RESEARCH PAPER

Deep Learning for Automated Coronary Artery Calcium Scoring in Chest CTs

Muhammad Hussain[1*], Hadi Yousuf Saeed[2], Muhammad Imran Farid[3], Ismoilov Ortiq Ismoilovich[4], Sanjay Kumar[5], Hafiz Usama Talha[6]

[1*]Assistant Professor, Department of Cardiology, Abwa Medical College, Faisalabad
Email: syedmhg@gmail.com
[2]Assistant Professor, Department of Cardiology, Chaudhary Pervaiz Elahi Institute of Cardiology, Multan
[3]Assistant Professor, Department of Electrical and Computer Engineering, Air University, Islamabad
[4]Department of Human Anatomy, Samarkand State Medical University, Samarkand, Uzbekistan
[5]Senior Registrar, Consultant Cardiologist and General Physician, Cardiology Department, Suleman Roshan Medical College and Hospital, Tando Adam, District Sanghar
[6]Postgraduation Trainee, Department of Pharmacology, Islamic International Medical College, Rawalpindi

Received: 15th Jul, 2026 | Revised: 25th Jul, 2026 | Accepted: 5th Aug, 2026 | Available Online: 12th Aug, 2026

Abstract

Background: Coronary artery calcium (CAC) scoring is an established marker of coronary atherosclerotic burden and cardiovascular risk.

Objective: To evaluate the diagnostic accuracy, agreement, and processing efficiency of deep-learning-based automated coronary artery calcium scoring compared with expert manual Agatston scoring.

Methodology: This prospective diagnostic accuracy study was conducted at Abwa Medical College Faisalabad from June 2024 to June 2025 and included 210 adult patients undergoing non-contrast chest CT. Demographic and clinical information including age, gender, body mass index, smoking status, hypertension, diabetes mellitus, dyslipidemia, family history of cardiovascular disease, and previous cardiovascular history was recorded using a standardized data collection form.

Results: The mean age of participants was 61.8 ± 10.4 years, and 128 (61.0%) were male. The mean manual CAC score was 238.6 ± 315.4 Agatston units compared with 231.9 ± 307.8 using automated scoring (p=0.184). The mean absolute difference was 18.7 ± 24.6 units. Automated and manual scores showed a very strong correlation (r=0.967; p<0.001) and excellent agreement (ICC=0.974, 95% CI: 0.966–0.981). Automated scoring correctly classified 195 (92.9%) patients into the appropriate CAC risk category, with a weighted kappa of 0.904 (95% CI: 0.861–0.947; p<0.001).

Conclusion: It is concluded that deep-learning-based automated CAC scoring demonstrates excellent agreement with expert manual Agatston scoring, high diagnostic accuracy, reliable risk-category classification, and substantially faster processing.

Keywords: Coronary artery calcium; deep learning; artificial intelligence; chest CT; Agatston score; cardiovascular risk.

How to cite this article: Muhammad Hussain1* Hadi Yousuf Saeed2 Muhammad Imran Farid3 Ismoilov Ortiq Ismoilovich4 Sanjay Kumar5 Hafiz Usama Talha6. Deep Learning for Automated Coronary Artery Calcium Scoring in Chest CTs. Int J Drug Deliv Technol. 2026;16(76s):1535-1541. DOI: 10.25258/ijddt.16.76s.140.

Introduction

Coronary artery disease (CAD) still contributes to a high burden of cardiovascular morbidity and mortality globally. Early identification of subclinical coronary atherosclerosis is crucial because early preventive measures can significantly lower future cardiovascular events [1]. The established imaging biomarker of CAC, which represents the accumulation of calcified atherosclerotic plaque in the coronary arteries, is a

clinically meaningful imaging biomarker of coronary atherosclerotic burden that can provide clinically meaningful information about cardiovascular risk beyond conventional risk factors [2]. The calcium score in the coronary arteries is traditionally measured on non-contrast cardiac computed tomography (CT) performed on a dedicated machine, with computed tomography (CT gated using an electrocardiogram (ECG), using the Agatston scoring system [3]. Patients are typically classified into different categories based on the amount of calcium: no calcium, some calcium, and severe calcium. A higher CAC score is strongly correlated with myocardial infarction, cardiovascular events, and death [4]. Although proven to be clinically useful, cardiac computed tomography (CAC) is not routinely ordered in all patients who could benefit from cardiovascular risk assessment. Chest computed tomography (CT) is commonly used in many clinical settings, such as assessing lung diseases, cancer screening, infections, trauma, and other thoracic conditions [5]. These scans often include the coronary arteries in the field of view and so provide potentially valuable data on coronary calcification. Manual calcium scoring, however, is time consuming and relies on trained readers, so that the routine chest CT scan is often not assessed for calcium or reported qualitatively. Thus, there is significant potential for incidental cardiovascular risk assessment to go untapped.

