

Fully Automated Sinogram-based Deep Learning Model for Detection and Classification of Intracranial Hemorrhage

Chitimireddy Sindhura¹, Mohammad Al Fahim¹,
Phaneendra K. Yalavarthy², and Subrahmanyam Gorthi¹

¹*Department of Electrical Engineering,
Indian Institute of Technology, Tirupati 517619, India and*
²*Department of Computational and Data Sciences,
Indian Institute of Science, Bengaluru 560012, India*

Abstract

Purpose: To propose an automated approach for detecting and classifying Intracranial Hemorrhages (ICH) directly from sinograms using a deep learning framework. This method is proposed to overcome the limitations of the conventional diagnosis by eliminating the time-consuming reconstruction step and minimizing the potential noise and artifacts that can occur during the Computed Tomography (CT) reconstruction process.

Methods: This study proposes a two-stage automated approach for detecting and classifying ICH from sinograms using a deep learning framework. The first stage of the framework is Intensity Transformed Sinogram Sythesizer, which synthesizes sinograms that are equivalent to the intensity-transformed CT images. The second stage comprises of a cascaded Convolutional Neural Network-Recurrent Neural Network (CNN-RNN) model that detects and classifies hemorrhages from the synthesized sinograms. The CNN module extracts high-level features from each input sinogram, while the RNN module provides spatial correlation of the neighborhood regions in the sinograms. The proposed method was evaluated on a publicly available RSNA dataset consisting of a large sample size of 8,652 patients.

Results: The results showed that the proposed method had a notable improvement as high as 27% in patient-wise accuracies when compared to state-of-the-art methods like ResNext-101, Inception-v3 and Vision Transformer. Furthermore, the sinogram-based approach was found to be more resilient to noise and offset errors in comparison to CT image-based approaches. The proposed model was also subjected to a multi-label classification analysis to determine the hemorrhage type from a given sinogram. The learning patterns of the proposed model were also examined for explainability using the activation maps.

Conclusion: The proposed sinogram-based approach can provide an accurate and efficient diagnosis of ICH without the need for the time-consuming reconstruction step and can potentially overcome the limitations of CT image-based approaches. The results show promising outcomes for the use of sinogram-based approaches in detecting hemorrhages, and further research can explore the potential of this approach in clinical settings.

Keywords: ICH, CNN, RNN, Sinograms, CT Scans, Hemorrhage Detection

I. INTRODUCTION

Intracranial Hemorrhage (ICH) is a critical and life-threatening condition leading to a high fatality rate and impairments. The causes of ICH might be direct or indirect, which include trauma (accidents), stroke, vascular malformation, high blood pressure, illegal substances, and blood coagulation abnormalities¹. It is vital to design a rapid and accurate Computer Aided Diagnosis (CAD) system that would aid neurologists or radiologists and improve the survival rate. X-ray-based Computed Tomography (CT) scan is the most often utilized imaging modality for ICH diagnosis because of its wide availability, short acquisition time, and its ability to distinguish skull fractures and hemorrhages in the brain efficiently^{2,3}.

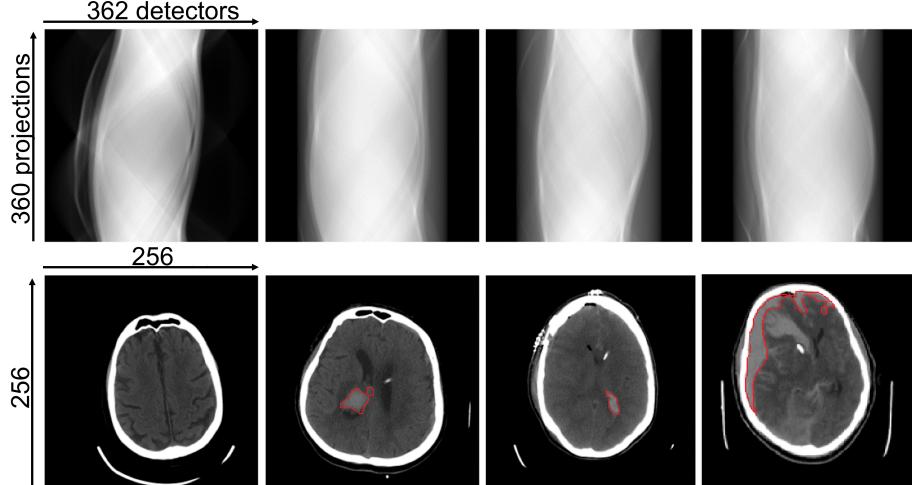


FIG. 1: A few examples of raw sinograms and corresponding CT images. The first row shows raw projections (sinograms), and the second row shows the corresponding CT images (brain window is displayed for better visualisation). The hemorrhages are contoured in red colour.

In X-ray CT imaging, a sinogram is first generated by obtaining the raw projections of an anatomical slice of interest from numerous angles. One axis of the sinogram represents the angle of the X-ray detector, while the other axis represents the distance from the rotation center in the detector row. These projections are uninterpretable for humans because the information about each spatial location is distributed throughout the sinogram. The reconstruction methods, such as analytical and iterative methods, are used to convert sinograms

into CT images to make this data understandable⁴. A few examples of raw sinograms and their reconstructed CT images are shown in Fig. 1.

Sinograms inherently contain all the information found in reconstructed CT images, even though they do not explicitly include anatomical details suitable for human interpretation. In addition, with the use of different reconstruction techniques, there is a risk of experiencing a variety of information loss, including interpolation-related resolution loss and artifacts from beam hardening, scattering, and motion⁵. Additionally, the reconstruction algorithms cause noise in the CT images depending on the technique used⁶. Sinogram acquisition is quick and simple, but reconstruction is computationally difficult and time-consuming, taking up to 10 minutes to compute^{7,8}. The above limitations of CT reconstruction approaches motivated the current proposal to use sinograms directly for ICH detection.

Convolutional Neural Networks (CNNs) have recently been investigated for medical applications due to their superior performance compared to conventional algorithms in many applications. Most of the methods in the literature use reconstructed CT images for ICH detection. Different CNN architectures are initially used to identify each type of haemorrhage^{9–12}. For the identification and classification of hemorrhages, only a few papers have recommended using 3D CNN architectures^{13–16}. The accuracy of identifying and categorizing hemorrhages in CT scans was significantly increased by combining CNN and a Recurrent Neural Network (RNN)^{17,18}.

Vision Transformer (ViT) is a recent neural network architecture for computer vision tasks that has attracted considerable interest due to its ability to capture fine-grained details and long-range dependencies in images¹⁹. ViT replaces convolutional layers with a self-attention mechanism, enabling the model to attend to different parts of the input and extract relevant features. ViT has achieved state-of-the-art performance on benchmark datasets and can be adapted to various vision tasks. Despite being a recent breakthrough in computer vision tasks, vision transformers have not been extensively studied in the medical image domain.

There have been very few studies that propose the use of sinograms to detect abnormalities. A study on the detection and characterization of blood vessels from sinograms using deep learning models is proposed²⁰. A recent study used intensity windowed sinograms created from windowed CT scans to examine the viability of using sinograms rather than CT scans to detect abnormalities²¹. CT images and sinograms are indeed different representations of the same underlying data. CT images represent tissue attenuation, while sinograms

represent accumulated attenuation along the X-ray path through the body. The windowing techniques can be applied to CT images to selectively highlight the specific regions of interest while suppressing others. However, the same approach cannot be applied to the sinogram domain. Similarly, neural networks can be employed to synthesize sinograms by preserving information relevant to specific tissues and discarding irrelevant information. To the best of our knowledge, no method has been proposed so far to synthesize a sinogram that is equivalent to a windowed CT scan. Further, RNN-based methods that consider neighborhood information have not been explored so far in the context of sinogram-based detection of abnormalities.

