ye *et al.* [22]	VGG-like + GRU	2D	256x256x(15 - 80) (0.625 - 10 mm)	493 (194)	493 (194)	8007	>0.98 [−]	-	-
2019	This work	VGG-like + bidirectional LSTM	2D	512x512x320∗ (0.5 mm)	386 (177)	386 (177)	-	0.96 [0.93 − 0.97]	0.98 [−]	0.78 [−]

the data in the form of segmentation or candidate selection to reduce data size while retaining the diagnostically relevant parts of the full image. This may considerably increase processing time, making the method less suitable for emergent clinical applications.

In this work we present a neural network for the diagnostic task of intracranial hemorrhage (ICH) identification in 3D non-contrast CT (NCCT). Previously, we have presented a method for identification of ICH in 2D axial NCCT images [10]. Here, we show how to train a neural network end-to-end using full high-resolution 3D NCCTs, which fully utilizes a single GPU. At inference, the network is capable of predicting the binary output classification of a single 3D image at once in approximately half a second.

This task is an important step in the imaging work-up of stroke and trauma patients. A fast diagnosis can help guide work-flow, treatment decisions and surgical interventions. For example, the presence of ICH is an important contra-indication for thrombolytic treatment in acute stroke patients. We show accurate performance of the network which was trained and validated on a large dataset of 1554 cranial NCCT exams.

A. RELATED WORK

Over the past several years, the number of other publications on automated evaluation of cranial non-contrast CT exams has increased. Several studies have proposed other methods for the identification of intracranial hemorrhage in an acute setting. Although these studies utilize different datasets for the creation and validation of the methods, an overview is shown in Table 1 to offer a general comparison between methodologies and performances.

Related work from recent years concerning automatic identification of cranial pathology in CT has focused on approaches that employ CNNs. Several studies have investigated the use of convolutional autoencoders for the identification of cerebral abnormalities. Sato *et al.* [20] use a 3D approach on a limited dataset and demonstrate a low sensitivity due to small areas of pathology that are missed. Pawlowski *et al.* [21] and Helwan *et al.* [14] employ similar 2D methods but both only evaluate the method on a small set of 2D axial slices. Multiple studies using large datasets of tens to hundreds of thousands of CT exams for the development and validation of the CNN-based methods have been presented [12], [16]–[19]. Generally, there is a lack of details pertaining to the CNN architectures and training procedures that have been used for the proposed methods presented in related work. The use of natural language processing by some for the creation of a reference standard for a large dataset may be more sensitive to errors than manual labeling. Also, substantial resampling of input data may lead to a loss of diagnostically relevant image data, further introducing errors to the system. A method is presented by Grewal *et al.* [15] that uses a comparable approach to obtain image level predictions. However, the context information modeled by the method is restricted to a small 3D neighborhood and not the entire input image. Therefore, image level predictions are made based on integration of limited regional data. Furthermore,

a limited dataset of unknown composition is used to train and test the method. A similar approach combining CNN and gated recurrent units (GRU) was recently presented by Ye *et al.*. The method uses a multi-channel input consisting of different intensity window normalizations to identify ICH and classify its subtype. Similar results are reported, however the processing time of approximately 30 seconds is substantially longer than this proposed method. Unfortunately, no implementations of the methods described in Table 1 are publicly available. Therefore, a direct comparison with this work is challenging. Furthermore, only a single dataset, used by Chilamkurthy *et al.* [16], is publicly available. However, their method was trained using a separate and much larger dataset, while the publicly available dataset was used for testing. For a direct comparison with our proposed method, training data must also be made available.

II. DATA

A. PATIENT DATA AND ACQUISITION

This study was approved by the institutional ethics committee and informed consent was waived. Data was acquired by retrospectively searching our research image database for patients that had received a cranial NCCT exam in the period January 1st 2012 - December 31st 2016. All images were acquired using a Canon Aquilion One scanner with a 320-detector row and reconstructed with a FC25 or FC26 kernel. All images had 512 x 512 voxels in-plane; the number of axial slices varied in the range 280-450. The mean voxel size for the whole dataset was 0.43 x 0.43 x 0.5 mm.

B. REFERENCE STANDARD

All images were visually inspected together with the official clinical radiology report. Cases with a confirmed ICH were labeled positive. Cases where no evidence of ICH was discovered were labeled as negative. The labeled positive and negative datasets were divided into training, validation and test dataset by random selection. An overview of the separate datasets is shown in Table 2.