AI, especially deep learning, has proven to be a potential solution to automated analysis of medical imaging data. CNNs can detect complex imaging characteristics, localize anatomical structures, segment coronary calcifications, and perform quantitative measurements with minimal human participation [7]. The use of these algorithms in routine chest CT could enable automated measurement of CAC scores without extra imaging or radiation dose or significant radiologist workload, as reported [8]. Automated CAC scoring may have some clinical benefits. A deep-learning system can analyse large volumes of CT exams with high consistency, limit interobserver variability, have the potential to report calcium in the coronary artery, and quantify cardiovascular risk assessment as part of a chest CT interpretation. [9] These systems could be especially useful in high-volume radiology practice and screening programs where manual CAC of all eligible CT exams is not feasible. Strong correlation has previously been shown between the deep-learning-based CAC measurements and manual Agatston scoring by experts, with promising sensitivity and specificity for the detection of clinically relevant coronary calcification [10]. Automated techniques have also shown potential to classify patients into known CAC risk groups and identify patients with moderate or severe coronary calcification that could benefit from other cardiovascular evaluations [11]. The performance of the algorithm, however, can be affected by the CT acquisition parameters, slice thickness, reconstruction technique, motion artifacts, contrast administration, scanner characteristics, and patient-related factors [12]. Thus, validation with expert manual assessment is still necessary before the use of automated CAC scoring in the clinical routine. Continuous calcium score correlation should not be the only parameter evaluated, but also agreement within clinically relevant CAC categories, diagnostic performance to detect coronary calcification, and factors associated with automated calcium score errors should also be evaluated [13].

Objective

To evaluate the diagnostic accuracy, agreement, and processing efficiency of deep-learning-based automated coronary artery calcium scoring compared with expert manual Agatston scoring.

Methodology

This prospective diagnostic accuracy study was conducted at Abwa Medical College Faisalabad from June 2024 to June 2025 and included 210 adult patients undergoing non-contrast chest computed tomography (CT). In patients older than 40 years, of both sexes, who had had clinically indicated non-contrast chest CT with good visualization of the coronary arteries, patients were included. The following criteria applied for examinations to be counted: Full computed tomography (CT) data sets of adequate image quality for automated and manual computed tomography analysis of coronary artery (CAC) assessment. Patients who had received any coronary artery bypass surgery, coronary stenting, or cardiac surgery that significantly changed the coronary anatomy were excluded, as were those with a history of severe motion or respiratory artifacts, patients with an insufficient visualization of the heart, those with technically poor-quality CT images, and those in whom clinical data were incomplete.

Data Collection

Once approved by the IEBC and after obtaining written informed consent, the eligible patients were recruited through a non-probability consecutive sampling technique. Demographic and clinical data (age, sex, BMI, smoking habits, hypertension, diabetes mellitus, dyslipidaemia, family history of

cardiovascular disease and previous cardiovascular history) were obtained on a standard data collection form. Chest CT was performed on a multidetector CT scanner and according to the institutional non-contrast chest CT protocol. Standardised slice thickness was used to reconstruct images and transferred to the imaging workstations. All scans were anonymised before analysis. Two highly experienced radiologists with no knowledge of the automated results independently scored the CAC images using manual techniques. The conventional threshold (which was used to detect the calcified coronary lesions) was ≥130 Hounsfield units, while the Agatston method was used to quantify the calcified coronary lesions. Consensus was reached when disagreements arose amongst readers. The consensus Agatston score was taken as the reference standard. These same CT scans were evaluated using a convolutional neural network (CNN) architecture-based deep-learning algorithm that was trained to detect coronary arteries, heart, calcified lesions, non-coronary calcifications, and automatically derive the total CAC score. All examinations were automatically processed by the algorithm, and the resulting automated Agatston score and processing time were noted. Patients were classified as having no, mild (CAC 0 to 99), moderate (CAC 100 to 399), and severe (CAC ≥400) coronary calcification for both manual and automated assessment. The basic outcome was the consensus of automated and manual CAC scores by expert eye. Secondary outcomes were the correlation between continuous automated and manual Agatston scores; the degree of agreement among the various CAC risk categories; sensitivity, specificity, positive predictive value, negative predictive value, and overall diagnostic accuracy for the detection of CAC. Processing time for automated and manual assessment and factors associated with clinically significant automated assessment errors were also assessed.