This work presents, for the first time, an end-to-end deep learning model for the detection of ICH directly from sinograms. The primary contribution is to propose a deep learning model to synthesize intensity-transformed sinograms from the acquired raw sinograms. The second contribution is the development of a cascaded deep learning model based on CNN and RNN to detect the presence of ICH in sinograms. This cascading model ensures that information from the adjacent slices is integrated with the current slice to reduce false predictions. The performance of the proposed model is compared with the state-of-the-art architectures. Finally, the robustness of the proposed sinogram-based approach versus the existing CT-based approach to offset and Poisson errors are also analyzed.

The rest of the paper is organized as follows: section II A presents the details of the dataset and the experimental setup used for evaluations. Section II B presents the details of the proposed method. Section III presents the results and ablation studies, as well as experiments on the robustness analysis. Section IV provides a detailed discussion of results, and finally, the conclusion is presented in section V.

II. MATERIALS AND METHODS

A. Dataset and Experimental Setup

The current study uses a subset of data from the huge publicly available dataset shared by the Radiological Society of North America (RSNA) that contains brain CT scans with intracranial hemorrhages²². Since the raw sinograms are not explicitly provided in this or any other publicly available dataset, synthetic sinograms were generated from CT scans using the

Radon transform with parallel beam geometry. To accomplish this, a CUDA implementation of the Radon transform, called Torch Radon²³, is used. Although modern CT machines use more advanced scanning geometries like helical fan beam or cone beam, this study aimed to explore the feasibility of utilizing sinograms over CT scans for hemorrhage detection regardless of the geometry used in sinogram generation. This work could be extended to other geometries in the future^{24,25}.

The dataset used for evaluation consists of a total of 361,838 CT slices (8,652 subjects). These subjects are categorized into five types of hemorrhages: Epidural Hematoma (EDH) with 170 subjects, Intraparenchymal Hemorrhage (IPH) with 2,832 subjects, IntraVentricular Hemorrhage (IVH) with 2,165 subjects, Subarachnoid Hemorrhage (SAH) with 2,810 subjects, and Subdural Hematoma (SDH) with 401 subjects. Out of the entire dataset, 267,279 slices (6,122 subjects) are used for training, 25,783 slices (765 subjects) are used for validation, and the remaining 24,786 slices (765 subjects) are used for testing. Additionally, 1000 healthy subjects are included in the testing phase for patient-wise analysis. Therefore, it roughly corresponds to grouping the entire data in the ratio of 7:1:2 among the training, validation, and testing respectively. In order to ensure that all slices from any given subject are part of only one of the three groups, the random grouping of data is made subject-wise and not slice-wise.

The dimension of each CT slice is 512×512 . All images are down-sampled by 2 to meet the memory constraints. Sinograms are generated with one projection for each degree from CT slices with dimensions 256×256 , as shown in Fig. 1. The dimension of the resulting sinograms is 360×362 . All implementations are performed on a workstation with NVIDIA Quadro RTX 8000 graphics card, and 96 GB of RAM. Adam optimizer with a learning rate of 0.1 is used for training the proposed method. The training is performed for 20 epochs for all the experiments, and the best model is saved. Each epoch took around 45 minutes on a 48 GB GPU. To have a fair comparison, all models were trained to 20 epochs and a stopping criterion where the model stops when there is no decrease in validation loss for five epochs. The training stopped at around 10-12 epochs for all experiments.

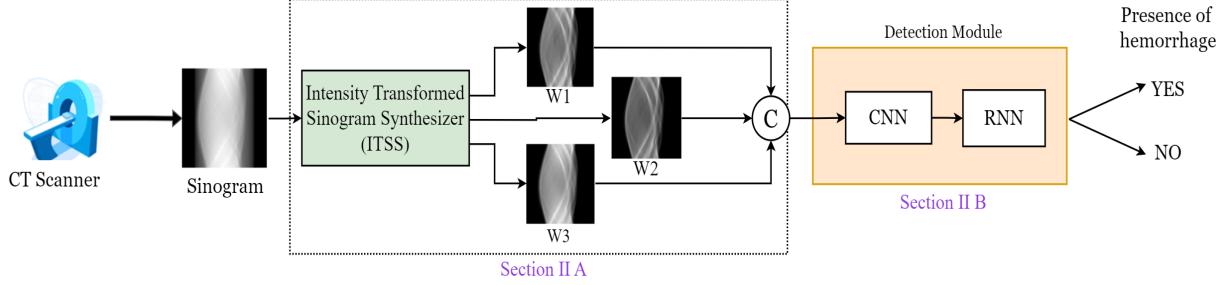


FIG. 2: The key modules of the proposed method for automated detection of intracranial hemorrhage from the sinogram data. The details of each module are given in the sections indicated in the figure.

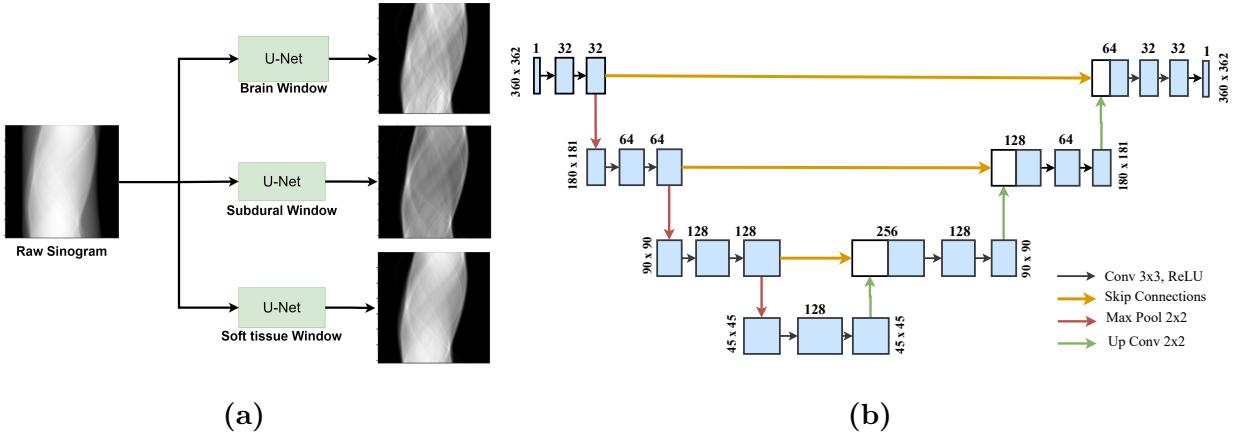


FIG. 3: The proposed Intensity Transformed Sinogram synthesizer (ITSS) Module (a) Sinogram synthesis (b) The U-Net configuration.

B. Proposed Method

The proposed sinogram-based ICH detection method is a two-stage end-to-end deep learning method and is summarized in Fig. 2. The first stage, referred to as the “Intensity Transformed Sinogram Synthesizer” (ITSS), generates intensity-transformed sinograms that correspond to the intensity windowed CT scans of the brain, subdural, and soft tissues. The second stage, referred to as the detection module, is a cascaded combination of CNN and RNN. More details of both these modules are presented in the rest of this section.

1. *Intensity Transformed Sinogram Synthesizer*

In general, the intensity range of a CT image is [-1000, 1000] Hounsfield Units (HU)²⁶. In order to visualize or highlight a specific hemorrhage, a few intensity windows are explicitly chosen instead of dealing with the whole range of intensities. Radiologists typically focus on three different ranges of HU values to detect various types of intracranial hemorrhages in head CT scans: [0, 80] HU for the brain, [80, 200] HU for the subdural, and [40, 380] HU for the soft tissue. Notice that the windowing operation is a non-linear transform due to the clipping of HU values within the specified lower and upper limits. Application of such windowing to the CT scans improves the contrast between the regions of interest and the rest of the image. Further, it has been demonstrated that the use of windowed CT scans instead of the original CT scans significantly improves the diagnostic accuracy of CT-based deep learning methods¹⁸.

