Canadian Journal of Cardiology 38 (2022) 160–168

Clinical Research

A Deep-Learning Algorithm-Enhanced System Integrating Electrocardiograms and Chest X-rays for Diagnosing Aortic Dissection

Wei-Ting Liu, MD,[a] Chin-Sheng Lin, MD, PhD,[b] Tien-Ping Tsao, MD,[b,c] Chia-Cheng Lee, MD, PhD,[d] Cheng-Chung Cheng, MD,[b] Jiann-Torng Chen, MD, PhD,[e] Chien-Sung Tsai, MD,[f] Wei-Shiang Lin, MD,[b] and Chin Lin, PhD[g]

[a] *Department of Internal Medicine, Tri-Service General Hospital, National Defense Medical Center, Taipei, Taiwan*

[b] *Division of Cardiology, Department of Internal Medicine, Tri-Service General Hospital, National Defense Medical Center, Taipei, Taiwan*

[c] *Division of Cardiology, Cheng Hsin General Hospital, Taipei, Taiwan*

[d] *Division of Colorectal Surgery, Department of Surgery, Tri-Service General Hospital, National Defense Medical Center, Taipei, Taiwan*

[e] *Department of Ophthalmology, Tri-Service General Hospital, National Defense Medical Center, Taipei, Taiwan*

[f] *Division of Cardiovascular Surgery, Department of Surgery, Tri-Service General Hospital, National Defense Medical Center, Taipei, Taiwan*

[g] *School of Medicine, National Defense Medical Center, Taipei, Taiwan*

ABSTRACT

Background: Chest pain is the most common symptom of aortic dissection (AD), but it is often confused with other prevalent cardiopulmonary diseases. We aimed to develop deep-learning models (DLMs) with electrocardiography (ECG) and chest x-ray (CXR) features to detect AD and evaluate their performance.

Methods: This study included 43,473 patients in the emergency department (ED) between July 2012 and December 2019 for retrospective DLM development. A development cohort including 49,071 ED records (120 AD type A and 64 AD type B) was used to train DLMs for ECG and CXR, and 9904 independent ED records (40 AD type A and 34 AD type B) were used to validate DLM performance. Human-machine competitions of ECG and CXR were conducted. Patient characteristics and laboratory results were used to enhance the diagnostic accuracy. The DLM-enabled AD diagnostic process was prospectively evaluated in 25,885 ED visits.

RÉSUMÉ

Contexte : La douleur thoracique est le symptôme le plus courant de la dissection aortique (DA), mais celle-ci est souvent confondue avec d'autres maladies cardiopulmonaires fréquentes. Notre objectif était de développer des modèles d'apprentissage profond (MAP) permettant l'analyse de données d'électrocardiographie (ECG) et de radiographie pulmonaire en vue de détecter la DA et d'évaluer leurs performances.

Méthodologie : Cette étude portait sur le développement rétrospectif de MAP à partir des données recueillies chez 43 473 patients admis aux urgences entre juillet 2012 et décembre 2019. Une cohorte de développement comportant 49 071 dossiers des urgences (120 cas de DA de type A et 64 cas de DA de type B) a servi à entraîner les MAP à l'analyse des données d'ECG et de radiographie pulmonaire, et 9 904 dossiers des urgences indépendants (40 cas de DA de type A et 34 cas de DA de type B) ont été utilisés pour valider les performances des

Chest pain is one of the most common complaints in patients visiting the emergency department (ED).[1] Life-threatening etiologies of chest pain—such as acute coronary syndrome, aortic dissection, pulmonary embolism, cardiac tamponade, and tension pneumothorax—pose a great challenge for physicians and require many screening tools, including electrocardiograms (ECGs), chest X-rays (CXRs), and laboratory testing for rapid differentiation.[2] Among these fatal causes, aortic dissection (AD) is associated with very high rates of morbidity and mortality, and the mortality rate increases 1% to 2% per hour after initial symptoms and increases to up to 50% if unrecognized within 48 hours.[3] However, several diagnostic difficulties—including low prevalence, atypical presentations, lack of specific biomarkers, and similarity to other acute conditions—lead to delayed diagnosis, with an average of 4.3 hours from arrival at the ED to diagnosis.[4,5] The misdiagnosis rates of AD range from 14% to 39%,

Received for publication May 26, 2021. Accepted September 27, 2021.

Corresponding author: Dr Chin Lin, School of Medicine, National Defense Medical Center, No. 161, Min-Chun E. Road, Sec. 6, Neihu, Taipei 114, Taiwan, ROC. Tel.: 886-2-87923100#18574.

