# Deep Residual Dilated Convolutional Learning for Detection of Large Vessel Occlusion in Ischemic Stroke Patients


Rohan Chatterjee
Department of Computer Science
California State University, Los Angeles
Los Angeles, CA, USA
rchatte@calstatela.edu

Dikshant Sagar
Department of Computer Science
California State University, Los Angeles
Los Angeles, CA, USA
dsagar2@calstatela.edu

Farnaz Mohammadi
Genentech Computational Sciences
Genentech, Inc.
South San Fransico, CA, USA
farnazmd95@gmail.com

Arpit Vaishya
Department of Computer Science
California State University, Los Angeles
Los Angeles, CA, USA
avaishy@calstatela.edu

Mohammad Pourhomayoun
Department of Computer Science
California State University, Los Angeles
Los Angeles, CA, USA
mpourho@calstatela.edu

Manveen Kaur
Department of Computer Science
California State University, Los Angeles
Los Angeles, CA, USA
jlim34@calstatela.edu

Jung Soo Lim
Department of Computer Science
California State University, Los Angeles
Los Angeles, CA, USA
jlim34@calstatela.edu

Navid Amini
Department of Computer Science
California State University, Los Angeles
Los Angeles, CA, USA
namini@calstatela.edu



***Abstract*— Rapid identification of large vessel occlusions (LVOs) is crucial when treating and managing patients with acute ischemic strokes (AIS). This urgency is due to the fact that LVOs are associated with high rates of post-stroke dependence and mortality. Furthermore, precise and prompt detection of such occlusions can assist in minimizing long-term functional and psychological ramifications. The conventional method for LVO detection requires manual examination of Computed Tomography Angiography (CTA) brain images. However, treatment delays frequently occur due to the unavailability of skilled radiologists specialized for these cases. In this study, we introduce a Three-Dimensional Deep Residual Dilated Convolutional Neural Network (DRDCNet-3D) that can harness the capabilities of dilated convolutions to improve spatial resolution. This method allows for rapid and precise detection of LVOs directly from three-dimensional CTA brain scans. Utilizing the recently published IACTA-EST dataset, our designed model has exhibited remarkable performance, boasting an AUC-ROC of 0.91 and an F1-score of 0.90. Such performance represents an average improvement of over 20% compared to existing state-of-the-art 3D medical image analysis models available in open-source platforms. This notable enhancement underscores the model's efficacy in diagnosing large vessel occlusions (LVOs). Our approach equips physicians with valuable insights for determining optimal treatment strategies in suspected AIS cases. Accordingly, timely interventions can be carried out, reducing the risk of permanent severe damage and benefiting patient outcomes.**



***Keywords*— *Large Vessel Occlusion, Convolutional Neural Network, Computer Tomography Angiography, Dilated Convolution, Residual Connections.***


## I. Introduction

Acute ischemic strokes (AIS), often caused by large vessel occlusion (LVO), comprise about 85% of all strokes and are currently the fifth leading cause of mortality in the US [1], [2]. Large vessel occlusion occurs when the proximal posterior, middle, and anterior cerebral arteries are obstructed or blocked [3]. Furthermore, LVOs result in three out of five cases of post-stroke dependence or disability and are responsible for more than 90% of post-stroke mortality [4]. Notably, advanced and modern treatments such as endovascular thrombectomy (EVT) have demonstrated substantial superiority in the treatment of patients with AIS [5], [6]. Nevertheless, rapid diagnosis and treatment are crucial for a favorable outcome, given that research indicates the efficacy of EVT may start to decline within merely 7.3 hours from the onset of stroke symptoms [5]–[8]. Moreover, facilities capable of performing EVT [9] are scarce, and patients may need to be transferred to more advanced healthcare centers for proper treatment once occlusions are identified. Hence, the timely identification of LVOs in AIS patients is both a vital and an urgent requirement within the healthcare system.

Computed tomography angiography (CTA) is a standard imaging technique used for LVO detection [6], [10]. Due to the detailed information they provide in visualizing the intracranial vascular structure, CTAs have demonstrated higher diagnostic accuracy compared to the use of other imaging methods [11]. However, manual detection and analysis consume a significant amount of time and demand the expertise of experienced radiologists, who are limited in number and availability. Faced with these obstacles, ongoing research is exploring alternative methods to accelerate and improve the accuracy of LVO detection [12]. Recently, through the integration of deep learning techniques and clinical data, researchers have made a notable advancement in the early identification of LVO cases, demonstrating the potential to streamline the diagnostic process and improve patient care [13].

Brain CTA data is inherently three-dimensional, providing a comprehensive view of the intricate vascular structures of the brain. Consequently, the task of LVO detection demands solutions that have the ability to extract and capture this spatial information to yield diagnostic insights. Lately, Convolutional Neural Networks (CNNs) have emerged as powerful tools in the realm of 3D image processing and analysis. These sophisticated algorithms, inspired by the human visual system, excel in recognizing intricate patterns and features within images. In certain instances, their ability to dissect complex visual data has revolutionized tasks that were once challenging for traditional algorithms or human experts alone. Utilizing their expertise in extracting complex hierarchical features, these networks can be employed to interpret a wide spectrum of information [14]. Their adaptability and capacity to learn from vast datasets have significantly enhanced the accuracy and speed of image-related tasks, leading to groundbreaking advancements in various domains, including healthcare, diagnostics, and treatment [15].