However, in the case of sinograms, the intensity windowing cannot be directly applied for the reasons discussed in the preceding section.

There are no works so far attempting to synthesize sinograms that are similar to windowed CT scans. One of the key objectives of this paper is to propose a model that can synthesize intensity-transformed sinograms without the need for the CT reconstruction phase. To this end, we propose a U-Net-based ensemble deep learning model to synthesize sinograms corresponding to the three intensity windows mentioned above.

U-Net architecture²⁷ is one of the widely used encoder-decoder models in medical imaging. It was initially proposed for performing segmentation; it was later extended to other applications like image denoising²⁸, image restoration²⁹, and image super-resolution³⁰. Inspired by its success in other applications, we propose a U-Net-based ensemble architecture for synthesizing intensity-transformed sinograms.

Figure 3 shows the proposed ensemble architecture of the ITSS module. Each branch of this module takes the raw sinogram as an input and outputs a sinogram equivalent to the specific windowing of the brain, subdural and soft tissue. To this end, the baseline U-Net model is modified to have three-level encoder and decoder blocks with [32, 64, 128] filters in each level. Each encoder level consists of two consecutive sequences of 3×3 Convolution, Rectified Linear Unit (ReLU) activation, and batch normalization. It is followed by a 2×2 max-pooling, where the spatial dimensions of feature maps are reduced by half. In a similar

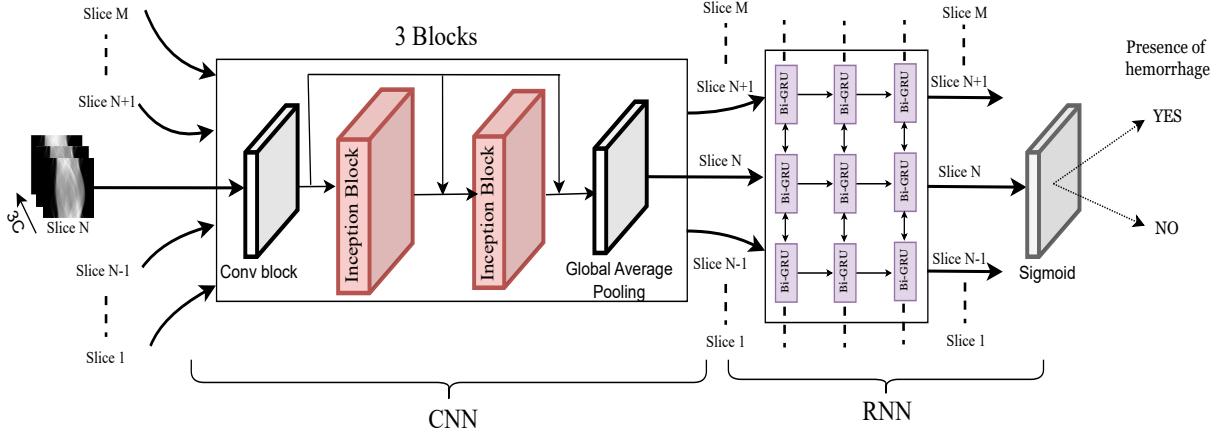


FIG. 4: The proposed model for the detection of ICH. The architectural details of the inception block are presented in Fig. 5.

manner, the decoder block has 2×2 transpose convolution at first and is concatenated with the corresponding feature map from the encoder block through a skip connection. The inclusion of skip connections addresses the challenge of retaining features from earlier layers, which can be lost as the network becomes deeper. Subsequently, two consecutive sequences of 3×3 convolutions are applied, concluding with a ReLU activation function. To enhance network generalization, a dropout layer is added at the end of each stage.

In order to generate the target sinograms (ground truth) for the training phase, a CT image is first reconstructed from the raw sinogram. The three windowed CT images corresponding to the brain, subdural and soft tissue are obtained by applying appropriate windowing to the generated CT image. Finally, the sinograms corresponding to the above three windowed CT scans are generated through the Radon transform. These three resulting sinograms are considered as the target images to be synthesized by the proposed network for the given input raw sinogram. Notice that the process mentioned above is needed only during the training phase for the ground truth generation. Thus after the training phase, the windowed sinograms are directly synthesized without performing CT image reconstruction, intensity windowing on the CT image, and the application of Radon transforms.

The proposed loss function is a weighted combination of Mean Structural SIMilarity (MSSIM) and Mean Square Error (MSE) appropriate for generating the transformed sinograms and is given by:

$$\text{Loss} = \Gamma * \text{MSSIM} + (1 - \Gamma) * \text{MSE}, \quad (1)$$

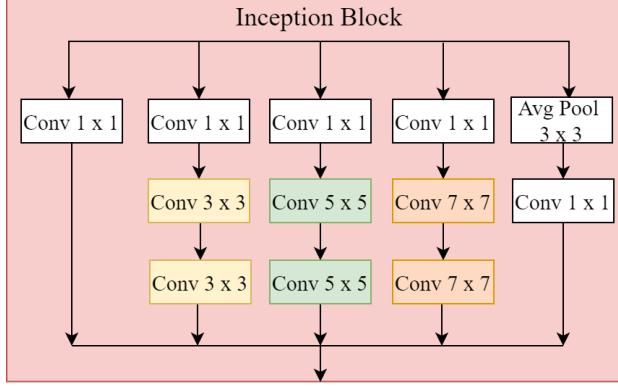


FIG. 5: The detailed architecture of the inception block mentioned in Fig. 4.

where the value of Γ is empirically set to 0.75. The MSSIM is a perceptual metric that quantifies visual perception through luminance, contrast, and structure across all pixels between the synthesized and the target images³¹. The mathematical expression for computing MSSIM is given by:

$$\text{MSSIM} = \frac{1}{M} \sum_{i=1}^M \frac{(2\mu_i^T \mu_i^S + C_1)(2\sigma_i^T \sigma_i^S + C_2)}{(\mu_i^T + \mu_i^S + C_1)(\sigma_i^T + \sigma_i^S + C_2)}, \quad (2)$$

where μ_i^T and μ_i^S are the local means, σ_i^T and σ_i^S are the standard deviations computed on a 11×11 window. The subscript i indicates the pixel number, while the superscript indicates whether the value is computed for the target image (T) or the synthesized image (S). C_1 and C_2 are the constants included to avoid instability of the fraction. Specifically, $C_1 = (K_1 L)^2$ and $C_2 = (K_2 L)^2$ where L is the dynamic range of the pixel values (255 for 8-bit gray-scale images), $K_1 = 0.01$ and $K_2 = 0.03$.

MSE computes the square of the intensity differences between the synthesized and the target images averaged across all pixels and is given by:

$$\text{MSE} = \frac{1}{M} \sum_{i=1}^M (I_i^T - I_i^S)^2 \quad (3)$$

where I_i^T and I_i^S are respectively the intensity values of the i^{th} pixel in the target and synthesized images.

The sinograms synthesized by this ITSS module are given as inputs to the deep learning-based ICH detection module described in the following subsection.

2. Detection Module

The proposed module aims to improve the accuracy of ICH detection by not only extracting the appropriate high-level features but also by incorporating the relevant information from the neighborhood slices. To this end, this paper proposes a cascaded CNN-RNN architecture as shown in Fig. 4.

The CNN part of the proposed architecture is inspired by the inception block that was first proposed in GoogleNet³². The architecture of the inception block is shown in Fig. 5. It contains a convolutional block followed by three dense inception blocks. The basic idea of the inception block is to extract information at different scales by convolving with kernels of different sizes and fuse the resulting feature maps to obtain better representations of the image. Each inception block has three convolutional kernels 3×3 , 5×5 , and 7×7 . Two such blocks are combined together and termed as dense inception block. Each dense inception block is followed by a batch normalization layer and a leaky ReLU activation function.