Statistical Analysis

Data were analysed using SPSS 26.0 and suitable statistical software for diagnostic accuracy analysis. The continuous variables were presented as mean ± standard deviation or median (interquartile range) depending on the distribution of the data and categorical variables were expressed as frequencies and percentages. Spearman's correlation coefficient was used to compare automated and manual CAC score. In order to assess agreement between continuous measurements, ICC was calculated and Bland–Altman analysis performed, and weighted Cohen's kappa was used to assess agreement between CAC categories. These measures (sensitivity, specificity, positive and negative predictive values, and diagnostic accuracy) were derived from expert manual scoring as the reference standard. The processing time has been compared between automated and manual methods, either by paired t-test or Wilcoxon signed-rank test, depending on the distribution. The p value of < 0.05 was taken as statistically significant.

Results

The study included 210 patients with a mean age of 61.8 ± 10.4 years and a mean BMI of 27.6 ± 4.3 kg/m². Males comprised 128 (61.0%) participants, while 82 (39.0%) were female. Hypertension was present in 118 (56.2%) patients, dyslipidemia in 91 (43.3%), diabetes mellitus in 74 (35.2%), and current smoking in 52 (24.8%). A family history of cardiovascular disease was reported in 47 (22.4%) patients, while 39 (18.6%) had previous cardiovascular disease.

Table 1. Baseline Demographic and Cardiovascular Risk Characteristics of Study Participants (n=210)

Characteristic	n (%) / Mean ± SD
Age (years), Mean ± SD	61.8 ± 10.4
Male gender	128 (61.0%)
Female gender	82 (39.0%)
BMI (kg/m²), Mean ± SD	27.6 ± 4.3
Hypertension	118 (56.2%)
Diabetes mellitus	74 (35.2%)
Dyslipidemia	91 (43.3%)
Current smoking	52 (24.8%)
Family history of cardiovascular disease	47 (22.4%)
Previous cardiovascular disease	39 (18.6%)

The mean manual CAC score was 238.6 ± 315.4 Agatston units compared with 231.9 ± 307.8 using the automated method, with no significant difference (p=0.184). Median scores were also similar at 112 (IQR: 24–318) and 108 (IQR: 21–306), respectively (p=0.216). The mean absolute difference was only 18.7 ± 24.6 Agatston units. Automated and manual scores demonstrated a very strong correlation (r=0.967; p<0.001) and excellent agreement, with an intraclass correlation coefficient of 0.974 (95% CI: 0.966–0.981).

Table 2. Comparison and Agreement Between Manual and Automated CAC Scores

Parameter	Result
Manual CAC score, Mean ± SD	238.6 ± 315.4
Automated CAC score, Mean ± SD	231.9 ± 307.8
Comparison of mean CAC scores	p=0.184
Manual CAC score, Median (IQR)	112 (24–318)
Automated CAC score, Median (IQR)	108 (21–306)
Comparison of median CAC scores	p=0.216
Mean absolute difference (Agatston units), Mean ± SD	18.7 ± 24.6
Spearman correlation coefficient	r=0.967; p<0.001
Intraclass correlation coefficient	0.974 (95% CI: 0.966–0.981)
Mean Bland–Altman bias	−6.7 Agatston units
95% limits of agreement	−55.1 to 41.7

Among patients with a manual CAC score of 0, 49 of 51 (96.1%) were correctly classified. Correct classification was achieved in 44 of 48 (91.7%) patients with scores of 1–99, 57 of 63 (90.5%) with scores of 100–399, and 45 of 48 (93.8%) with scores ≥400. Automated category distributions closely matched manual scoring, with 53 (25.2%), 49 (23.3%), 61 (29.0%), and 47 (22.4%) patients assigned to the four categories, respectively.