The convergence of medical expertise and artificial intelligence marks a transformative era in healthcare, where the synthesis of advanced technologies and human knowledge is leading to breakthroughs in patient care and medical understanding [16]. These state-of-the-art techniques not only meet but exceed the challenges posed by the diagnosis of conditions such as large vessel occlusions. By closely examining the intricate details of CTA scans, CNNs can identify subtle patterns, irregularities, and occlusions directly from the image. This innovative approach represents a significant leap forward in medical diagnostics. It ensures not only timely but also highly accurate detection of LVOs, enabling prompt intervention and significantly improving the outcomes for patients experiencing acute ischemic strokes.

Aligned with these significant technological strides, in this paper, we introduce a novel three-dimensional Deep Residual Dilated Convolutional Neural Network (DRDCNet-3D) architecture designed to detect LVOs on CTA brain scans. The proposed model, along with its intricate underlying architecture, significantly amplifies the efficiency of the detection process, leading to a substantial acceleration in the diagnostic procedure.

Our approach utilizes the power of dilated convolutions to expand the receptive field of CNNs, resulting in enhanced computational efficiency and achieving remarkable detection accuracy. Moreover, the integration of residual connections significantly amplifies information propagation within the architecture, facilitating data transfer between layers to preserve lower-level information, culminating in enhanced LVO prediction capabilities.

To the best of our knowledge, this is the first study that exploits the potential of 3D dilated convolutions to enhance spatial resolution when investigating multi-dimensional medical neuroimaging data for LVO detection. Using the publicly available IACTA-EST dataset [17], we showcase our model's potential for precise LVO detection with an improvement of over 20% when compared to existing state-of-the-art 3D models. Furthermore, in addition to its superior performance, our model demonstrates its efficiency with average inference times less than one-third of those required by existing models, significantly enhancing its practical utility.

The remainder of this paper is organized as follows: Section II provides an overview of the relevant literature. Section III describes the IACTA-EST dataset, outlines the preprocessing steps we employ, provides an explanation of our DRDCNet-3D architecture, and discusses our training process. Section IV covers the experimental results and their corresponding discussions, and finally, Section V concludes this study.

## II. Related Work

Several studies in the field of medical imaging and deep learning demonstrate the utility of cutting-edge computational methods in addressing critical neurological medical conditions. These studies are an important foundation for our work, which shares the goal of leveraging advanced deep learning methodology to improve diagnosis and treatment. One such recent noteworthy paper by Zhao et al. [18] presents a specific deep learning model trained on 1177 multi-dimensional CTA scans for detecting intracranial aneurysms. The model can predict 99% of the negative cases accurately with high confidence and demonstrate a potential reduction in human workload.

Furthermore, several studies harness the potential of dilated convolutions for the analysis of multi-dimensional neurological data. An approach by Nazir et al. [19] proposes an optimized end-to-end architecture based on convolutional neural networks utilizing Magnetic Resonance Imaging (MRI) and CTA brain scans. Similar to our approach, their model employs varying dilated convolutions with residual connections to enhance information flow between layers. However, their work is not focused on detection of neurological diseases, but rather it primarily focuses on the segmentation of intracranial vascular structures within neuroimaging data.

Moreover, multiple studies address the challenge of LVO detection. One such approach is MBH-Net by Yao et al. [20], which achieves satisfactory accuracy using a three-branch hybrid network and an auxiliary attention guidance module. However, their approach exclusively processes 2D images, neglecting the value of 3D spatial data. Another approach by Barman and colleagues [21] proposes a deep learning architecture named DeepSymNet. Inspired by Siamese Networks, their proposed model utilizes the two brain hemispheres as parallel input to detect changes in the symmetry of vascular and brain tissue, aiding in the detection of LVO. However, the model primarily focuses on detecting differences in symmetrical differences, potentially limiting its ability to capture other essential aspects of LVO.

In the field of computer vision, conventional images are typically viewed as 2D data. However, researchers have increasingly delved into the realm of 3D spatial data, such as videos or three-dimensional objects. Adapting to this shift, scientists have sought to extend existing 2D algorithms to process richer 3D data. To accomplish so, they reconstruct 3D versions of familiar models by incorporating specialized 3D convolutional layers. This innovative approach enables the adaptation of established algorithms to the complexities of 3D data, fostering advancements in various applications within the realm of computer vision. Two of the well-known models are 3D Resnet50 [22], [23] and 3D EfficientNetB0 [24].

## III. Methods

Our process for detecting LVOs begins with the collection of relevant CTA brain scans, followed by the application of essential preprocessing steps to the acquired samples, such as normalization and rescaling. At the core of our methodology is the use of our proposed DRDCNet-3D, meticulously engineered to achieve precise and dependable LVO detection. Figure 1 depicts a high-level overview of our complete LVO detection workflow. Our comprehensive approach is designed to enhance the efficiency and effectiveness of LVO diagnosis.

