Published in final edited form as:
Proc SPIE Int Soc Opt Eng. 2022 ; 12032: . doi:10.1117/12.2613147.

Calcium scoring in low-dose ungated chest CT scans using convolutional long-short term memory networks

K. Pieszko[1,*], **A. Shanbhag**[1,*], **A. Killekar**[1], **M. Lemley**[1], **Y. Otaki**[1], **Serge Van Kriekinge**[1], **Paul Kavanagh**[1], **Robert JH Miller**[1,3], **Edward J Miller**[2], **Tim Bateman**[4], **D. Dey**[1], **D. Berman**[1], **P. Slomka**[1]

[1.]Departments of Imaging and Medicine, Cedars-Sinai Medical Center, Los Angeles, CA, USA

[2.]Section of Cardiovascular Medicine, Department of Internal Medicine, Yale University School of Medicine, New Haven, CT, USA

[3.]Department of Cardiac Sciences, University of Calgary, Calgary AB, Canada

[4.]Cardiovascular Imaging Technologies LLC, Kansas City, MO, USA

Abstract

We aimed to develop a novel deep-learning based method for automatic coronary artery calcium (CAC) quantification in low-dose ungated computed tomography attenuation correction maps (CTAC). In this study, we used convolutional long-short -term memory deep neural network (conv-LSTM) to automatically derive coronary artery calcium score (CAC) from both standard CAC scans and low-dose ungated scans (CT-attenuation correction maps). We trained convLSTM to segment CAC using 9543 scans. A U-Net model was trained as a reference method. Both models were validated in the OrCaCs dataset (n=32) and in the held-out cohort (n=507) without prior coronary interventions who had CTAC standard CAC scan acquired contemporarily. Cohen's kappa coefficients and concordance matrices were used to assess agreement in four CAC score categories (very low: <10, low:10–100; moderate:101–400 and high >400). The median time to derive results on a central processing unit (CPU) was significantly shorter for the conv-LSTM model- 6.18s (inter quartile range [IQR]: 5.99, 6.3) than for UNet (10.1s, IQR: 9.82, 15.9s, p<0.0001). The memory consumption during training was much lower for our model (13.11Gb) in comparison with UNet (22.31 Gb). Conv-LSTM performed comparably to UNet in terms of agreement with expert annotations, but with significantly shorter inference times and lower memory consumption

Keywords

Coronary calcium scoring; deep learning; PET CTAC; Convolutional LSTM

[*]- indicates that authors contributed equally to this work

1. INTRODUCTION

Positron emission tomography (PET) is well established in diagnostic evaluation known and suspected coronary artery disease[1]. Studies have shown that the information on coronary artery calcification severity (CAC) predicts major adverse cardiac events independently from perfusion deficits and therefore, can improve risk stratification in comparison with PET alone. The common practice of CT attenuation correction (CTAC) provides technical possibility for contemporary coronary Calcium Score (CAC) assessment in patients undergoing PET study with acceptable agreement with standard CAC scores[2] without the need for a dedicated CAC scan, but at a cost of more difficult and time-consuming manual quantification. Several methods allowing for fully automatic CAC scoring in low dose chest CT[3], including CTAC maps [4] have been developed. We aimed to evaluate the performance of a convolutional long-short memory deep neural network (convLSTM) in CAC score quantification from CTAC maps, an architecture that was not previously used in such application.

2. METHODS

Our method involved the use of two conv-LSTM models, the first of which was trained to segment calcium, while the second model was trained to segment the heart silhouette. The combination of two models allowed us to reduce the model overcalling that was frequently caused by bone structures. The same two-stage approach was used with the reference UNet model with 3 adjacent slices for fair comparison, that has been successfully used in this application by several research teams[3,5,6].

Proposed architecture (Figure 1) is based on a modified version of hierarchical multi-scale attention[7] for semantic segmentation. The temporal information from 3 adjacent slices is aggregated in form of an attention map and fed along with the spatial information from the main segmentation block. The network was built using the PyTorch library[12].

Training data

the data originated from three centers (The University of Calgary, Calgary, Canada; Yale University, New Heaven, CT, USA and Saint Luke Hospital, Kansas City, MO, USA) participating in the multi-center REgistry of Fast Myocardial Perfusion Imaging with NExt generation SPECT (REFINE SPECT) [8] that included both standard ECG-gated non-contrast scans (n=1827) and ungated CT attenuation correction maps (n=7716).

All the training cases have been annotated on-site by two readers the with at least 5 years of experience in CAC scoring. A train-validation-internal testing spilt of 80: 10: 10 was used with equal number of cases from each CAC category in the internal validation and testing sets. The heart segmentation models were trained on a separate dataset with 419 ungated chest CT with expert annotated heart masks.