TABLE 2. Overview of datasets used for training, validation and test of the full model.

	Training (75%)	Validation (5%)	Test (20%)	Total
Positive	666	44	177	887
Negative	782	52	209	1043
Total	1448	96	386	1930

A subset of the training data described in Table 2 was annotated for the presence of ICH at axial slice level. This dataset consisted of 170 cranial NCCT exams, of which half were positive for the presence of ICH. For each study in the entire dataset the cranial cavity was segmented using multi-atlas registration and levelset refinement [23]. For each positive case, a bounding box was drawn by a trained observer around the connected and certain part of the hemorrhage. Each axial slice where the bounding box was present was labeled as positive. All axial slices in the negative cases where cranial cavity was visible were labeled as negative. This resulted in a large number of positive and negative samples, as shown in Table 3.

TABLE 3. Overview of data used for pre-training the CNN part of the full method. Number of positive and negative patients and labeled 2D axial slices for training and validation sets.

	Patients		Slices	
	Positive	Negative	Positive	Negative
Training	75	75	10831	22205
Validation	10	10	1258	3085
Total	85	85	12089	25290

III. 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.

A. 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.

The LSTM model can be described by the following equations:

$$\begin{aligned} i_t &= \sigma(W_i x_t + U_i h_{t-1} + b_i) \\ f_t &= \sigma(W_f x_t + U_f h_{t-1} + b_f) \\ o_t &= \sigma(W_o x_t + U_o h_{t-1} + b_o) \\ \check{C}_t &= \tanh(W_C x_t + U_C h_{t-1} + b_C) \\ C_t &= f_t \odot C_{t-1} + i_t \odot \check{C}_t \\ h_t &= o_t \odot \tanh(C_t), \end{aligned} \tag{1}$$

where the input at time point t is x_t and h_{t-1}, with x_t as the input sequence data at time point t and h_{t-1} the previous hidden state. The current cell state, C_t, is modified by contributions from the current input and previous hidden state through the input gate, forget gate, output gate and cell

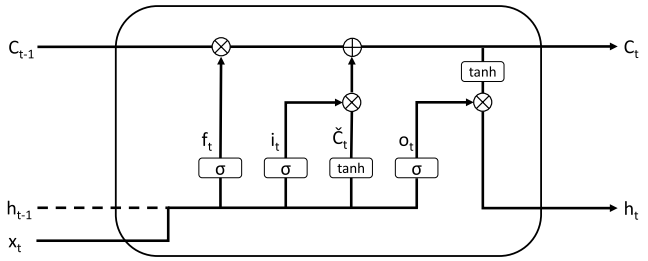

FIGURE 1. Schematic overview of update mechanism of LSTM cell state. At time point t, the current cell state C_t is modified by the current input data, x_t, and previous hidden state, h_{t-1}, through the input gate, i_t, forget gate, f_t, output gate, o_t and cell updates, $\check{C}_t$. $\otimes$ and $\oplus$ denote point-wise multiplication and addition respectively. σ and tanh are sigmoid and hyperbolic tangent functions respectively. The LSTM model is described by Equation 1.

updates, denoted as i_t, f_t, o_t and $\check{C}_t$ respectively. These four layers contribute through the weight matrices and biases, denoted as W, U and b respectively. σ and tanh are sigmoid and hyperbolic tangent functions respectively and $\odot$ is the element-wise product.

Bidirectional LSTM has shown to be more effective than unidirectional processing of sequential data [27]. In this approach, a given input with N sequence data points is processed by two recurrent networks in opposite directions, producing two separate outputs. Therefore, recurrent network A will produce N outputs h: $(h_1^A, h_2^A, \ldots, h_{N-1}^A, h_N^A)$ and recurrent network B will produce outputs H: $(H_N^B, H_{N-1}^B, \ldots, H_2^B, H_1^B)$. Finally, the summation of the outputs is taken as: $(h_1^A + H_N^B, h_2^A + H_{N-1}^B, \ldots, h_{N-1}^A + H_2^B, h_N^A + H_1^B)$.