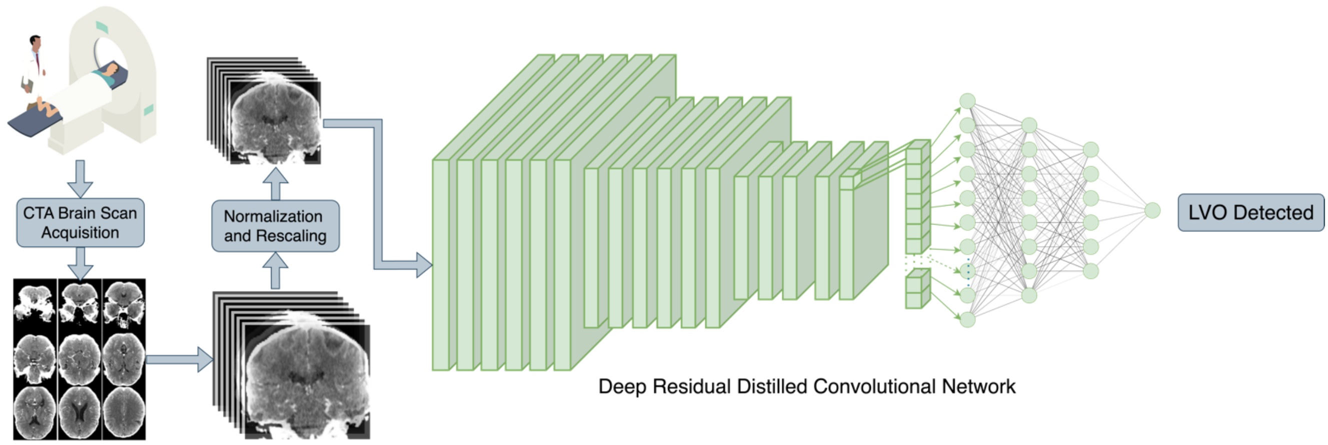


Fig. 1. The workflow of our model for detecting large vessel occlusions. Collected brain computed tomography angiography scans undergo essential preprocessing steps such as skull stripping, normalization, and rescaling. The proposed DRDCNet-3D model is subsequently trained using the prepared dataset, leveraging its robust architecture and carefully tuned hyperparameters to achieve accurate and reliable LVO detection.

### A. *Dataset*

To evaluate our model, we utilize the publicly available IACTA-EST dataset [17] from the University of Texas Health Science Center at Houston (UTHealth), made accessible by the IEEE Symposium on Biomedical Imaging (ISBI 2023). The dataset is comprised of 301 CTA brain scans with 159 non-LVO and 142 LVO samples. While this sampling may not reflect the real-world distribution of such cases, utilizing balanced data is advantageous and necessary for training deep learning models without introducing bias.

All CTA samples in the IACTA-EST dataset are sourced from stroke centers located in Houston, Texas. Before public distribution, the samples in the dataset are skull-stripped, resampled, and linearly registered to a standard image template using rigid transformations. Moreover, each scan is manually labeled to determine the presence of LVO, according to the neuroradiologist's reports from the stroke center. Finally, the scans are converted from DICOM to NIfTI format for easy access and use. Figure 2 visualizes a descending vertical slice from a sample 3D CTA brain scan at varying depths.

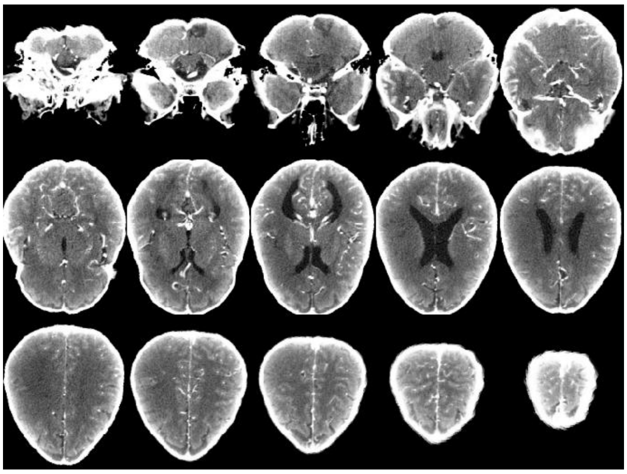

Fig. 2. Example of descending vertical slices of a brain computed tomography angiography (CTA).

The initial resolution of all shared CTA scans is $146 \times 182 \times 133$. Voxel values are clipped between the range of 0 to 100 Hounsfield units, and the voxel size is $1 \times 1 \times 1$ mm. The dataset curators define an LVO as an occlusion in the internal carotid artery (ICA), M1, M2, or A1 brain vasculature. High-grade stenosis or near-complete occlusions are not considered LVOs for the purposes of this dataset. Posterior circulation stroke and LVOs in other segments are also excluded in this study.

### B. *Preprocessing*

The preprocessing of the images involve multiple essential steps to standardize the training process and increase overall efficiency. Initially, border cropping is applied to remove extraneous pixels that are irrelevant to the image content and task at hand. Subsequently, spline interpolation is used to decrease the size of the samples by 10 percent in each dimension, improving the efficiency of the training process. This method is utilized as research indicates that spline interpolation effectively minimizes information loss while incurring only minimal computational costs [25]. As a result, the final spatial size for each CTA scan is significantly reduced to $133 \times 163 \times 119$ while losing minimal information and effectively reducing the computational cost of the model. Lastly, the dataset undergoes normalization to be confined between 0 and 1. For initial training and testing purposes, the dataset is randomly split into training and validation sets with a ratio of 75:25, respectively, which is used to compare our proposed model to other well-known 3D CNN architectures.