Notice that the CNN part of the proposed architecture provides the probability of the presence of hemorrhage based on the information of the current slice alone. The false positives obtained in the individual slice-wise predictions can be reduced further by additionally incorporating the information from the adjacent slices through the feature vector obtained at the last layer of the CNN. For this purpose, the CNN module is cascaded with an RNN module. It can be noted from Fig. 4 that while the CNN module extracts one feature map for each slice, the RNN module is fed with the feature maps extracted from all slices of a patient.

Gated Recurrent Units (GRU) are one of the widely used RNNs that were initially proposed for language processing tasks³³ and later extended to image processing tasks³⁴ due to their improved accuracies and computational efficiency. Since the slices above and below the current slice of interest can potentially share information about the hemorrhage in the current slice, we use Bidirectional GRU (Bi-GRU)³⁵. The last convolutional layer of the CNN gives a feature vector of size 120 for each slice. The feature vectors of all slices of a patient’s scan are then fed to the three Bi-GRU layers as illustrated in Fig. 4. It is followed by a dropout layer of 0.2 and a sigmoid activation function for binary classification. The number of filters in the three Bi-GRU layers is set to 256. Inspired by the existing deep learning methods for detection, we chose to use binary cross-entropy between the predicted and the

TABLE I: Quantitative evaluation of ITSS module on slices with and without hemorrhages for three windows.

Windowing Type	Slices With Hemorrhages			Slices Without Hemorrhages		
	MSE	PSNR (dB)	MSSIM (%)	MSE	PSNR (dB)	MSSIM (%)
Brain	0.0067	22.90	93.53	0.0041	26.95	93.01
Subdural	0.0221	17.20	90.36	0.0120	21.74	90.72
Soft Tissue	0.0023	28.29	96.80	0.0014	31.95	96.28

ground truth labels as the loss function. The Bidirectional Long Short-Term Memory (Bi-LSTM) is another widely used RNN model in the literature³⁶. Bi-LSTM is also evaluated in this paper in place of Bi-GRU during the ablation study presented in section III D.

III. EXPERIMENTS AND RESULTS

In this section, we first present the results of each module of the proposed method, followed by the ablation studies. Then we present a study of the robustness of sinogram-based detection and CT-based detection of abnormalities to offset errors and to the Poisson noise. We also present a study of the adaptability of the proposed model from parallel-beam geometry to cone-beam geometry. Finally, we examine the interpretability of the deep learning model in learning the patterns of sinograms.

A. Results of ITSS Module

The ability of the proposed ITSS module to synthesize the sinograms is evaluated both qualitatively and quantitatively by comparing the results with the ground truth sinograms generated directly from the windowed CT scans. Figure 6 visually presents a sample result for the synthesized sinograms, ground truth, and the errors in the synthesized sinogram for all three windows of the brain. It can be noted that the sinograms generated from the proposed synthesizer are very similar to the ground truth sinograms.

In order to assess the generality of the synthesizer, the quantitative evaluation is performed separately on slices with and without hemorrhages. MSE, PSNR and MSSIM are

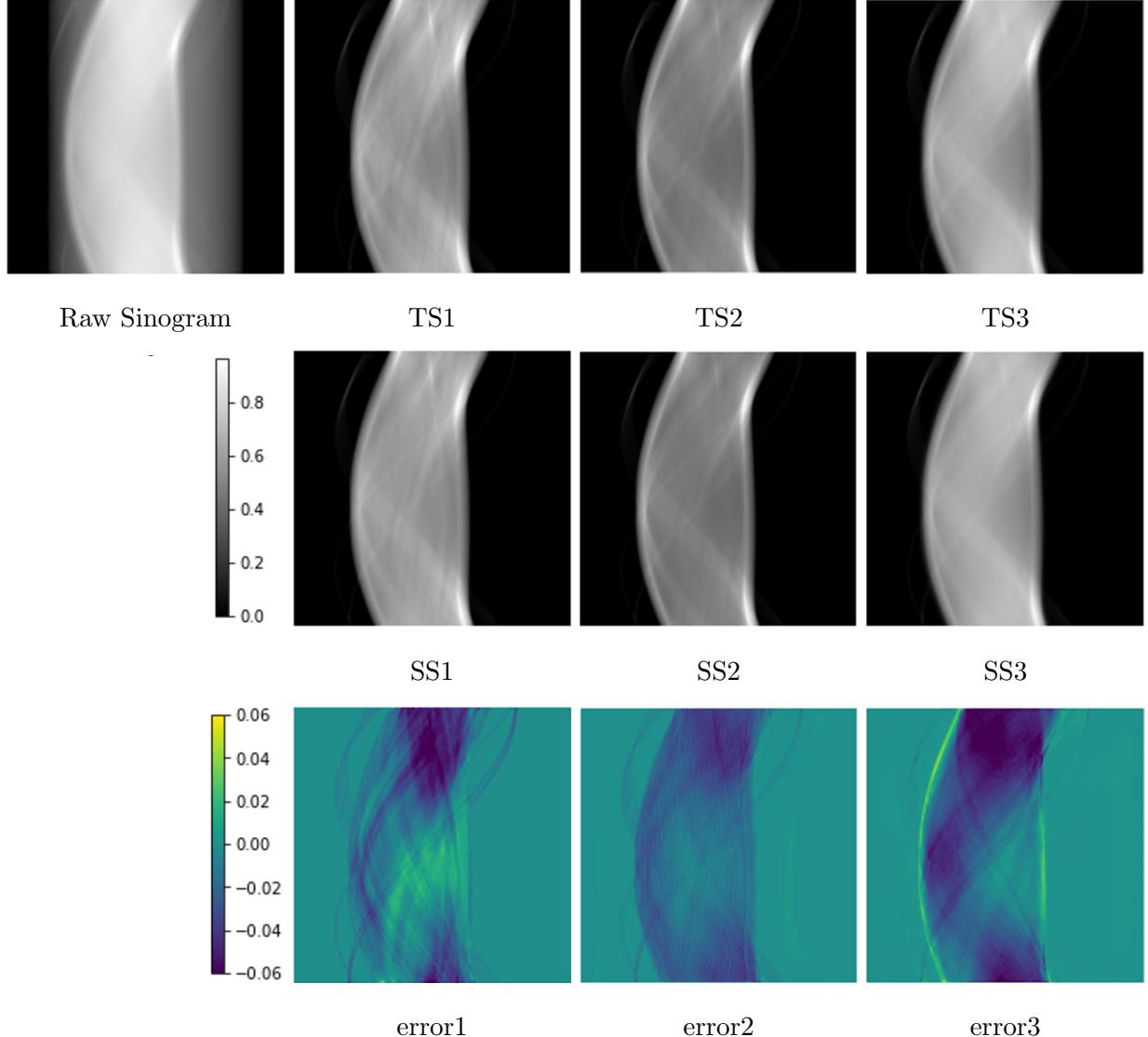


FIG. 6: Sample result of ITSS module: Target Sinograms (TS), Synthesized Sinograms (SS) and the error are presented in rows one, two and three respectively.

used for the evaluation. As discussed in section II B 1, the higher the values of MSSIM and PSNR, the better the accuracy of the synthesized sinograms. Similarly, small values of MSE indicate more accurate results. Table. I presents the results for all three windows. It can be noted that the proposed synthesizer performed well in terms of all three metrics with MSE in the range of 0.1-2.2%, PSNR of 17-32 dB, and MSSIM of 90-97%. It can also be noted that the proposed module synthesized the sinograms with and without hemorrhages equally well.

TABLE II: Evaluation of the proposed and state-of-the-art detection methods. Patient-wise and slice-wise results are mentioned outside and inside the brackets, respectively.