B. 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.

C. 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.

All hyperparameters were set following pilot experiments performed on the separate validation sets described in Tables 2 and 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 2x2 was applied to each slice, reducing the input dimensions to 256x256 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 x- and y-direction 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. Dropout of 50% was applied before the first and last dense layer to prevent overfitting. Performance of the model was evaluated by calculation of classification accuracy of 500 randomly selected patches obtained from the validation dataset.

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

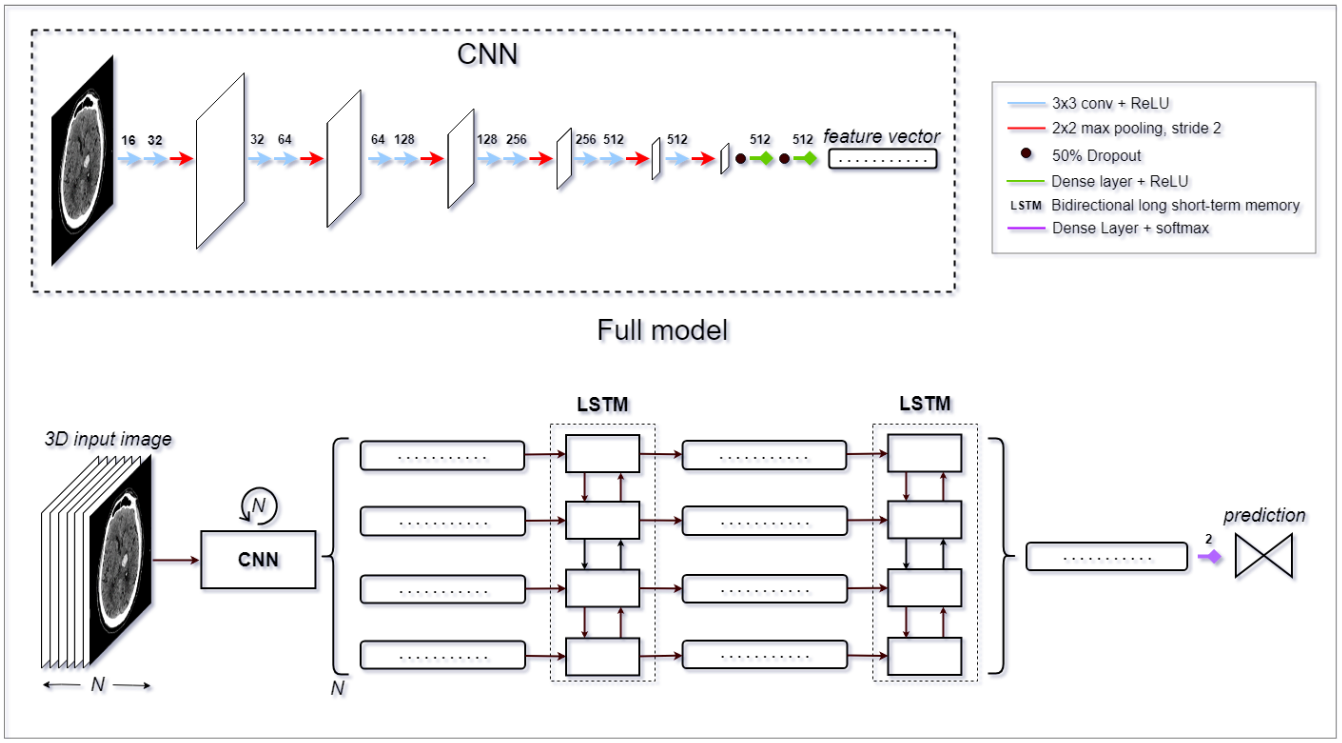

FIGURE 2. Schematic overview of the full architecture. Colored arrows represent layers of filters. Numbers define the number of filters in the corresponding layer. A full 3D input image consisting of *N* axial slices is passed to the model. Each axial slice in the image is sequentially processed by the CNN, finally producing a stack of *N* feature vectors. Therefore, *N* does not have to be explicitly set but is dependent on the input image. The feature vectors are passed to multiple bi-directional long short-term memory layers to produce a final feature vector that represents the 3D input image. The final densely connected layer with softmax produces the class probabilities for the given input image.

size of 2x2x2 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) 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.