Preprocessing involved min-max thresholding between −800 Hounsfield Units (HU) and 1200 HU. The final output was a 3 class segmentation mask (Background, Coronary Calcium, Non Coronary Calcium). To counter the large class imbalance, we used subset sampling of the majority class as well as focal loss[9] as cost function between the ground truth expert reader annotation and network generated mask.

Training was performed with the Adam optimizer[10] and Kaiming He initialization[11]. The learning rate started at 10^{-3} and was dynamically reduced as the validation loss plateaued. Early stopping was utilized to reduce any form of overfitting. We used the internal testing set (10% of the training data) to optimize the model's hyperparameters.

Testing data

Data was evaluated on a held-out cohort of 507 consecutive patients that underwent PET scans with CT attenuation correction in 2018 in Cedars-Sinai Medical Center, Los Angeles, USA. The inclusion criterion was availability of contemporarily performed ECG-gated CAC scan with expert readers annotation. The exclusion criteria were history of percutaneous coronary intervention (PCI) or coronary artery bypass grafting. Additionally, the models were evaluated on a publicly available dataset with dedicated CAC scans (OrCaSc challenge[13], n=32).

Linearly weighted Cohen's kappa coefficients, concordance indices and concordance matrices were used to assess agreement between expert readers annotations and model's predictions in four CAC score categories (CAC=0, very low: CAC 1–100; moderate: CAC 101–400 and high: CAC >400). Z-test was used to assess the statistical significance of differences in Kappa values. Time to of inference was assessed on 1024 samples,16 slices each and the medians were compared using Wilcoxon rank-sum test. P-values <0.05 considered as statistically significant.

3. RESULTS

The median time to derive results on a central processing unit (CPU) was significantly shorter for the conv-LSTM model- 6.18s (inter quartile range [IQR]: 5.99, 6.3) than for UNet (10.1s, IQR: 9.82, 15.9s, p<0.0001). The time to derive results on a graphics processing unit (GPU; NVIDIA RTX 2080 Ti) was also shorter for convLSTM 111.23 milliseconds (IQR 110.8, 111.7) vs. 113.6 milliseconds (IQR 113.5,113.5) for UNet. The preprocessing took 2.24s on average regardless of the used model. The memory consumption during training was much lower for our model (13.11Gb) in comparison with UNet (22.31 Gb).

The agreement did not differ significantly between two evaluated approaches (Figure 2). Linearly weighted Kappa values were comparable between convLSTM and UNet. For the agreement on standard CAC scans the Cohen's Kappa of our model was 0.84 (95% Confidence Interval [CI]: 0.81, 0.88 while UNet achieved Kappa of 0.86 (95% CI: 0.83, 0.89, p=0.47). For the CTAC scans, the Cohen's Kappa for convLSTM was 0.64 (95% CI: 0.64, 0.72) while for UNet it was 0.67 (95% CI: 0.62.0.71, p=0.61). The agreement on publicly available OrCaCs dataset (n=32) yielded higher Kappa values for our model - 0.89 (95% CI 0.79 0.99) in comparison with UNet (0.77, 95% CI 0.64, 0.90) but the difference was not statistically significant at p=0.31)

4. NEW OR BREAKTHROUGH WORK PRESENTED

We worked on imitating radiologist approach of aggregating information from adjacent slices and creating an attention network in a hierarchical format. This approach is designed to avoid overfitting, due to the reduced number of internal weights. As a result, the model trains and infers faster with much lower memory consumption. The application of convolutional LSTM to CAC scoring has not been previously described.

5. CONCLUSIONS

We have developed a novel architecture for calcium scoring and evaluated the agreement with expert readers annotations in both gated and ungated cardiac CT scans, as well as in a publicly available OrCaCs dataset. The main advantage of this approach is the significantly reduced memory consumption for training and almost 2x faster inference times on CPU. This is crucial for the deployment of calcium scoring software on clinical workstations, that may not always be equipped with a high-end GPU. The agreement between our model's predicted CAC scores and expert annotations was not significantly different from that of UNet.

6. INFORMATION ON OTHER SUBMISSIONS

We are currently working on evaluation of prognostic values of DL-CAC scores in CTAC scans in the population of patients undergoing PET scan. This work will share the common architecture and will include the Figure 1 from the current abstract as a supplementary material.