### C. *Model Architecture Design*

Our proposed DRDCNet-3D model consists of a total of thirty-three layers, divided into five convolutional blocks, followed by three pairs of fully connected layers, both utilizing the rectified linear unit (ReLU) activation function. The model concludes with a final dense layer with a sigmoid activation for a binary classification output to detect LVO. A detailed architectural design of our model can be seen in Figure 3. Each convolutional block contains a batch normalization layer along with multiple pairs of 3D dilated convolutional layers using the ReLU activation and a 3D max pooling layer. All 3D max pooling layers are implemented with the same pooling resolution of $2 \times 2 \times 2$.

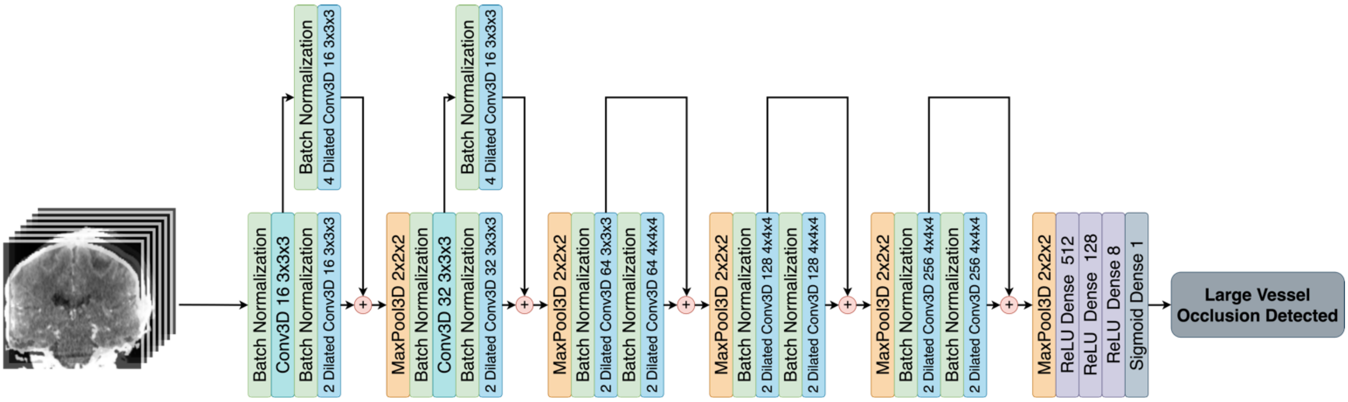


Fig. 3. Model architecture for the proposed DRDCNet-3D. Dilated convolution layers are utilized to obtain a high spatial resolution and residual connections enhance information flow. Padding is used to maintain consistent input-output shapes, and max pooling layers are used to reduce dimensionality and capture essential features.

The first two blocks are designed to extract low-level volume features using basic $3 \times 3 \times 3$ convolutional layers and two varying dilation factors focusing on different spatial resolutions. Dilated convolutional layers allow for features to be extracted with a larger receptive field, resulting in the integration of more extensive contextual information. This results in an enhancement of the model's ability to identify complex patterns and relationships within the data without dramatically expanding computational demands [26]. Furthermore, distinct dilation factors result in distinct receptive fields for the convolutional kernel extracting multi-scale information.

After the initial convolutional block, the number of kernels is doubled for each subsequent block to facilitate the extraction of higher-level features. Residual connections are integrated to provide a new path between layers to improve information flow and ensure low-level features are available to the subsequent layers. Deep neural networks often suffer from vanishing gradients during training, hindering the convergence process. Residual connections can allow gradients to flow more easily through the network by providing shortcuts [27]. This helps mitigate the vanishing gradient problem and facilitates smoother and faster convergence.

The padding of each convolutional layer is dynamically adjusted based on the size of the kernel and dilation factors, ensuring that the layer's input and output maintain the same spatial size. Maintaining a constant spatial size in CNNs has been shown to yield several benefits that contribute to the effectiveness of the network architecture and the quality of learned features, as it preserves spatial information irrespective of the network depth and allows for an efficient architecture design. Batch normalization is used in each block, as shown in Figure 3, to mitigate internal covariate shifts due to variance in data distribution between batches, leading to faster convergence and improved model accuracy. Furthermore, batch normalization acts as a regularizing agent, which prevents overfitting during training [28].

### D. Training

The proposed model architecture is implemented using PyTorch 2.0.1 and trained using an Nvidia RTX 4070 GPU. For training and evaluation, we employ two separate strategies. The first strategy involves randomly dividing the data into dedicated train and test subsets, with 75% of the data allocated for training. In addition, we employ a cross-validation approach to ensure a comprehensive model assessment. Binary Cross Entropy is used as the loss function with a learning rate of 0.0001 across 150 epochs. We utilize the Adam Optimizer [29], which stands for Adaptive Moment Estimation, a widely used optimization algorithm based on stochastic gradient descent.

Batch size is an essential hyperparameter in deep learning models, and different batch sizes can often result in different accuracies and run times [30]. Due to the memory constraints of a single GPU and the relatively large input size, gradient accumulation is performed after every four batches of size 4. Gradient accumulation is a common splitting mechanism used to overcome memory limitations when training large models or processing large batches of data. This mechanism divides the training data into smaller batches and accumulates the gradients before applying them to the model [31]. By performing gradient accumulation, the effective batch size is increased to 16, addressing the GPU memory limitation. These optimal parameter values, including batch size, are achieved after multiple rounds of testing to maximize both accuracy and efficiency.