D. IMPLEMENTATION

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

IV. EXPERIMENTS AND 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.

A. 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 *Ntimes*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.

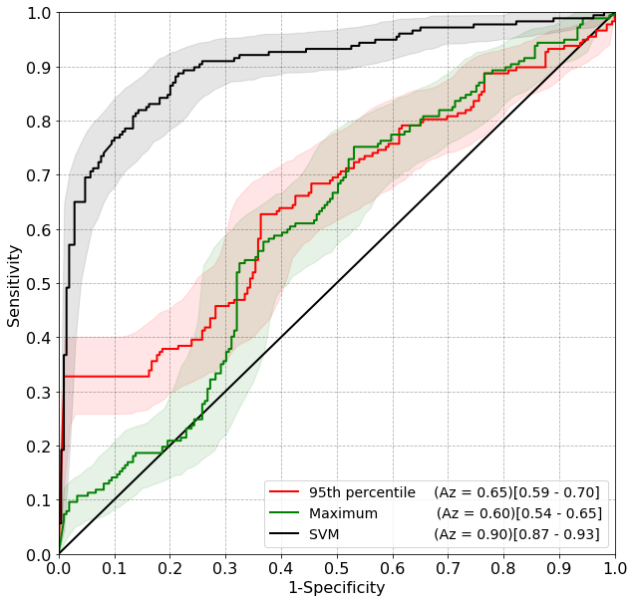

FIGURE 3. Receiver operating characteristic curves with 95% confidence interval bands for the prediction of ICH on the test dataset. ROC curve of combination of all 2D slices within each case by taking the maximum predicted probability (green) or 95th percentile (red). ROC curve of support vector machine (SVM) fitted using the feature vectors produced by the CNN (black).

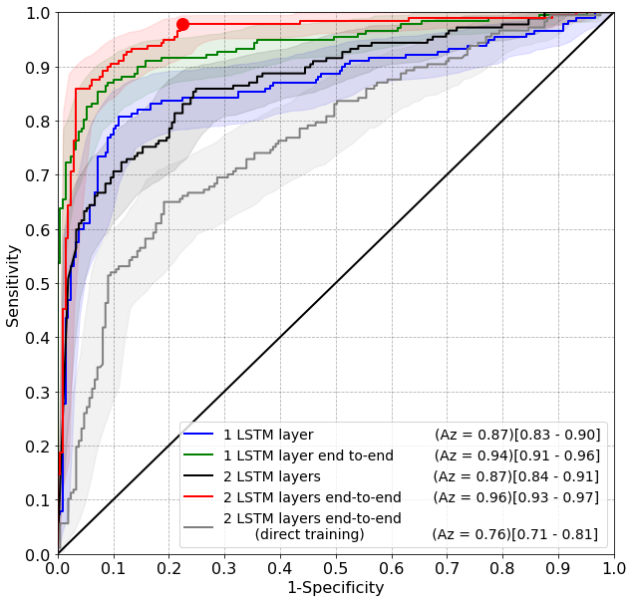

FIGURE 4. Receiver operating characteristic curves with 95% confidence interval bands for the prediction of ICH on the test dataset. ROC curve after pre-trained initialization of the individual parts of the network for models with one (yellow) or two LSTM layers (black). ROC curve after the addition of end-to-end training of the method with a single LSTM layer (green) and full model (red). ROC after training the full model end-to-end without pre-training of the individual components (gray).

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.

B. 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.

Chosen operating point results in a sensitivity of 0.98, specificity of 0.78 and accuracy of 0.87 (red circle).