	Accuracy (%)	Sensitivity (%)	Specificity (%)
ResNeXt-101 ³⁷	55.73 (89.45)	98.86 (70.87)	26.42 (92.64)
Inception-V3 ³²	68.70 (91.71)	98.14 (68.29)	48.69 (95.73)
ViT ¹⁹	57.22 (85.04)	94.71 (98.18)	26.90 (94.33)
Proposed Method	95.50 (94.30)	93.16 (87.81)	97.10 (96.83)

B. Results of Detection Module

The performance of the proposed detection module was compared with other popular deep learning models: ResNeXt-101³⁷, Inception-v3³² and Vision Transformer (ViT)¹⁹. ResNeXt is a part of the ResNet family of neural network models. It is known for its increased expressiveness and accuracy, which is achieved by using Split-Transform-Aggregate blocks that allow complex feature representations. Inception-v3 architecture, on the other hand, is designed to handle multi-scale information in the input data and is computationally efficient for image classification tasks. ViT is a recent neural network model that uses self-attention mechanisms to process input data, and it has shown promising results for image classification and other vision tasks. The inputs to all these methods are the synthesized sinograms obtained from the ITSS module.

Table II presents both slice-wise and patient-wise quantitative results for the proposed and existing methods. It can be noted from slice-wise evaluations that the proposed method has resulted in an accuracy of 94.3%, with an improvement of around 3% compared to the best result from the existing methods. Similarly, the slice-wise sensitivity and specificity for the proposed method are 87.8% and 96.8% with an improvement of around 19% and 1%, respectively, compared to the best results from the existing methods.

Notice from the patient-wise evaluation results in Table II that the proposed method achieves an overall accuracy of 95.5%, showcasing an improvement of approximately 27% compared to the best result from the existing methods. The sensitivity and specificity values for the proposed method are respectively 93.2% and 97.1%. While there is a decrease in sensitivity by around 5.7% and 1.5% compared to the ResNeXt-101 and ViT, respectively,

the specificity has increased by around 70.7%. Similarly, while there is a decrease in sensitivity by around 5% compared to the Inception-V3, the specificity has increased by around 48.4%. This indicates that the proposed method is effective in detecting the condition in individual patients and can help minimize the risk of false positives and false negatives. The significant improvements in the overall accuracy and specificity of the proposed method can be attributed to the drastic reduction in the false negatives due to the cascading of the RNN network with the CNN network. It can be noted that while Vision Transformers (ViTs) have achieved success in various computer vision applications, they did not perform well when applied to sinogram data. This is because sinogram data lacks spatial information, making it difficult for the vision transformer to capture spatial relationships between different parts of the image. In contrast, CT images contain spatial information and are better suited for vision transformer.

C. Multi-label Classification of hemorrhages

It is often observed that multiple hemorrhages of various types can be present in the brain CT scans. This is known as a multi-label classification problem. Our proposed model is trained to classify five types of hemorrhages like Epidural Hematoma (EDH), Intraparenchymal Hematoma (IPH), Intraventricular Hematoma (IVH), Subarachnoid Hemorrhage (SAH), and Subdural Hematoma (SDH). The training dataset includes samples for each of the five types of hemorrhages; however, the number of samples with EDH is much lower than those of the other classes. The following minor modifications are made to our proposed model in order to adapt it as a multi-label hemorrhage classifier. The dense layer at the end is adjusted to have six units corresponding to the number of hemorrhage classes. Additionally, the loss function is changed to class-weighted binary cross-entropy, which takes into consideration the class imbalance in the training data. These modifications allowed our model to classify multiple types of hemorrhages effectively. Alternative loss functions, such as focal loss, can also be considered to improve performance further.

Table III shows the performance of a proposed model for classifying different types of hemorrhages based on metrics like accuracy, sensitivity, specificity, precision, and AUC. The overall performance of the proposed approach for detecting any ICH type was reasonably good, with an AUC of 80.69% and accuracy of 76.92%. However, when considering

TABLE III: Multi-label classification results for all intracranial hemorrhage subtypes in terms of accuracy, AUC, sensitivity and specificity on the test set for the proposed model

ICH type	AUC	Accuracy	Sensitivity	Specificity
Any	80.69	76.92	86.05	75.34
EDH	55.73	99.45	11.76	99.70
IPH	65.49	96.34	31.40	99.60
IVH	78.12	96.44	58.36	97.88
SAH	74.11	85.75	61.24	86.97
SDH	78.21	88.67	66.11	90.31

individual ICH types, the results varied widely. The results show that the model performs well for detecting IPH, IVH, SAH, and SDH, with AUCs ranging from 65.49% to 78.21% and accuracies ranging from 85.75% to 96.44%. The performance of the model for detecting EDH is suboptimal, with an AUC of 55.73% and a sensitivity of only 11.76%. This is primarily due to class imbalance present in the training data. Overall, the results suggest that the deep learning-based approach shows promise for detecting and classifying different types of ICH.

D. Ablation Studies

We first analyze the detection accuracy when using a raw sinogram versus using synthesized sinograms. Secondly, the impact of cascading an RNN network to the CNN network in the proposed detection module is studied. Finally, the effect of using Bi-LSTM in place of the currently used Bi-GRU is also evaluated.

Figure 7 presents ROC curves for all the aforementioned combinations, and the following observations can be made from them. The use of three-channel synthesized sinograms with a CNN detection module (denoted as “CNN+3C” in the figure) has significantly improved the results compared to feeding the same detection module with the raw sinograms (denoted as “CNN+raw”). The best results are obtained by cascading the CNN network with Bi-GRU and feeding it with the 3-channel synthesized sinograms (denoted as “CNN+3C+Bi-GRU”).

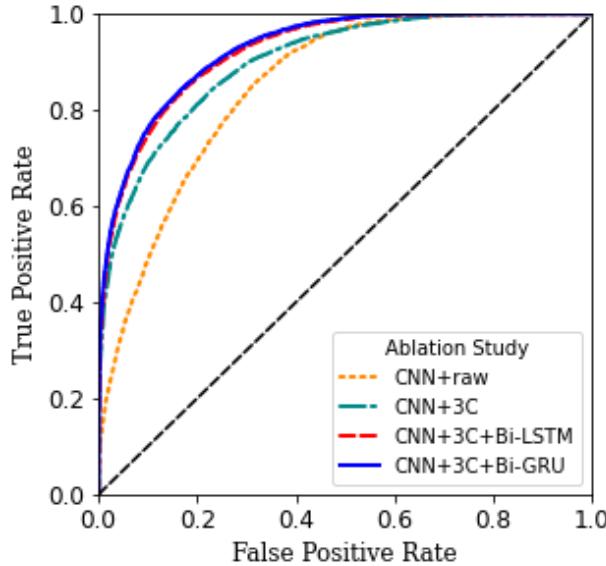


FIG. 7: Ablation Study: ROC curves are presented for varying numbers of channels (1C, 3C) and different Recurrent Neural Networks (Bi-LSTM, Bi-GRU).

TABLE IV: Robustness analysis of CT and sinogram-based detection to Poisson noise and offset errors in terms of accuracy.

	CT-based Detection	Sinogram-based Detection
Initial Result	94.41%	94.30%
Offset Error	92.20% (2.21% ↓)	94.15% (0.15% ↓)
Poisson Noise	91.55% (2.86% ↓)	94.20% (0.10% ↓)

The use of Bi-LSTM (denoted as “CNN+3C+Bi-LSTM”) in place of Bi-GRU has given almost similar results, with slightly better results with Bi-GRU.

E. Robustness Analysis

This subsection evaluates the robustness of sinogram-based detection versus CT-based detection to the offset errors in the projection angles and also to the Poisson noise in the image acquisition. A brief description of these two errors is first presented and is followed by the evaluation results.

Offset Error: Offset error in sinograms refers to systematic errors in the measurement of

the position of the detected photons or the projection angles. This occurs due to mechanical or electronic misalignment of the detectors or the gantry, and it can result in a shift or tilt in the sinogram data. During the CT reconstruction phase, offset errors can result in artifacts such as ghosting, streaking, or blurring of the image³⁸.