## IV. Results and Discussion

Our evaluation process consists of two major components. Firstly, we conduct extensive testing by utilizing dedicated training and validation sets crafted from the IACTA-EST dataset. This approach allows us to compare the results achieved by our DRDCNet-3D model with those of other well-established architectures. Subsequently, to ensure the robustness and generalization of our model, we employ a four-fold cross-validation strategy, providing a more comprehensive assessment of our model's performance.

Following the rigorous training process, a thorough assessment is conducted using the designated validation set consisting of 25% of the original IACTA-EST dataset. The results are shown in Table I. Although the train-test split is class-stratified, given the inherent imbalance of real-world LVO distribution, it is important to refrain from using accuracy as a sole evaluation metric. Instead, we employ five metrics to evaluate the models, including the Area Under the Receiver Operating Characteristic Curve (AUC-ROC), F1-Score, Precision, Sensitivity, and Specificity. These metrics allow us to measure and prevent the models' bias towards the majority class. This is especially important when working with medical datasets, as they are more susceptible to distribution bias. The comparative analysis is centered around three models, 3D ResNet50, 3D EfficientNetB0, and our proposed DRDCNet-3D, all evaluated on the validation set. The architecture of both comparative models is similar to their general 2D variants [24] [22] [23]; however, they are comprised of 3D modules using TensorFlow and are modified for binary detection. Comparison to the previously published approaches, such as MBH-Net [20] and DeepSymNet [21] cannot be performed due to the unavailability of code.

TABLE I. COMPARISON OF MODEL PERFORMANCE METRICS

| *Model Architecture* | *AUC-ROC* | *F1-Score* | *Precision* | *Sensitivity* | *Specificity* |
|---|---|---|---|---|---|
| 3D ResNet50 | 0.68 | 0.65 | 0.76 | 0.56 | 0.81 |
| 3D EfficientNet-B0 | 0.77 | 0.72 | 0.68 | 0.77 | 0.62 |
| **DRDCNet-3D (Our Model)** | **0.91** | **0.90** | **1.00** | **0.82** | **1.00** |

Remarkably, our proposed DRDCNet-3D model outperforms the other models across all metrics, displaying an AUC-ROC of 0.91, an F1-score of 0.90, a precision of 1.0, a sensitivity of 0.82, and a specificity of 1.0 on the dedicated validation set. In contrast, the 3D ResNet50 model yields less favorable performance, characterized by an AUC-ROC of 0.68, an F1-score of 0.65, a precision of 0.76, a sensitivity of 0.56, and a specificity of 0.81. Furthermore, the 3D EfficientNet-B0 model demonstrates slightly superior performance in some metrics, registering an AUC-ROC of 0.77, an F1-score of 0.72, and a sensitivity of 0.77. However, it exhibits reduced performance in terms of precision and specificity, with scores of 0.68 and 0.62, respectively. In Figure 4, we plot the receiver operating characteristic curves of the three models for comparison, further indicating the superior performance of our proposed model.

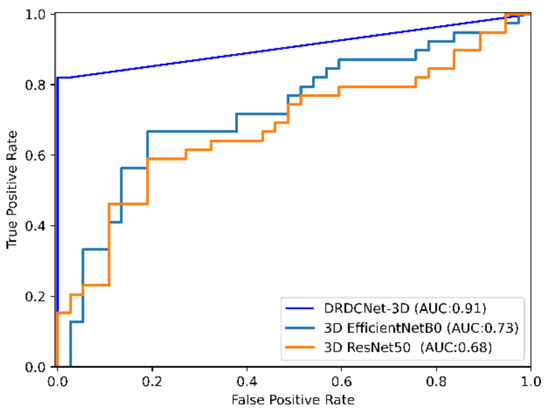


Fig. 4. The receiver operating characteristic (ROC) curve for LVO detection highlights the superior performance of our DRDCNet-3D model, outperforming the 3D Resnet50 and 3D EfficientNet-B0 models on the validation set.

Figure 5 visualizes the confusion matrix for predictions generated by our proposed DRDCNet-3D model on the validation set. Particularly noteworthy is the absence of false positives, indicating that our model is not likely to predict the presence of LVO in non-LVO patients erroneously. Equally remarkable is the high count of true negatives, as all 37 of the negative samples are accurately identified by the model, illustrating the model's ability to classify patients without LVO as accurately. These attributes hold significant value within the medical realm, given the escalating costs associated with unnecessary medical interventions. Furthermore, the confusion matrix shows that our model has a remarkably low number of false negatives and notably high true positives. This outcome highlights the model's effectiveness in optimally detecting LVO cases and facilitating timely treatment. Consequently, the model holds the potential to substantially optimize stroke center resources, improving patients' survival rate and quality of life.