V. 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

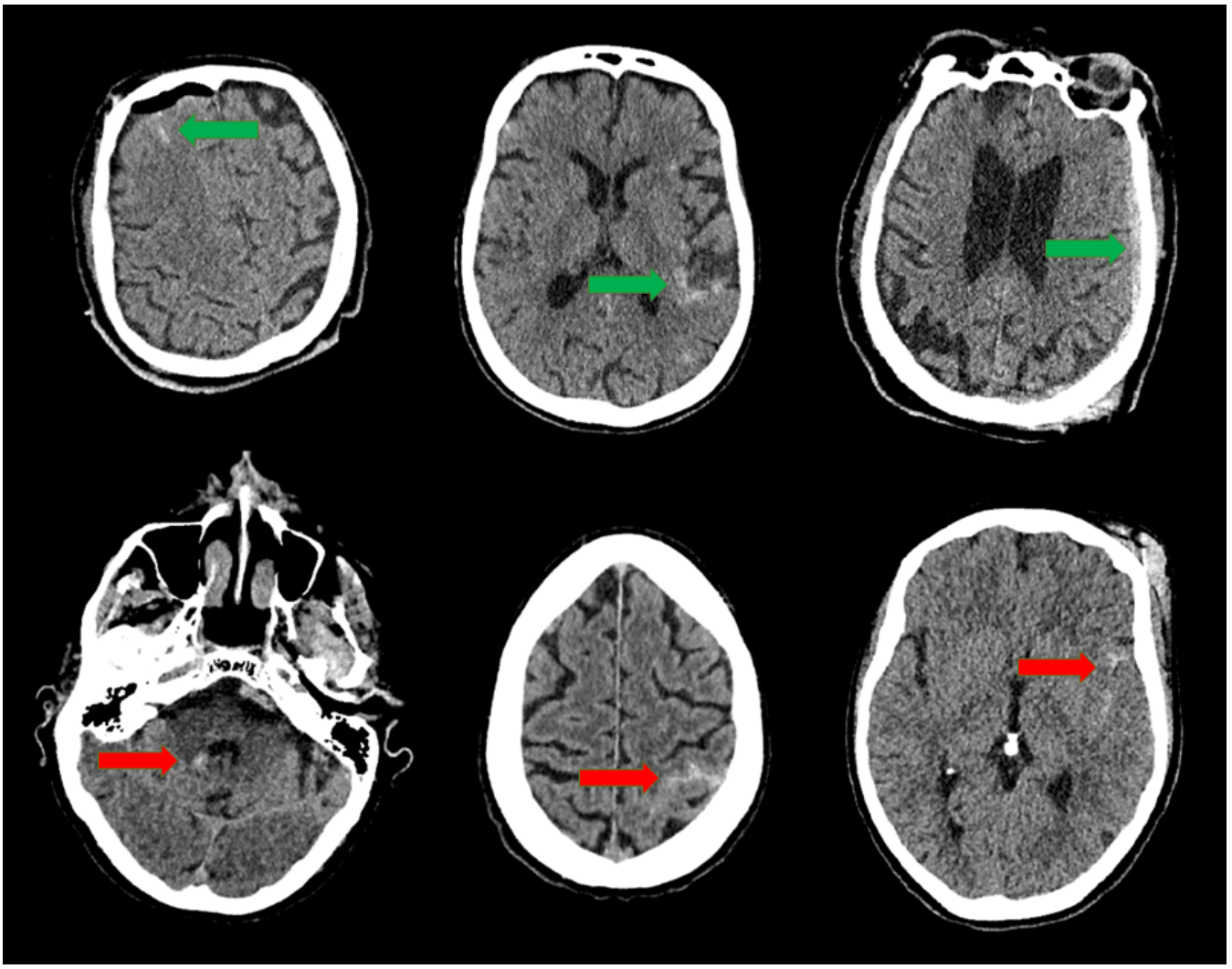

FIGURE 5. Overview of predicted results. Single axial slices shown for individual cases. True positive predictions with predicted probabilities 0.97, 0.89 and 0.98 for hemorrhages indicated with green arrows (top row, from left to right). False negative predictions with predicted probabilities 0.02, 0.08 and 0.09 for hemorrhages indicated with red arrows (bottom row, from left to right).

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]–[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]–[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.

VI. 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

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[2] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 779–788.

[3] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2015, pp. 3431–3440.

[4] G. Hinton *et al.*, "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 82–97, Nov. 2012.

[5] G. Litjens *et al.*, "anchez, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017.

[6] A. Esteva *et al.*, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, p. 115, 2017.

[7] V. Gulshan *et al.*, "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *J. Amer. Med. Assoc.*, vol. 316, no. 22, pp. 2402–2410, 2016.

[8] D. Wang, A. Khosla, R. Gargeya, H. Irshad, and A. H. Beck, "Deep learning for identifying metastatic breast cancer," 2016, *arXiv:1606.05718*. [Online]. Available: https://arxiv.org/abs/1606.05718

[9] B. E. Bejnordi *et al.*, "Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer," *JAMA*, vol. 318, no. 22, pp. 2199–2210, Dec. 2017.

[10] A. Patel and R. Manniesing, "A convolutional neural network for intracranial hemorrhage detection in non-contrast CT," *Med. Imag. Comput.-Aided Diagnosis*, vol. 10575, 2018, Art. no. 105751B.