Error due to Poisson Noise: There is usually noise added during the sinogram acquisition due to random fluctuations in the number of photons received by the detector and is typically modeled with Poisson distribution³⁸. These noise levels can be higher, particularly in regions with lower signal intensities.

Notice that the proposed sinogram-based detection module can also be adapted to perform CT-based abnormality detection by simply feeding it with windowed CT scans instead of sinograms. Hence, in order to compare the robustness of sinogram versus CT-based approaches, the same detection module is first trained separately with the noise-free sinograms and the corresponding CT scans. Later, for testing the sinogram-based detection module, two test sets of sinograms are generated by introducing an offset of 0.1°, and by adding a Poisson noise with a mean of 0.1. The CT images reconstructed from those two sets of sinograms are used as testing data for CT-based detection.

Table IV summarizes the robustness analysis results for both approaches. It can be noticed that the change in detection accuracy in the sinogram-based approach is very minimal, with less than 0.5% for both offset and Poisson errors. In the case of CT based approach, the detection accuracies are reduced significantly by 2% for Poisson noise and 3% for offset error.

F. Extension to Fan-beam Geometry

The evaluations performed so far are on the sinograms generated from the parallel-beam geometry. The objective of this subsection is to analyze the adaptability of our proposed network trained on sinograms from parallel-beam geometry to fan-beam geometry. The equivalent sinograms for the testing data associated with fan-beam are generated using the publicly available Tomographic Iterative GPU-based Reconstruction (TIGRE) toolbox³⁹.

The network that was originally trained on parallel-beam geometry is directly used without any additional training for evaluating the detection accuracy on fan-beam geometry-based test data. It is found that the accuracies obtained for fan-beam geometry are com-

parable to those achieved with parallel-beam geometry. This result is as expected since sinograms associated with both parallel-beam and fan-beam are very similar in terms of their image characteristics. The conversion of fan-beam to parallel-beam requires rearranging of the rays. Thus, both of them contain the same information.

The CT machines have recently evolved towards cone-beam geometry. While fan-beam CT utilizes a fan-shaped X-ray beam and rotating detectors, cone-beam CT uses a cone-shaped X-ray beam and rotating detectors. The primary difference between these geometries is in the projection dimension: fan-beam CT captures 1D projection data, whereas cone-beam CT acquires 2D projection data during the scanning process. Notice that although cone-beam geometry-based CT acquisition is made, its equivalent parallel-beam geometry sinograms can be obtained through appropriate rebinning of data, as discussed in ^{40,41}. Hence, the proposed model that is trained on parallel-beam geometry can still be adapted to cone-beam geometry-based image acquisition with appropriate rebinning of data and training.

G. Interpretability

This subsection provides the results of interpretability by showing that the attention of the proposed method is towards the most relevant regions of the sinogram that highlight the hemorrhage to enable accurate automated detection of ICH. To this end, the Gradient-weighted Class Activation Map (Grad-CAM)⁴² obtained in the penultimate layer of the CNN architecture is visually analyzed.

The first and second rows of Fig. 8 respectively present sample results for intraventricular hematoma and intraparenchymal hematoma. The raw sinograms and the windowed CT scans with injuries overlaid are respectively shown in columns (a) and (b). The sinograms generated for the spatial masks containing those hemorrhages alone are presented in column (c). The synthesized sinograms obtained from the ITSS module are shown in column (d). Finally, column (e) presents Grad-CAM heat maps obtained from the penultimate layer of the CNN architecture. It can be noticed by comparing results in columns (a), (c), and (d) that, in the synthesized sinograms, a distinct contrast can be seen in the regions associated with the hemorrhages, which are otherwise not visible in the raw sinograms. It can also be noted from columns (c) and (e) that the proposed deep learning method is indeed able to automatically give more importance (weight) to the regions in the sinograms that correspond

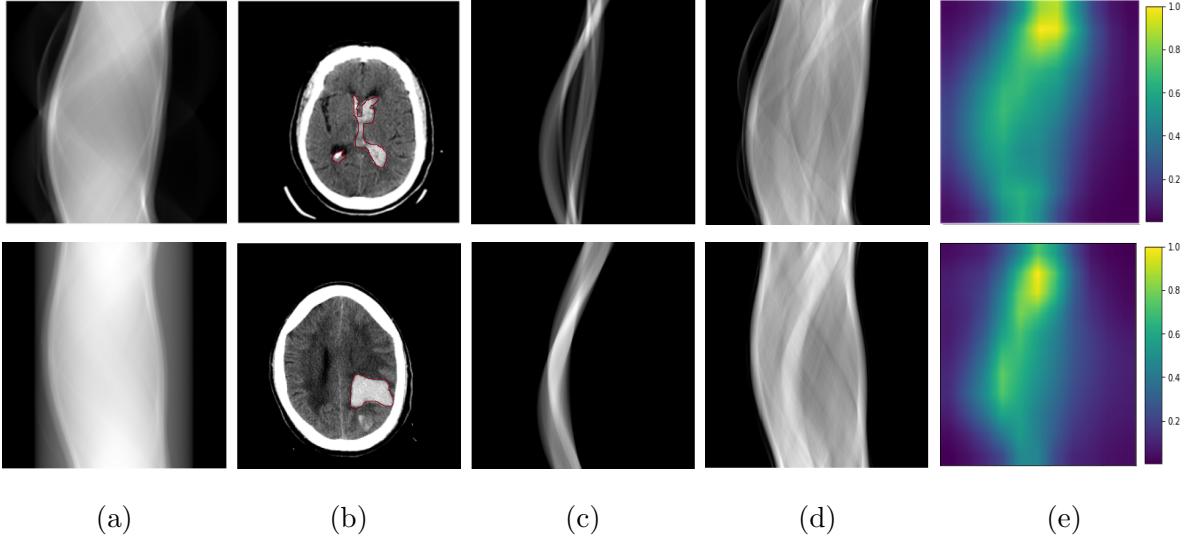


FIG. 8: Interpretation of the detection results on sinograms through Grad-CAM heatmaps. Each row presents results for different types of hemorrhages. Column (a) shows the raw sinograms. Column (b) shows the reconstructed CT images with the contours of hemorrhages overlaid. Column (c) illustrates the sinograms generated for the hemorrhage mask alone. Column (d) shows synthesized sinograms. Column (e) shows the Grad-CAM heat-maps.

to a given hemorrhage.

IV. DISCUSSION

Deep learning-based analysis of images for various tasks of detection and classification of pathologies, mimicking the function of a radiologist, has seen tremendous growth in the last few years. These data-driven methods that use reconstructed images have been largely successful due to the latent space representation offered by neural networks. Latent space is an important idea since “deep learning” relies on it to understand data characteristics and simplify data representations to identify patterns. The power of neural networks lies in their ability to create this latent space and perform the task. The developments so far have relied on image data, which requires an additional step of image reconstruction after sinogram acquisition. The image reconstruction step can take as much as ten minutes, even though the image acquisition can be completed in less than two minutes. As sinograms are challenging for humans to interpret, our method uses neural network-based analysis for ICH

detection directly from sinograms. It mainly consists of two modules: Intensity Transformed Sinogram Synthesizer (ITSS) and detection module.

ITSS module is a novel contribution where a U-Net-based deep learning model is proposed to synthesize sinograms equivalent to windowed CT scans. The synthesizer has shown good performance both visually and quantitatively in terms of PSNR, MSE and MSSIM. We notice that, while the raw sinograms are not human interpretable, distinct contrast patterns associated with the hemorrhage can be seen in the synthesized sinograms. Further, the Grad-CAM heat maps have confirmed that the proposed detection algorithm is able to focus on the injury-related areas in the sinogram automatically.