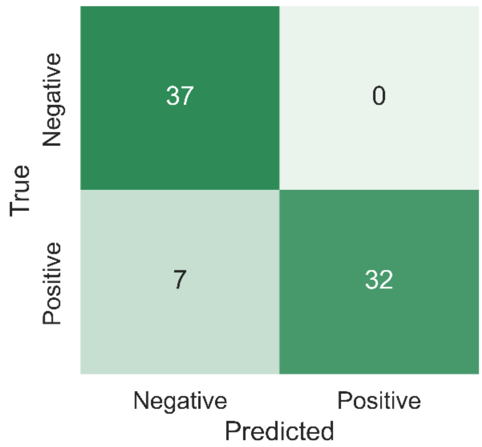


Fig. 5. The confusion matrix for our DRDCNet-3D model's performance in the detection of LVOs.

A critical point of interest in our comparison is inference time, which refers to the time required for the architecture to process a data sample and make a prediction. Inference time stands as an essential performance metric that compares the efficiency and computational cost associated with the model. Minimizing inference time is essential for reducing latency and optimizing resource utilization, directly impacting real-world usability for the architecture. The average inference time for the validation set has been summarized in Table II. Once again, our model outperforms both 3D ResNet50 and 3D EfficientNet-B0 architectures with an inference time of less than 36 milliseconds. This significantly lower inference time of our model demonstrates its resource efficiency while outperforming the other models in terms of LVO detection.

TABLE II. COMPARISON OF MODEL INFERENCE TIME

| *Model Architecture* | *Inference Time (ms)* |
|---|---|
| 3D ResNet50 | 121.314 |
| 3D EfficientNet-B0 | 439.875 |
| **DRDCNet-3D (Our Model)** | **35.827** |

Moreover, in order to meticulously avoid any potential data split bias and substantiate the integrity of our findings, we conduct four-fold cross-validation to assess our model's consistent performance and the data split's validity thoroughly. By dividing the dataset into distinct subsets/folds and systematically rotating the training and validation data, we provide a more comprehensive understanding of our model's ability to generalize unseen data and ensure a balanced evaluation of our architecture, ensuring the model's effectiveness across different data subsets. This process not only ensures the model's reliability but also strengthens the credibility of our research findings.

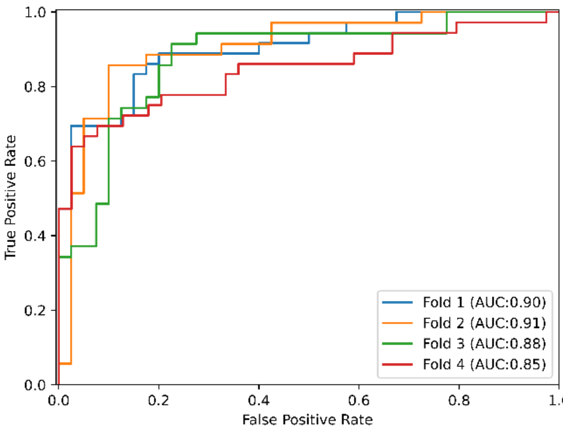


Fig. 6. The receiver operating characteristic (ROC) curve for LVO detection for each iteration within the four-fold cross-validation process. The results showcase the model's robustness and generalization providing a more comprehensive evaluation of its performance.

We deliberately select a four-fold approach for this study, as it allows for a substantial representation of data in both training and validation sets for each iteration. Despite the modest size of the IACTA-EST dataset, this approach allows for both reliable and stable results during validation. The results from these iterations are then averaged to provide a more accurate and reliable estimation of the model's performance. The outcomes of this experiment are comprehensively presented in Table III, while Figure 6 provides a graphical representation of the AUC-ROC results for each iteration within the four-fold cross-validation process. We observe that variations of our model trained in these experiments perform consistently well, achieving accuracy metrics higher than existing model architectures.

TABLE III. DRDCNET-3D CROSS-VALIDATION PERFORMANCE

| *Fold* | *AUC-ROC* | *F1-Score* | *Precision* | *Sensitivity* | *Specificity* |
|---|---|---|---|---|---|
| 1 | 0.90 | 0.83 | 0.83 | 0.83 | 0.85 |
| 2 | 0.91 | 0.86 | 0.86 | 0.86 | 0.88 |
| 3 | 0.88 | 0.83 | 0.78 | 0.89 | 0.78 |
| 4 | 0.85 | 0.75 | 0.92 | 0.64 | 0.95 |
| **Mean ± SD** | **0.88 ± 0.03** | **0.82 ± 0.04** | **0.85 ± 0.06** | **0.80 ± 0.12** | **0.86 ± 0.07** |

Every innovative approach inevitably comes with its own set of limitations. One notable constraint we encounter stems from computational and memory limitations, which necessitated compressing the 3D images within the dataset dimensionally by approximately 10%. While we use caution to minimize the potential loss of valuable information, it is important to acknowledge that this reduction in image resolution could potentially impact the overall performance of our approach. Furthermore, another limitation arises from the relatively limited number of samples available within the dataset. This constraint could potentially influence the model's learned distribution, hindering its ability to capture the intricate variations present in real-world cases effectively. These variations can emerge due to various factors, including variances in CTA scan acquisition methodologies and differences in equipment used to acquire CTA images. The scarcity of samples might limit the model's adaptability to these diverse real-world conditions, potentially constraining its generalizability.

Our overarching goal is to significantly reduce the mortality rate of AIS patients and alleviate the burden of long-term disabilities they may face. We recognize that as we move forward, refining our methodology and addressing these limitations for future work is essential to ensure that our model is not only reliable but also highly accurate while enhancing generalizability for rapid and accurate LVO detection.