[11] L. M. Prevedello, B. S. Erdal, J. L. Ryu, K. J. Little, M. Demirer, S. Qian, R. D. White, "Automated critical test findings identification and online notification system using artificial intelligence in imaging," *Radiology*, vol. 285, no. 3, pp. 923–931, 2017.

[12] J. Merkow, R. Luftkin, K. Nguyen, S. Soatto, Z. Tu, and A. Vedaldi, "DeepRadiologyNet: Radiologist level pathology detection in CT head images," 2017, *arXiv:1711.09313*. [Online]. Available: https://arxiv.org/abs/1711.09313

[13] X. W. Gao, R. Hui, and Z. Tian, "Classification of CT brain images based on deep learning networks," *Comput. Methods Programs Biomed.*, vol. 138, pp. 49–56, Jan. 2017.

[14] A. Helwan, G. El-Fakhri, H. Sasani, and D. Uzun Ozsahin, "Deep networks in identifying CT brain hemorrhage," *J. Intell. Fuzzy Syst.*, vol. 1, pp. 1–9, Jan. 2018.

[15] M. Grewal, M. M. Srivastava, P. Kumar, and S. Varadarajan, "RADnet: Radiologist level accuracy using deep learning for hemorrhage detection in CT scans," in *Proc. IEEE 15th Int. Symp. Biomed. Imag.*, Apr. 2018, pp. 281–284.

[16] S. Chilamkurthy, "Deep learning algorithms for detection of critical findings in head CT scans: A retrospective study," *Lancet*, vol. 392, no. 1, pp. 2388–2396, 2018.

[17] J. J. Titano *et al.*, "Automated deep-neural-network surveillance of cranial images for acute neurologic events," *Nature Med.*, vol. 24, no. 9, pp. 1337–1341, 2018.

[18] M. R. Arbabshirani, B. K. Fornwalt, G. J. Mongelluzzo, J. D. Suever, B. D. Geise, A. A. Patel, and G. J. Moore, "Advanced machine learning in action: Identification of intracranial hemorrhage on computed tomography scans of the head with clinical workflow integration," *NPJ Digit. Med.*, vol. 1, no. 1, p. 9, Apr. 2018.

[19] P. Chang, E. Kuoy, J. Grinband, B. D. Weinberg, M. Thompson, R. Homo, J. Chen, H. Abcede, M. Shafie, L. Sugrue, and C. G. Filippi, "Hybrid 3D/2D convolutional neural network for hemorrhage evaluation on head CT," *Amer. J. Neuroradiol.*, vol. 39, no. 9, pp. 1609–1616, 2018.

[20] D. Satos, "A primitive study on unsupervised anomaly detection with an autoencoder in emergency head CT volumes," *Med. Imag. Comput.-Aided Diagnosis*, vol. 10575, Feb. 2018, Art. no. 105751P.

[21] N. Pawlowski, "Unsupervised lesion detection in brain CT using Bayesian convolutional autoencoders," in *Proc. Med. Imag. Deep Learn. (MIDL)*, Amsterdam, The Netherlands, 2018.

[22] H. Ye, F. Gao, Y. Yin, D. Guo, P. Zhao, Y. Lu, X. Wang, J. Bai, K. Cao, Q. Song, and H. Zhang, "Precise diagnosis of intracranial hemorrhage and subtypes using a three-dimensional joint convolutional and recurrent neural network," *Eur. Radiol.*, vol. 1, pp. 1–11, Apr. 2019.

[23] A. Patel, B. van Ginneken, F. J. A. Meijer, E. J. van Dijk, M. Prokop, and R. Manniesing, "Robust cranial cavity segmentation in CT and CT perfusion images of trauma and suspected stroke patients," *Med. Image Anal.*, vol. 36, pp. 216–228, Feb. 2016.

[24] S. Hochreiter, "Gradient flow in recurrent nets: The difficulty of learning long-term dependencies," Piscataway, NJ, USA: IEEE Press, 2001. doi: 10.1109/9780470544037.

[25] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[26] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with LSTM," in *Proc. 9th Int. Conf. Artif. Neural Netw.*, 1999, pp. 850–855.

[27] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional LSTM and other neural network architectures," *Neural Netw.*, vol. 18, no. 5, pp. 602–610, 2005.