Table 3. Agreement of Manual and Automated Scoring Across CAC Risk Categories

CAC category (Agatston units)	Manual scoring, n (%)	Automated scoring, n (%)	Correctly classified, n (%)
0	51 (24.3%)	53 (25.2%)	49 (96.1%)
1–99	48 (22.9%)	49 (23.3%)	44 (91.7%)
100–399	63 (30.0%)	61 (29.0%)	57 (90.5%)
≥400	48 (22.9%)	47 (22.4%)	45 (93.8%)
Overall	210 (100%)	210 (100%)	195 (92.9%)

The automated system demonstrated strong diagnostic performance for detecting coronary calcium, with a sensitivity of 96.2%, specificity of 94.1%, positive predictive value of 98.1%, negative predictive value of 88.9%, and overall accuracy of 95.7%. For clinically relevant CAC ≥100, sensitivity and specificity were 94.6% and 95.9%, respectively, with an AUC of 0.972 (95% CI: 0.950–0.994), indicating excellent discrimination.

Table 4. Diagnostic Performance and Processing Efficiency of Automated CAC Scoring

Parameter	Result
Sensitivity for detecting CAC >0	96.2%
Specificity for detecting CAC >0	94.1%
Positive predictive value	98.1%
Negative predictive value	88.9%
Overall diagnostic accuracy	95.7%
Sensitivity for CAC ≥100	94.6%
Specificity for CAC ≥100	95.9%
AUC for detecting CAC ≥100	0.972 (95% CI: 0.950–0.994)
Manual processing time (minutes), Mean ± SD	4.8 ± 1.3
Automated processing time (seconds), Mean ± SD	18.6 ± 5.4
Processing-time comparison	p<0.001

Discussion

The current study compared the accuracy of automated coronary artery calcium scoring using a deep-learning algorithm on 210 chest computed tomography (CT) scans. The mean age of the study group was 61.8 ± 10.4 years and the male predominance was 61.0%. The prevalence of hypertension (56.2%), dyslipidaemia (43.3%), diabetes mellitus (35.2%) and current smoking (24.8%) suggests that there is a significant burden of cardiovascular risk factors in the population. Similarly, previous studies have assessed automated CAC scoring in primarily in middle-aged and older adults with typical cardiovascular comorbidities, in which incidental coronary calcification may be clinically pertinent [14]. In automated CAC measurements, the agreement between the measurements and the expert manual scoring was very close. The mean manual score was 238.6 ± 315.4 Agatston score vs 231.9 ± 307.8 using the deep-learning method; this difference was not statistically significant. The mean absolute difference was only 18.7 ± 24.6 Agatston units. Furthermore, a high level of agreement between automated and manual scores

was found, both for the correlation (r=0.967) and the intraclass correlation coefficient (0.974). Similar high level of intraclass agreement and strong correlation between the deep-learning derived CAC scores and the manual Agatston scores have been previously reported, indicating the reliability of automated quantification [15][16]. Bland–Altman analysis showed only a small mean bias of −6.7 Agatston units with 95% limits of agreement between −55.1 and 41.7. This suggests the automated system underestimated calcium burden in general, but the bias was small compared to the large variation in CAC among the study population. Past studies also have revealed small systematic underestimation or overestimation, typically without significantly altering clinical risk categorization, from automated CAC algorithms [17]. The deep-learning model also exhibited excellent predictive performance for the clinically relevant classification of CAC categories. In total, 92.9% (195/210) of patients were classified in the same risk group as expert manual scoring. Correct classification was achieved in 96.1% of patients with a CAC score of 0, 91.7% with scores of 1–99, 90.5% with scores of 100–399, and 93.8% with scores ≥400. There was an excellent categorical agreement as measured by weighted Cohen's kappa (0.904). Similarly, there is high agreement for the risk category classification, which is clinically relevant as treatment decisions are often made based on risk category classification, not on actual CAC scores [18].