ACKNOWLEDGEMENTS

EJM has served as a consultant for GE Healthcare and Bracco Inc, and he and his institution have received grant support from Bracco Inc. DSB, PK, SVK and PJS participated in software royalties for QPS software at Cedars-Sinai Medical Center. PJS has received research grant support from Siemens Medical Systems. DSB's institution has received grant support from HeartFlow. KP received financial support from Polish National Agency for International Academic Exchange (NAWA). All other authors have reported that they have no relationships relevant to the contents of this paper to disclose. Research reported in this publication was supported by the National Heart, Lung, And Blood Institute of the National Institutes of Health under Award Number R01HL089765. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

REFERENCES

[1]. Dilsizian V, Bacharach SL, Beanlands RS, Bergmann SR, Delbeke D, Dorbala S, et al. ASNC imaging guidelines/SNMMI procedure standard for positron emission tomography (PET) nuclear cardiology procedures 2016.

[2]. Mylonas I, Kazmi M, Fuller L, De Kemp RA, Yam Y, Chen L, et al. Measuring coronary artery calcification using positron emission tomography-computed tomography attenuation correction images. Eur Heart J Cardiovasc Imaging 2012;13(9):786–92. [PubMed: 22511812]

[3]. Zeleznik R, Foldyna B, Eslami P, Weiss J, Alexander I, Taron J, et al. Deep convolutional neural networks to predict cardiovascular risk from computed tomography. Nat Commun [Internet] 2021;12(1). Available from: 10.1038/s41467-021-20966-2

[4]. Išgum I, de Vos BD, Wolterink JM, Dey D, Berman DS, Rubeaux M, et al. Automatic determination of cardiovascular risk by CT attenuation correction maps in Rb-82 PET/CT. J Nucl Cardiol 2018;25(6):2133–42. [PubMed: 28378112]

[5]. Singh G, Al'Aref S, Lee B, Lee J, Tan S, Lin F, et al. End-to-End, Pixel-Wise Vessel-Specific Coronary and Aortic Calcium Detection and Scoring Using Deep Learning. Diagnostics 2021;11(2):215. [PubMed: 33540660]

[6]. Gogin N, Viti M, Nicodème L, Ohana M, Talbot H, Gencer U, et al. Automatic coronary artery calcium scoring from unenhanced-ECG-gated CT using deep learning. Diagn Interv Imaging 2021;(xxxx).

[7]. Tao A, Sapra K, Catanzaro B. Hierarchical Multi-Scale Attention for Semantic Segmentation 2020 [cited 2021 Aug 23];Available from: https://arxiv.org/abs/2005.10821

[8]. Slomka PJ, Betancur J, Liang JX, Otaki Y, Hu LH, Sharir T, et al. Rationale and design of the REgistry of Fast Myocardial Perfusion Imaging with NExt generation SPECT (REFINE SPECT). J Nucl Cardiol 2020;27(3):1010–21. [PubMed: 29923104]

[9]. Lin T-Y, Goyal P, Girshick R, He K, Dollár P. Focal Loss for Dense Object Detection 2017 [cited 2021 Aug 23];Available from: http://arxiv.org/abs/1708.02002

[10]. Kingma DP, Ba JL. Adam: A method for stochastic optimization 3rd Int Conf Learn Represent ICLR 2015 - Conf Track Proc 2015;1–15.

[11]. He K, Zhang X, Ren S, Sun J. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification 2015 [cited 2021 Aug 23];Available from: https://arxiv.org/abs/1502.01852

[12]. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, ... & Chintala S (2019). Pytorch: An imperative style, high-performance deep learning library. Advances in neural information processing systems, 32, 8026–8037.

[13]. Wolterink Jelmer & Leiner Tim & De Vos Bob & Coatrieux Jean-Louis & Kelm Bernd & Kondo Satoshi & Salgado Rodrigo & Shahzad Rahil & Shu Huazhong & Snoeren Miranda & Takx Richard & Van Vliet Lucas & Walsum Theo & Willems Tineke & Yang Guanyu & Zheng Yefeng & Viergever Max & Isgum Ivana. (2016). An evaluation of automatic coronary artery calcium scoring methods with cardiac CT using the orCaScore framework. Medical physics 43. 2361–2373. 10.1118/1.4945696. [PubMed: 27147348]