[28] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: https://arxiv.org/abs/1412.6980

[29] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. 13th Int. Conf. Artif. Intell. Statist.*, 2010, pp. 249–256.

[30] R. Jozefowicz, W. Zaremba, and I. Sutskever, "An empirical exploration of recurrent network architectures," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 2342–2350.

[31] T. Tieleman and G. Hinton, "Lecture 6.5-RMSPROP: Divide the gradient by a running average of its recent magnitude," *COURSERA, Neural Netw. Mach. Learn.*, vol. 4, no. 2, pp. 26–31, 2012.

[32] R. Al-Rfou *et al.*, "Theano: A Python framework for fast computation of mathematical expressions," 2016, *arXiv:1605.02688*. [Online]. Available: https://arxiv.org/abs/1605.02688

[33] F. Chollet *et al.* (2015). *Keras*. [Online]. Available: https://keras.io

[34] J. C. Platt, "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods," *Adv. Large Margin Classifiers*, vol. 10, no. 3, pp. 61–74, 1999.

[35] E. R. DeLong, D. M. DeLong, and D. L. Clarke-Pearson, "Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach," *Biometrics*, vol. 44, pp. 837–845, Sep. 1988.

[36] A. M. Demchuk and S. B. Coutts, "Alberta stroke program early CT score in acute stroke triage," *Neuroimag. Clinics*, vol. 15, no. 2, pp. 409–419, 2005.

[37] J. Y.-H. Ng, M. Hausknecht, S. Vijayanarasimhan, O. Vinyals, R. Monga, and G. Toderici, "Beyond short snippets: Deep networks for video classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 4694–4702.

[38] J. Donahue, L. A. Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko, and T. Darrell, "Long-term recurrent convolutional networks for visual recognition and description," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 2625–2634.

[39] Y. Jin, Q. Dou, H. Chen, L. Yu, J. Qin, C.-W. Fu, and P.-A. Heng, "SV-RCNet: Workflow recognition from surgical videos using recurrent convolutional network," *IEEE Trans. Med. Imag.*, vol. 37, no. 5, pp. 1114–1126, May 2018.

[40] G. Swapna, S. Kp, and R. Vinayakumar, "Automated detection of diabetes using CNN and CNN-LSTM network and heart rate signals," *Procedia Comput. Sci.*, vol. 132, pp. 1253–1262, Mar. 2018.

[41] S. L. Oh, E. Y. k. Ng, R. S. Tan, and U. R. Acharya, "Automated diagnosis of arrhythmia using combination of CNN and LSTM techniques with variable length heart beats," *Comput. Biol. Med.*, vol. 102, no. 1, pp. 278–287, Nov. 2018.

[42] D. Liang, L. Lin, H. Hu, Q. Zhang, Q. Chen, Y. lwamoto, X. Han, and Y.-W. Chen, "Combining convolutional and recurrent neural networks for classification of focal liver lesions in multi-phase CT images," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, Sep. 2018, pp. 666–675.

[43] I. Shahzadi, T. B. Tang, F. Meriadeau, and A. Quyyum, "CNN-LSTM: Cascaded framework for brain Tumour classification," in *Proc. IEEE-EMBS Conf. Biomed. Eng. Sci. (IECBES)*, Dec. 2018, pp. 633–637.

[44] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," Apr. 2014, *arXiv:1409.1556*. [Online]. Available: https://arxiv.org/abs/1409.1556

[45] C. Feng, A. Elazab, P. Yang, T. Wang, F. Zhou, H. Hu, X. Xiao, and B. Lei, "Deep learning framework for AlzheimerâĂŹs disease diagnosis via 3D-CNN and FSBi-LSTM," *IEEE Access*, vol. 7, pp. 63605–63618, 2019.

[46] M. F. Stollenga, W. Byeon, M. Liwicki, and J. Schmidhuber, "Parallel multi-dimensional LSTM, with application to fast biomedical volumetric image segmentation," in *Proc. Adv. Neural Inf. Process. Syst.*, May 2015, pp. 2998–3006.