## V. CONCLUSIONS

The occurrence of strokes is on the rise, reaching a point where someone in the United States experiences a stroke approximately every 40 seconds. Ischemic strokes resulting from LVOs account for the majority of these cases. Accordingly, rapid detection of LVOs is a pressing need in ischemic stroke care, playing a significant role in decreasing the mortality rate and improving the quality of life for affected patients. In this paper, we demonstrate the remarkable potential of deep neural networks in accurately predicting LVOs through the analysis of CTA brain scans by proposing a residual dilated CNN-based model called DRDCNet-3D. We also elucidate the efficacy of advanced techniques such as batch normalization, residual connections, and dilated convolutions, which allow for efficient training of neural networks. The proposed model exhibits superior performance to that of existing models, underlining its potential as a valuable tool in enhancing diagnostic accuracy and stroke patient care. In essence, the outcomes of our study not only accentuate the potential of deep convolutional neural networks in addressing LVO prediction but also emphasize the significance of employing a diversified set of evaluation metrics, particularly within medical contexts where accurate predictions are pivotal. As we continue to build on this research, our ongoing objective remains to significantly reduce the mortality rate of AIS patients and alleviate the burden of long-term disabilities they may face, striving to meet our ultimate goal of advancing the field of stroke care and improving the lives of those affected by this critical medical condition.

## REFERENCES

[1] W. S. Smith, "Endovascular Stroke Therapy," *Neurotherapeutics*, vol. 16, no. 2, pp. 360–368, Apr. 2019, doi: 10.1007/s13311-019-00724-5.

[2] F. Herpich and F. Rincon, "Management of Acute Ischemic Stroke," *Critical Care Medicine*, vol. 48, no. 11, pp. 1654–1663, Nov. 2020, doi: 10.1097/CCM.0000000000004597.

[3] R. C. Rennert *et al.*, "Epidemiology, Natural History, and Clinical Presentation of Large Vessel Ischemic Stroke," *Neurosurg.*, vol. 85, no. suppl_1, pp. S4–S8, Jul. 2019, doi: 10.1093/neuros/nyz042.

[4] K. Malhotra, J. Gornbein, and J. L. Saver, "Ischemic Strokes Due to Large-Vessel Occlusions Contribute Disproportionately to Stroke-Related Dependence and Death: A Review," *Front. Neurol.*, vol. 8, p. 651, Nov. 2017, doi: 10.3389/fneur.2017.00651.

[5] M. Goyal *et al.*, "Endovascular thrombectomy after large-vessel ischaemic stroke: a meta-analysis of individual patient data from five randomised trials," *The Lancet*, vol. 387, no. 10029, pp. 1723–1731, Apr. 2016, doi: 10.1016/S0140-6736(16)00163-X.

[6] M. A. Almekhlafi *et al.*, "Imaging of Patients with Suspected Large-Vessel Occlusion at Primary Stroke Centers: Available Modalities and a Suggested Approach," *AJNR Am J Neuroradiol*, p. ajnr;ajnr.A5971v1, Jan. 2019, doi: 10.3174/ajnr.A5971.

[7] T. G. Jovin *et al.*, "Thrombectomy within 8 Hours after Symptom Onset in Ischemic Stroke," *N Engl J Med*, vol. 372, no. 24, pp. 2296–2306, Jun. 2015, doi: 10.1056/NEJMoa1503780.

[8] J. L. Saver *et al.*, "Time to Treatment With Endovascular Thrombectomy and Outcomes From Ischemic Stroke: A Meta-

analysis," *JAMA*, vol. 316, no. 12, p. 1279, Sep. 2016, doi: 10.1001/jama.2016.13647.

[9] D. Aguiar De Sousa *et al.*, "Access to and delivery of acute ischaemic stroke treatments: A survey of national scientific societies and stroke experts in 44 European countries," *European Stroke Journal*, vol. 4, no. 1, pp. 13–28, Mar. 2019, doi: 10.1177/2396987318786023.

[10] S. A. Mayer *et al.*, "CTA-for-All: Impact of Emergency Computed Tomographic Angiography for All Patients With Stroke Presenting Within 24 Hours of Onset," *Stroke*, vol. 51, no. 1, pp. 331–334, Jan. 2020, doi: 10.1161/STROKEAHA.119.027356.

[11] M. J. Becks *et al.*, "Brain CT perfusion improves intracranial vessel occlusion detection on CT angiography," *Journal of Neuroradiology*, vol. 46, no. 2, pp. 124–129, Mar. 2019, doi: 10.1016/j.neurad.2018.03.003.

[12] S. A. Amukotuwa, M. Straka, S. Dehkharghani, and R. Bammer, "Fast Automatic Detection of Large Vessel Occlusions on CT Angiography," *Stroke*, vol. 50, no. 12, pp. 3431–3438, Dec. 2019, doi: 10.1161/STROKEAHA.119.027076.

[13] G. Tarkanyi, A. Tenyi, R. Hollos, P. J. Kalmar, and L. Szapary, "Optimization of Large Vessel Occlusion Detection in Acute Ischemic Stroke Using Machine Learning Methods," *Life*, vol. 12, no. 2, p. 230, Feb. 2022, doi: 10.3390/life12020230.