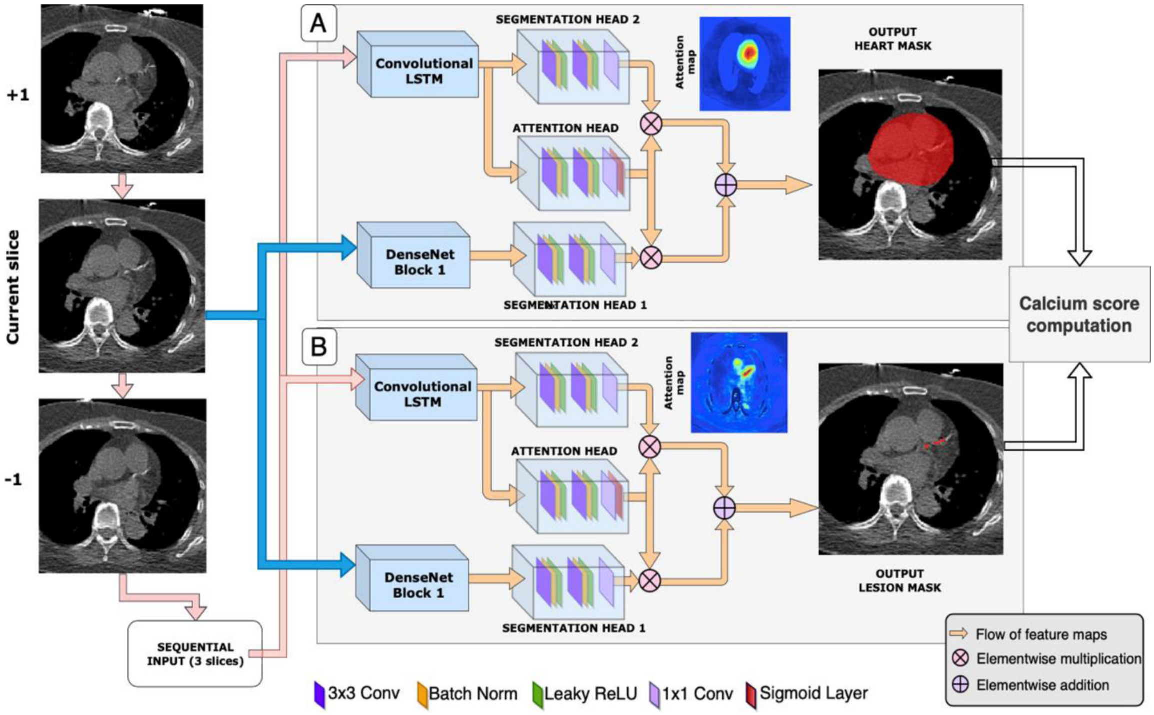

Figure 1. Architecture of the convolutional long-short term deep neural network.

Our solution consists of two networks, first of which (A) was trained for segmentation of the heart silhouette while the second network (B) was trained to segment the coronary artery calcium using expert annotations. The heart mask was applied to the final CAC prediction to reduce spurious bone overcalling.