[47] J. Cai, L. Lu, Y. Xie, F. Xing, and L. Yang, "Improving deep pancreas segmentation in CT and MRI images via recurrent neural contextual learning and direct loss function," 2017, *arXiv:1707.04912*. [Online]. Available: https://arxiv.org/abs/1707.04912
[48] J. Chen, L. Yang, Y. Zhang, M. Alber, and D. Z. Chen, "Combining fully convolutional and recurrent neural networks for 3D biomedical image segmentation," in *Proc. Adv. Neural Inf. Process. Syst.*, Sep. 2016, pp. 3036–3044.
[49] N. Braman, D. Beymer, and E. Dehghan, "Disease detection in weakly annotated, volumetric medical images using a convolutional LSTM network," Dec. 2018, *arXiv:1812.01087*. [Online]. Available: https://arxiv.org/abs/1812.01087
[50] S. Xingjian, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-C. Woo, "Convolutional lstm network: A machine learning approach for precipitation nowcasting," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 802–810.
[51] A. A. Tarnutzer, S.-H. Lee, K. A. Robinson, Z. Wang, J. A. Edlow, and D. E. Newman-Toker, "ED misdiagnosis of cerebrovascular events in the era of modern neuroimaging: A meta-analysis," *Neurology*, vol. 88, no. 15, pp. 1468–1477, Apr. 2017.
[52] M. J. Vermeulen and M. J. Schull, "Missed diagnosis of subarachnoid hemorrhage in the emergency department," *Stroke*, vol. 38, no. 4, pp. 1216–1221, 2007.

AJAY PATEL received the B.Sc. degree in biomedical sciences and M.Sc. degree in biomedical image sciences from the University of Utrecht, The Netherlands, in 2010 and 2014, respectively. In 2015, he joined the Diagnostic Image Analysis Group, Department of Radiology and Nuclear Medicine, Radboud University Medical Center, Nijmegen, The Netherlands, as a Ph.D. candidate working on computer-aided diagnosis in acute stroke.

SIL. C. VAN DE LEEMPUT received the Bachelor of Technology degree in design for virtual theater and games from the University of the Arts Utrecht, The Netherlands, in 2010, and the bachelor's and master's degrees in artificial intelligence from Radboud University, Nijmegen, The Netherlands, in 2013 and 2015, respectively. In 2015, he joined the Diagnostic Image Analysis Group, Department of Radiology and Nuclear Medicine, Radboud University Medical Center, Nijmegen, The Netherlands, as a Ph.D. candidate to work on computer-aided diagnosis in acute stroke.

MATHIAS PROKOP received the bachelor's degree in physics from Philipps-University Marburg, Germany. He is a Professor of radiology with Radboud University, Nijmegen, The Netherlands, and the Chairman of the Department of Radiology, since 2009. He came to the Netherlands in 2002, when he was an appointed Professor of radiology with UMC Utrecht in 2004. From 1998, he had been working as an Associate Professor of radiology with the University of Vienna Medical School, Austria. He trained as a Radiologist at Hanover Medical School, Germany. He is an expert in body imaging with a special focus on multislice CT and new imaging technologies. As one of the first users of the various generations of multislice CT scanners, he is working on new and improved imaging applications. In the past decade, he concentrated on chest screening with CT (cancer, cardiovascular disease, COPD) and has been a major player in the Dutch-Belgian lung cancer screening trial (NELSON). He is currently focusing on high-resolution CT perfusion imaging.

BRAM VAN GINNEKEN studied physics at the Eindhoven University of Technology and Utrecht University, and received the Ph.D. degree in chest radiography from the Image Sciences Institute on Computer-Aided Diagnosis, in 2001. He is a Professor of medical image analysis with Radboud University Medical Center, Nijmegen, The Netherlands, and chairs the Diagnostic Image Analysis Group. He also works for Fraunhofer MEVIS, Bremen, Germany, and is a Founder of Thirona, a company that develops software and provides services for medical image analysis. He has (co-)authored over 200 publications in international journals. He is an Associate Editor of the IEEE TRANSACTIONS ON MEDICAL IMAGING and a member of the Editorial Board of Medical Image Analysis. He pioneered the concept of challenges in medical image analysis.

RASHINDRA MANNIESING received the master's degree in electrical engineering from the Delft University of Technology, in 1999, and the Ph.D. degree in radiology from Utrecht University, University Medical Center Utrecht, in 2006. He is an Assistant Professor leading the neuroimaging group with the Diagnostic Image Analysis Group, Department of Radiology and Nuclear Medicine, Radboud University Medical Center, Nijmegen, The Netherlands. His research is on deep learning in stroke and trauma imaging.