The diagnostic performance was also very good. The automated system has a perfect sensitivity of 96.2%, a specificity of 94.1%, a positive predictive value of 98.1%, a negative predictive value of 88.9%, and an overall accuracy of 95.7% for detecting any coronary calcium. The sensitivity of the CAC ≥100 was 94.6%, the specificity was 95.9%, and the AUC was 0.972 for clinically significant CAC ≥100. These studies have shown similar diagnostic accuracy, with high sensitivity and specificity for moderate to severe coronary calcification on routine chest CT, and deep-learning systems generally have been shown to achieve high sensitivity and specificity for moderate to severe coronary calcification in previous studies [19]. One of the most significant practical conclusions was that there was a significant decrease in processing time. It took 18.6 ± 5.4 seconds to score automatically, versus 4.8 ± 1.3 minutes for manual scoring. This difference was very significant and equates to a significant increase in workflow efficiency. In high-volume CT services and screening programs, AI-driven CAC scoring has also been found to significantly cut down radiologist workload by analyzing and processing a vast array of examinations quickly and accurately, potentially saving time and manpower [20]. The low number of misclassifications could be associated with technical and anatomical issues, including motion artifacts, partial-volume effects, calcification of nearby cardiac structures, and coronary calcification which is confused with aortic or valvular calcium. All of these limitations have been previously documented and have been stressing the need for external validation between various scanners, different reconstruction parameters, and different patient groups before widespread clinical use [21]. The present study was a single-centre study with a small number of patients, which could limit the generalizability of the results. The deep-learning model was tested using a particular CT acquisition protocol, so results might differ in other scanners, CT reconstruction methods and patient groups. Automated scoring errors may also be caused by motion artifacts and calcification of cardiac structures adjacent to the area being scored. Additionally, there was no assessment of the effect of automated CAC scoring on subsequent clinical care or long-term cardiovascular outcomes.

Conclusion

It is concluded that deep-learning-based automated coronary artery calcium scoring demonstrated excellent agreement with expert manual Agatston scoring, high diagnostic accuracy, and reliable classification across clinically relevant CAC risk categories. The automated system also markedly reduced processing time, highlighting its potential to improve radiology workflow efficiency. Integration of automated CAC scoring into routine chest CT interpretation may facilitate opportunistic cardiovascular risk assessment without requiring additional imaging or radiation exposure.

References

1. Eng, D., Chute, C., Khandwala, N., Rajpurkar, P., Long, J., Shleifer, S., ... & Patel, B. N. (2021). Automated coronary calcium scoring using deep learning with multicenter external validation. *NPJ digital medicine*, *4*(1), 88.
2. van Velzen, S. G., Lessmann, N., Velthuis, B. K., Bank, I. E., van den Bongard, D. H., Leiner, T., ... & Išgum, I. (2020). Deep learning for automatic calcium scoring in CT: validation using multiple cardiac CT and chest CT protocols. *Radiology*, *295*(1), 66-79.
3. Wang, W., Wang, H., Chen, Q., Zhou, Z., Wang, R., Zhang, N., ... & Xu, L. (2020).

Coronary artery calcium score quantification using a deep-learning algorithm. *Clinical radiology*, *75*(3), 237-e11.