The proposed detection module contains a cascade of CNN and RNN architectures. While the CNN architecture learns slice-wise probabilities of the hemorrhage, the Bi-GRU-based RNN architecture further incorporates the relevant information from the neighborhood slices. This is the first work to use RNNs in the context of sinogram-based injury detection. The proposed CNN-RNN method has resulted in an overall patient-wise accuracy of 95.5%, with a significant improvement of over 27% compared to the best results from the existing methods. Compared to the recent vision transformer architecture, the proposed model based on CNN has demonstrated excellent performance. The attention mechanism used in vision transformers is computationally expensive, leading to a large number of parameters, indirectly resulting in the requirement of a significant amount of training data for effective learning. More importantly, vision transformers are not translation invariant by design. Thus, they may not be effective at capturing local features in the sinograms to provide the classification required. Experiments performed in this work showed that vision transformers might not be effective for the task at hand, where sinograms were utilized as input to the model.

Robustness analyses of sinogram-based detection and CT-based detection approaches are also performed. It is found that compared to CT-based hemorrhage detection, sinogram-based detection is relatively more robust to offset errors in projection angles and also to the noise during the image acquisition. Thus, the results in this paper have demonstrated that the accuracy of sinogram-based detection methods is not only on par with CT-based methods but also more robust to offset and noise errors. It can be noted that as noise increases in the images, the accuracy of the model may deteriorate. Since the current work is focused on evaluating the model’s performance on sinogram-based methods over CT-based

methods, we did not investigate computing the extreme limits of noise.

In addition to detecting the presence of a certain condition, the proposed method includes a multi-label classification analysis to classify the type of hemorrhages in a given sinogram. The results show that the proposed model accurately distinguishes various types of hemorrhages, providing valuable information for guiding treatment decisions and monitoring patient outcomes. The activation maps are used to visualize and interpret the areas in the sinogram that are most relevant to the classification decision, which improves transparency and trust in the model’s predictions. The proposed methods will hopefully take us one step closer to creating more straightforward scanner hardware to quickly detect important findings, which is the need in emergency medicine for timely intervention.

The proposed approach can potentially enhance the clinical workflow as an efficient triage tool. For example, it can be used to rapidly analyze the acquired sinograms of each patient in less than a second for the detection of ICH. The cases identified as positive for ICH can then be prioritized for CT reconstruction and for immediate further assessment by the clinical expert. It can thus enable quick diagnosis and treatment planning. Leveraging the power of deep learning, this method enables rapid detection of critical findings directly from sinograms, eliminating the need for time-consuming CT image reconstruction.

The current evaluations have certain limitations due to the unavailability of data. Firstly, because of the lack of availability of raw sinogram data, they are currently generated by applying the inverse Radon transform to CT data. In future work, we plan to perform evaluations directly on the acquired raw sinograms. Secondly, the current study did not include the detection of ICH directly from 2D projections obtained from cone-beam geometry due to the unavailability of ground truth, and we plan to study the same in future work.

V. CONCLUSION

In this work, new deep learning method is proposed for detecting and classifying ICH directly from sinograms without requiring cumbersome CT image reconstruction procedures. As sinograms are difficult to interpret by humans, the proposed method is the first of its kind to utilize the neural network-based analysis of sinograms to detect hemorrhages. It mainly consists of two modules: Intensity Transformed Sinogram Synthesizer (ITSS) and detection module. ITSS module is a novel contribution where a deep learning model is proposed to

synthesize sinograms equivalent to windowed CT scans. The proposed detection module contains a cascade of CNN and RNN architectures. While the CNN architecture learns slice-wise probabilities of the hemorrhage, the Bi-GRU-based RNN architecture further incorporates the relevant information from the neighborhood slices. The proposed method achieves an overall patient-wise accuracy of 95.5%, reflecting a substantial improvement of over 27% compared to the best results from existing methods. The method also performs well in classifying the type of hemorrhages. It is found that compared to CT-based hemorrhage detection, sinogram-based detection is relatively more robust to offset errors in projection angles and also to the noise during the image acquisition. The proposed methods will hopefully take us one step closer to creating more straightforward scanner hardware to quickly detect important findings, which is the need in emergency medicine for timely intervention.

ACKNOWLEDGEMENTS

This work is partly supported by SERB Startup research grant with Grant No: SRG/2019/00122, Semiconductor Research Corporation (SRC) with Grant No: 2022-IRP-26694272 and WIPRO-GE Collaborative Laboratory on Artificial Intelligence in Healthcare and Medical Imaging as well as Qatar National Research Fund (QNRF) with Grant No: NPRP12S-0228-190182.

CONFLICTS OF INTEREST

There are no conflicts of interest declared by Authors.

DATA AVAILABILITY

The brain CT scans with intracranial hemorrhages were obtained from the <https://www.kaggle.com/c/rsna-intracranial-hemorrhage-detection>. The code will be made available for enthusiastic users at the time of publication at <https://github.com/sindhura234/Sino-ICH>.

LIST OF FIGURES

1	A few examples of raw sinograms and corresponding CT images. The first row shows raw projections (sinograms), and the second row shows the corresponding CT images (brain window is displayed for better visualisation). The hemorrhages are contoured in red colour.	3
2	The key modules of the proposed method for automated detection of intracranial hemorrhage from the sinogram data. The details of each module are given in the sections indicated in the figure.	7
3	The proposed Intensity Transformed Sinogram synthesizer (ITSS) Module (a) Sinogram synthesis (b) The U-Net configuration.	7
4	The proposed model for the detection of ICH. The architectural details of the inception block are presented in Fig. 5.	9
5	The detailed architecture of the inception block mentioned in Fig. 4.	10
6	Sample result of ITSS module: Target Sinograms (TS), Synthesized Sinograms (SS) and the error are presented in rows one, two and three respectively.	13
7	Ablation Study: ROC curves are presented for varying numbers of channels (1C, 3C) and different Recurrent Neural Networks (Bi-LSTM, Bi-GRU).	17
8	Interpretation of the detection results on sinograms through Grad-CAM heatmaps. Each row presents results for different types of hemorrhages. Column (a) shows the raw sinograms. Column (b) shows the reconstructed CT images with the contours of hemorrhages overlaid. Column (c) illustrates the sinograms generated for the hemorrhage mask alone. Column (d) shows synthesized sinograms. Column (e) shows the Grad-CAM heat-maps.	20

LIST OF TABLES

I	Quantitative evaluation of ITSS module on slices with and without hemorrhages for three windows.	12
II	Evaluation of the proposed and state-of-the-art detection methods. Patient-wise and slice-wise results are mentioned outside and inside the brackets, respectively.	14

III	Multi-label classification results for all intracranial hemorrhage subtypes in terms of accuracy, AUC, sensitivity and specificity on the test set for the proposed model	16
IV	Robustness analysis of CT and sinogram-based detection to Poisson noise and offset errors in terms of accuracy.	17

REFERENCES

- ¹ J. Elliott and M. Smith, “The acute management of intracerebral hemorrhage: a clinical review,” *Anesthesia & Analgesia*, vol. 110, no. 5, pp. 1419–1427, 2010.
- ² F. Amyot, D. B. Arciniegas *et al.*, “A review of the effectiveness of neuroimaging modalities for the detection of traumatic brain injury,” *Journal of Neurotrauma*, vol. 32, no. 22, pp. 1693–1721, 2015.
- ³ S. Rincon, R. Gupta, and T. Ptak, “Imaging of head trauma,” *Handbook of Clinical Neurology*, vol. 135, pp. 447–477, 2016.
- ⁴ M. J. Willemink and P. B. Noël, “The evolution of image reconstruction for CT—from filtered back projection to artificial intelligence,” *European Radiology*, vol. 29, no. 5, pp. 2185–2195, 2019.
- ⁵ J. F. Barrett and N. Keat, “Artifacts in CT: Recognition and avoidance,” *Radiographics*, vol. 24, no. 6, pp. 1679–1691, 2004.
- ⁶ M. Diwakar and M. Kumar, “A review on CT image noise and its denoising,” *Biomedical Signal Processing and Control*, vol. 42, pp. 73–88, 2018.
- ⁷ S. Geyer, Lucas L *et al.*, “State of the art: Iterative CT reconstruction techniques,” *Radiology*, vol. 276, no. 2, pp. 339–357, 2015.
- ⁸ S. Zhang and Y. Xia, “CT image reconstruction algorithms: A comprehensive survey,” *Concurrency and Computation: Practice and Experience*, vol. 33, no. 8, p. e5506, 2021.
- ⁹ Y. Li, J. Wu *et al.*, “Automatic detection of the existence of subarachnoid hemorrhage from clinical CT images,” *Journal of Medical Systems*, vol. 36, no. 3, pp. 1259–1270, 2012.