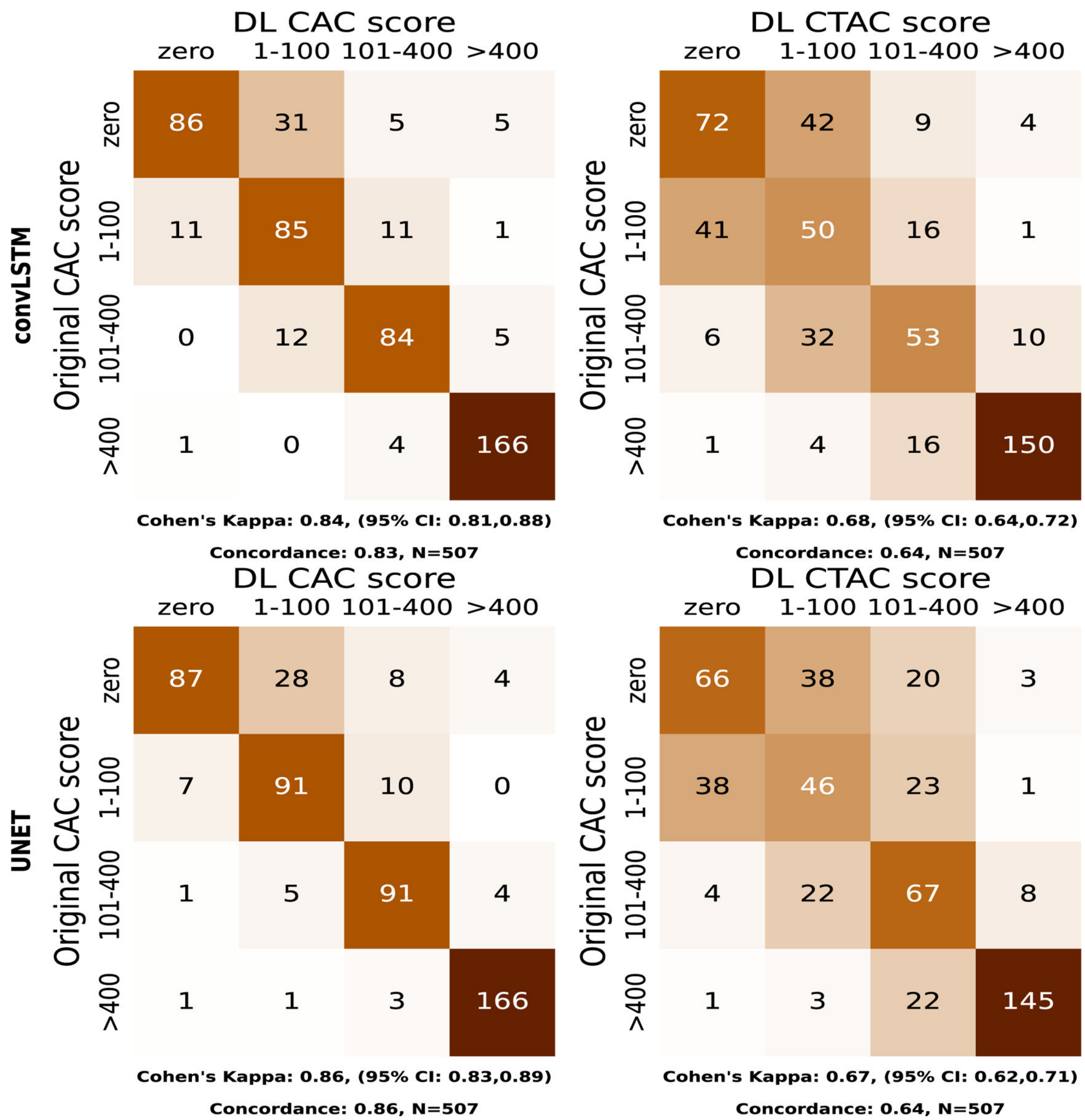

Figure 2. Agreement between deep learning models predictions and expert readers in four calcium score classes for convLSTM and Unet.

First column – agreement between expert annotated CAC scans and DL scores obtained from the CAC scans.

Second column – agreement between expert annotated CAC scans and DL prediction on contemporarily acquired CTAC maps CAC – coronary artery calcium, CTAC – CT attenuation correction, DL – deep learning

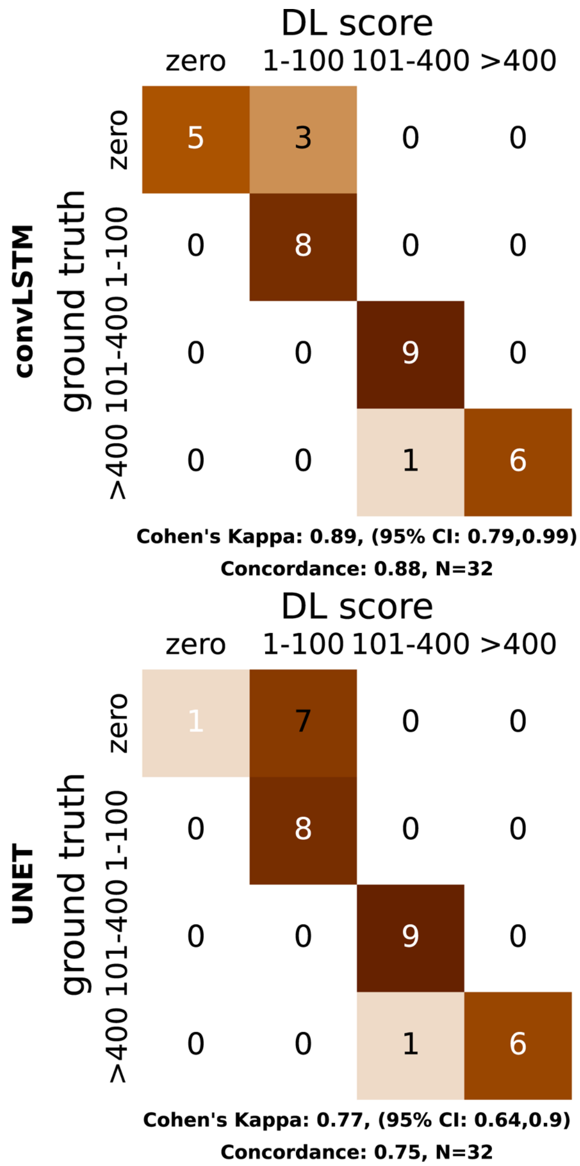

Figure 3. Agreement between expert annotations and DL predictions on the OrCaCs dataset.
CAC – coronary artery calcium, CTAC – CT attenuation correction, DL – deep learning