4. Abdelrahman, K., Shiyovich, A., Huck, D. M., Berman, A. N., Weber, B., Gupta, S., ... & Blankstein, R. (2024). Artificial intelligence in coronary artery calcium scoring detection and quantification. *Diagnostics*, *14*(2), 125.
5. Kim, S., Park, E. A., Ahn, C., Jeong, B., Lee, Y. S., Lee, W., & Kim, J. H. (2025). Performance of fully automated deep-learning-based coronary artery calcium scoring in ECG-gated calcium CT and non-gated low-dose chest CT. *European Radiology*, *35*(11), 7084-7095.
6. Martin, S. S., van Assen, M., Rapaka, S., Hudson Jr, H. T., Fischer, A. M., Varga-Szemes, A., ... & Schoepf, U. J. (2020). Evaluation of a deep learning–based automated CT coronary artery calcium scoring algorithm. *Cardiovascular Imaging*, *13*(2_Part_1), 524-526.
7. van Assen, M., Martin, S. S., Varga-Szemes, A., Rapaka, S., Cimen, S., Sharma, P., ... & Schoepf, U. J. (2021). Automatic coronary calcium scoring in chest CT using a deep neural network in direct comparison with non-contrast cardiac CT: a validation study. *European journal of radiology*, *134*, 109428.
8. Mineo, E., Assuncao-Jr, A. N., da Silva, C. F. G., Liberato, G., Dantas-Jr, R. N., de Vita Graves, C., ... & Nomura, C. H. (2025). Automated coronary artery calcium scoring using deep learning: validation across diverse chest CT protocols. *Academic Radiology*.
9. Lee, H., Martin, S., Burt, J. R., Bagherzadeh, P. S., Rapaka, S., Gray, H. N., ... & Schoepf, U. J. (2020). Machine learning and coronary artery calcium scoring. *Current Cardiology Reports*, *22*(9), 90.
10. Gogin, N., Viti, M., Nicodème, L., Ohana, M., Talbot, H., Gencer, U., ... & Mousseaux, E. (2021). Automatic coronary artery calcium scoring from unenhanced-ECG-gated CT using deep learning. *Diagnostic and interventional imaging*, *102*(11), 683-690.
11. Winkel, D. J., Suryanarayana, V. R., Ali, A. M., Görich, J., Buß, S. J., Mendoza, A., ... & Rapaka, S. (2022). Deep learning for vessel-specific coronary artery calcium scoring: validation on a multi-centre dataset. *European Heart Journal-Cardiovascular Imaging*, *23*(6), 846-854.
12. Wang, T. W., Tzeng, Y. H., Wu, K. T., Liu, H. R., Hong, J. S., Hsu, H. Y., ... & Wu, Y. T. (2024). Meta-analysis of deep learning approaches for automated coronary artery calcium scoring: Performance and clinical utility AI in CAC scoring: A meta-analysis: AI in CAC scoring: A meta-analysis. *Computers in biology and medicine*, *183*, 109295.
13. Rathore, S., Gautam, A., Raghav, P., Subramaniam, V., Gupta, V., Rathore, M., ... & Iyengar, S. (2025). Fully automated coronary artery calcium score and risk categorization from chest CT using deep learning and multiorgan segmentation: a validation study from National Lung Screening Trial (NLST). *IJC Heart & Vasculature*, *56*, 101593.
14. Xu, C., Guo, H., Xu, M., Duan, M., Wang, M., Liu, P., ... & Wang, Y. (2022). Automatic coronary artery calcium scoring on routine chest computed tomography (CT): comparison of a deep learning algorithm and a dedicated calcium scoring CT. *Quantitative Imaging in Medicine and Surgery*, *12*(5), 2684.
15. Choi, J. H., Cha, M. J., Cho, I., Kim, W. D., Ha, Y., Choi, H., ... & Chang, J. S. (2022). Validation of deep learning-based fully automated coronary artery calcium scoring using non-ECG-gated chest CT in patients with cancer. *Frontiers in Oncology*, *12*, 989250.
16. de Vos, B. D., Lessmann, N., de Jong, P. A., & Išgum, I. (2021). Deep learning–quantified calcium scores for automatic cardiovascular mortality prediction at lung screening low-dose CT. *Radiology: Cardiothoracic Imaging*, *3*(2), e190219.
17. Vonder, M., Zheng, S., Dorrius, M. D., van der Aalst, C. M., de Koning, H. J., Yi, J., ... & Oudkerk, M. (2022). Deep learning for automatic calcium scoring in population-based cardiovascular screening. *Cardiovascular Imaging*, *15*(2), 366-367.
18. Kim, J., Wu, H., Lee, J., Hoori, A., Hu, T., Makhlouf, M., & Wilson, D. (2025). Cardiovascular Risk Prediction Using End-toend Deep Learning And Hand-crafted Features On Ct Calcium Score Scans. *Journal of Cardiovascular Computed Tomography*, *19*(4), S14.
19. Kamel, P. I., Yi, P. H., Sair, H. I., & Lin, C. T. (2021). Prediction of coronary artery calcium and cardiovascular risk on chest

radiographs using deep learning. *Radiology: Cardiothoracic Imaging*, *3*(3), e200486.

20. Hu, J., Hao, G., Xu, J., Wang, X., & Chen, M. (2024). Deep learning-based coronary artery calcium score to predict coronary artery disease in type 2 diabetes mellitus. *Heliyon*, *10*(6).
21. Yuan, N., Kwan, A. C., Duffy, G., Theurer, J., Chen, J. H., Nieman, K., ... & Ouyang, D. (2023). Prediction of coronary artery calcium using deep learning of echocardiograms. *Journal of the American Society of Echocardiography*, *36*(5), 474-481.