[14] D. R. Sarvamangala and R. V. Kulkarni, "Convolutional neural networks in medical image understanding: a survey," *Evol. Intel.*, vol. 15, no. 1, pp. 1–22, Mar. 2022, doi: 10.1007/s12065-020-00540-3.

[15] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects," *IEEE Trans. Neural Netw. Learning Syst.*, vol. 33, no. 12, pp. 6999–7019, Dec. 2022, doi: 10.1109/TNNLS.2021.3084827.

[16] A. Barragán-Montero *et al.*, "Artificial intelligence and machine learning for medical imaging: A technology review," *Physica Medica*, vol. 83, pp. 242–256, Mar. 2021, doi: 10.1016/j.ejmp.2021.04.016.

[17] 2023 IEEE International Symposium on Biomedical Imaging (ISBI), "The Image Analysis for CTA Endovascular Stroke Therapy Dataset (IACTA-EST)." 2023.

[18] Z. Shi *et al.*, "A clinically applicable deep-learning model for detecting intracranial aneurysm in computed tomography angiography images," *Nat Commun*, vol. 11, no. 1, p. 6090, Nov. 2020, doi: 10.1038/s41467-020-19527-w.

[19] A. Nazir *et al.*, "OFF-eNET: An Optimally Fused Fully End-to-End Network for Automatic Dense Volumetric 3D Intracranial Blood Vessels Segmentation," *IEEE Trans. on Image Process.*, vol. 29, pp. 7192–7202, 2020, doi: 10.1109/TIP.2020.2999854.

[20] R. Yao *et al.*, "MBH-Net: Multi-branch Hybrid Network with Auxiliary Attention Guidance for Large Vessel Occlusion Detection," in *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Las Vegas, NV, USA: IEEE, Dec. 2022, pp. 872–876. doi: 10.1109/BIBM55620.2022.9995268.

[21] A. Barman, M. E. Inam, S. Lee, S. Savitz, S. Sheth, and L. Giancardo, "Determining Ischemic Stroke From CT-Angiography Imaging Using Symmetry-Sensitive Convolutional Networks," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, Venice, Italy: IEEE, Apr. 2019, pp. 1873–1877. doi: 10.1109/ISBI.2019.8759475.

[22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA: IEEE, Jun. 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.

[23] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," 2019, doi: 10.48550/ARXIV.1905.11946.

[24] A. Ebrahimi, S. Luo, and R. Chiong, "Introducing Transfer Learning to 3D ResNet-18 for Alzheimer's Disease Detection on MRI Images," in *2020 35th International Conference on Image and Vision Computing New Zealand (IVCNZ)*, Wellington, New Zealand: IEEE, Nov. 2020, pp. 1–6. doi: 10.1109/IVCNZ51579.2020.9290616.

[25] M. Gabbouj, P. Kuosmanen, Tampereen teknillinen korkeakoulu, and European Association for Signal Processing, Eds., *Signal processing X: theories and applications ; proceedings of EUSIPCO 2000, Tenth European Signal Processing Conference, Tampere, Finland, 4 - 8 September 2000*. Tampere, 2000.

[26] W. Li, G. Wang, L. Fidon, S. Ourselin, M. J. Cardoso, and T. Vercauteren, "On the Compactness, Efficiency, and Representation of 3D Convolutional Networks: Brain Parcellation as a Pretext Task," in *Information Processing in Medical Imaging*, vol. 10265, M. Niethammer, M. Styner, S. Aylward, H. Zhu, I. Oguz, P.-T. Yap, and D. Shen, Eds., in Lecture Notes in Computer Science, vol. 10265. , Cham: Springer International Publishing, 2017, pp. 348–360. doi: 10.1007/978-3-319-59050-9_28.

[27] K. He, X. Zhang, S. Ren, and J. Sun, "Identity Mappings in Deep Residual Networks," in *Computer Vision – ECCV 2016*, vol. 9908, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds., in Lecture Notes in Computer Science, vol. 9908. , Cham: Springer International Publishing, 2016, pp. 630–645. doi: 10.1007/978-3-319-46493-0_38.

[28] N. Pattanajak and H. Malekmohamadi, "Improving a 3-D Convolutional Neural Network Model Reinvented from VGG16 with Batch Normalization," in *2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, Leicester, United Kingdom: IEEE, Aug. 2019, pp. 45–50. doi: 10.1109/SmartWorld-UIC-ATC-SCALCOM-IOP-SCI.2019.00052.

[29] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," 2014, doi: 10.48550/ARXIV.1412.6980.

[30] R. Lin, "Analysis on the Selection of the Appropriate Batch Size in CNN Neural Network," in *2022 International Conference on Machine Learning and Knowledge Engineering (MLKE)*, Guilin, China: IEEE, Feb. 2022, pp. 106–109. doi: 10.1109/MLKE55170.2022.00026.

[31] Z. Huang, B. Jiang, T. Guo, and Y. Liu, "Measuring the Impact of Gradient Accumulation on Cloud-based Distributed Training," in *2023 IEEE/ACM 23rd International Symposium on Cluster, Cloud and Internet Computing (CCGrid)*, Bangalore, India: IEEE, May 2023, pp. 344–354. doi: 10.1109/CCGrid57682.2023.00040.