- ¹⁰ Y.-H. Li, L. Zhang, Q.-M. Hu, H.-W. Li, F.-C. Jia, and J.-H. Wu, “Automatic subarachnoid space segmentation and hemorrhage detection in clinical head CT scans,” *International Journal of Computer Assisted Radiology and Surgery*, vol. 7, no. 4, pp. 507–516, 2012.
- ¹¹ L. M. Prevedello, B. S. Erdal, J. L. Ryu, K. J. Little, M. Demirer, S. Qian, and 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 S. Chilamkurthy, R. Ghosh, S. Tanamala, M. Biviji, N. G. Campeau, V. K. Venugopal, V. Mahajan, P. Rao, and P. Warier, “Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study,” *The Lancet*, vol. 392, no. 10162, pp. 2388–2396, 2018.
- ¹³ K. Jnawali, M. R. Arbabshirani, N. Rao, and A. A. Patel, “Deep 3D convolution neural network for CT brain hemorrhage classification,” in *Medical Imaging 2018: Computer-Aided Diagnosis*, vol. 10575. SPIE, 2018, pp. 307–313.
- ¹⁴ 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 Digital Medicine*, vol. 1, no. 1, pp. 1–7, 2018.
- ¹⁵ M. Vidya, Y. Mallya, A. Shastry, and J. Vijayananda, “Recurrent sub-volume analysis of head CT scans for the detection of intracranial hemorrhage,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 864–872.
- ¹⁶ A. Patel, S. C. Van De Leemput, M. Prokop, B. Van Ginneken, and R. Manniesing, “Image level training and prediction: intracranial hemorrhage identification in 3D non-contrast CT,” *IEEE Access*, vol. 7, pp. 92 355–92 364, 2019.
- ¹⁷ M. Grewal, M. M. Srivastava, P. Kumar, and S. Varadarajan, “RADnet: Radiologist level accuracy using deep learning for hemorrhage detection in CT scans,” in *IEEE International Symposium on Biomedical Imaging*, 2018, pp. 281–284.
- ¹⁸ M. Burduja, R. T. Ionescu, and N. Verga, “Accurate and efficient intracranial hemorrhage detection and subtype classification in 3D CT scans with convolutional and long short-term memory neural networks,” *Sensors*, vol. 20, no. 19, p. 5611, 2020.
- ¹⁹ A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, “An image is worth 16x16 words: Transformers for image recognition at scale,” *arXiv preprint arXiv:2010.11929*, 2020.

- ²⁰ Q. De Man, E. Haneda, B. Claus, P. Fitzgerald, B. De Man, G. Qian, H. Shan, J. Min, M. Sabuncu, and G. Wang, “A two-dimensional feasibility study of deep learning-based feature detection and characterization directly from CT sinograms,” *Medical Physics*, vol. 46, no. 12, pp. e790–e800, 2019.
- ²¹ H. Lee, C. Huang, S. Yune, S. H. Tajmir, M. Kim, and S. Do, “Machine friendly machine learning: interpretation of computed tomography without image reconstruction,” *Scientific Reports*, vol. 9, no. 1, pp. 1–9, 2019.
- ²² A. E. Flanders, L. M. Prevedello, G. Shih, S. S. Halabi, J. Kalpathy-Cramer, R. Ball, J. T. Mongan, A. Stein, F. C. Kitamura, M. P. Lungren *et al.*, “Construction of a machine learning dataset through collaboration: the rsna 2019 brain CT hemorrhage challenge,” *Radiology: Artificial Intelligence*, vol. 2, no. 3, p. e190211, 2020.
- ²³ M. Ronchetti, “Torchradon: Fast differentiable routines for computed tomography,” *arXiv preprint arXiv:2009.14788*, 2020.
- ²⁴ J. Leuschner, M. Schmidt, D. O. Baguer, and P. Maass, “Lodopab-ct, a benchmark dataset for low-dose computed tomography reconstruction,” *Scientific Data*, vol. 8, no. 1, p. 109, 2021.
- ²⁵ T. M. Buzug, “Computed tomography: from photon statistics to modern cone-beam ct,” *Soc Nuclear Med*, 2009.
- ²⁶ G. N. Hounsfield, “Computed medical imaging,” *Science*, vol. 210, no. 4465, pp. 22–28, 1980.
- ²⁷ O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2015, pp. 234–241.
- ²⁸ C. Tian, L. Fei, W. Zheng, Y. Xu, W. Zuo, and C.-W. Lin, “Deep learning on image denoising: An overview,” *Neural Networks*, vol. 131, pp. 251–275, 2020.
- ²⁹ Y. Han and J. C. Ye, “Framing U-Net via deep convolutional framelets: Application to sparse-view CT,” *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1418–1429, 2018.
- ³⁰ X. Hu, M. A. Nael, A. Wong, M. Lamm, and P. Fieguth, “RUNet: A robust UNet architecture for image super-resolution,” in *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2019, pp. 505–507.
- ³¹ D. M. Rouse and S. S. Hemami, “Understanding and simplifying the structural similarity metric,” in *IEEE International Conference on Image Processing*. IEEE, 2008, pp. 1188–1191.

- ³² C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the inception architecture for computer vision,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2818–2826.
- ³³ J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, “Empirical evaluation of gated recurrent neural networks on sequence modeling,” in *NIPS Workshop on Deep Learning*, 2014.
- ³⁴ L. Mou, P. Ghamisi, and X. X. Zhu, “Deep recurrent neural networks for hyperspectral image classification,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 7, pp. 3639–3655, 2017.
- ³⁵ M. Schuster and K. Paliwal, “Bidirectional recurrent neural networks,” *IEEE Transactions on Signal Processing*, vol. 45, no. 11, pp. 2673–2681, 1997.
- ³⁶ A. Graves and J. Schmidhuber, “Framewise phoneme classification with bidirectional lstm networks,” in *Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005.*, vol. 4, 2005, pp. 2047–2052 vol. 4.
- ³⁷ S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, “Aggregated residual transformations for deep neural networks,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1492–1500.
- ³⁸ M. Diwakar and M. Kumar, “A review on CT image noise and its denoising,” *Biomedical Signal Processing and Control*, vol. 42, pp. 73–88, 2018.
- ³⁹ A. Biguri, M. Dosanjh, S. Hancock, and M. Soleimani, “Tigre: a matlab-gpu toolbox for cbct image reconstruction,” *Biomedical Physics & Engineering Express*, vol. 2, 2016.
- ⁴⁰ M. Kachelriess, M. Knaup, and O. Bockenbach, “Hyperfast parallel-beam and cone-beam back-projection using the cell general purpose hardware,” *Medical Physics*, vol. 34, no. 4, pp. 1474–1486, 2007.
- ⁴¹ G. H. Chen, S. Leng, and C. A. Mistretta, “A novel extension of the parallel-beam projection-slice theorem to divergent fan-beam and cone-beam projections,” *Medical physics*, vol. 32, no. 3, pp. 654–665, 2005.
- ⁴² R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, “Grad-CAM: Visual explanations from deep networks via gradient-based localization,” in *IEEE International Conference on Computer Vision*, 2017, pp. 618–626.