E-mail: xup6fup@mail.ndmctsgh.edu.tw

See page 167 for disclosure information.

https://doi.org/10.1016/j.cjca.2021.09.028

Results: The area under the curves (AUCs) of the ECG and CXR models were 0.918 and 0.857 for detecting AD in a human-machine competition, respectively, which were better than those of the participating physicians. In the validation cohort, the AUCs of the integrated model were 0.882, 0.960, and 0.813 in all AD, AD type A, and AD type B patients, respectively, with a sensitivity of 100.0% and a specificity of 81.7% for AD type A. In patients with chest pain and D-dimer tests, the DLM could predict more precisely, achieving a positive predictive value of 62.5% in the prospective evaluation.

Conclusions: DLMs may serve as decision-supporting tools for identification of AD and facilitate differential diagnosis in patients with acute chest pain.

MAP. L'analyse des données d'ECG et de radiographie pulmonaire a fait l'objet de compétitions entre humains et machines. Les caractéristiques des patients et les résultats des analyses de laboratoire ont été utilisés pour améliorer la précision du diagnostic. Le processus de diagnostic de la DA faisant appel aux MAP a été évalué de manière prospective sur 25 885 consultations aux urgences.

Résultats : La surface sous la courbe (SSC) des modèles d'analyse des données d'ECG et de radiographie pulmonaire était respectivement de 0,918 et 0,857 pour la détection de la DA dans une compétition entre humains et machine, ce qui constituait un meilleur résultat que celui obtenu par les médecins participants. Au sein de la cohorte de validation, la SSC du modèle intégré était de 0,882 dans l'ensemble des cas de DA, de 0,960 dans les cas de DA de type A et de 0,813 dans les cas de DA de type B, la sensibilité et la spécificité ayant atteint respectivement 100,0 % et 81,7 % dans les cas de DA de type A. Chez les patients présentant des douleurs thoraciques et dont le taux de D-dimères avait été mesuré, les MAP pouvaient donner des prévisions plus précises, une valeur prédictive positive de 62,5 % ayant été obtenue dans le cadre de l'évaluation prospective.

Conclusions : Les MAP peuvent servir d'outils d'aide à la décision dans le dépistage de la DA et faciliter le diagnostic différentiel chez les patients présentant des douleurs thoraciques aiguës.

leading to inappropriate treatment, such as administration of antiplatelet or anticoagulant agents, which results in an increased risk of major bleeding and mortality.[6,7] Early recognition of AD followed by prompt surgical or medical treatment are key to improving patient survival.

Although the diagnostic imaging techniques for AD recommended in current guidelines, including contrast-enhanced computed tomography (CT) and transesophageal echocardiography, provide > 95% sensitivity and specificity,[8] these examinations are expensive and invasive. The decision to perform advanced imaging is challenging in uncertain clinical circumstances.[4,5] In patients with AD, ECGs exhibited nonspecific ST-T changes, ST-segment depression, T-wave inversion, ST-segment elevation, or new Q waves,[9] whereas CXRs exhibited a widened mediastinum or aortic knob with sensitivities of 67% to 81%.[10,11] ECGs and CXRs have limited sensitivity to AD and cannot completely exclude the possibility of AD.[5,9] To reduce the misdiagnosis of AD and to avoid the overuse of potentially harmful tests, a precise method with high sensitivity is needed urgently.

Deep-learning techniques, a subfield of artificial intelligence (AI), have rapidly evolved in the past decade and provide a novel capability for disease diagnosis.[12] Compared with the traditional machine-learning algorithm with manual feature engineering, deep-learning models (DLMs) can extract features unrecognizable by humans with automatic feature engineering.[13] Incorporating ECGs with DLMs has been reported to enable identification of atrial fibrillation during sinus rhythm and estimation of a person's sex and age, which demonstrates the potential superhuman capability of AI in clinical diagnosis and prediction.[14,15] AI algorithms have also shown remarkable progress in image-recognition tasks with broad applications.[16] With the advancement of AI models, we aimed to use deep-learning techniques to identify features of AD on ECGs and CXRs to improve unmet needs.

In our study, we trained an ECG-based and CXR-based DLM to predict different types of AD. The performance of both trained models and physicians on AD prediction was compared, and the predictors of AD diagnosis in the study were evaluated. Finally, we analyzed the applications and predictive values of DLMs in different subgroups.

Methods

Data source

The ethical review was approved by the Institutional Review Board of Tri-Service General Hospital, Taipei, Taiwan (IRB A202005151 and C202105049). Each sample in this study was based on 1 ED record, which received at least 1 ECG test or at least 1 CXR test. We only used the first ECG or CXR test for each record and excluded the following test to simulate the emergency situation. The ECG signal was recorded in a digital format with a sampling frequency of 500 Hz and 10 s for each lead. The CXR image was recorded in DICOM format with a resolution of > 3000 × 3000 pixels. The details of the collection of patient characteristics and laboratory tests are given in the Supplemental Appendix S1.

Aortic dissection

The definite diagnosis of AD was made using the following criteria: the presence of an intimal flap with a true lumen and a false lumen or the presence of intramural hematoma in the aorta as shown on CT angiography; involvement of the ascending aorta (defined as type A) or involvement confined to the aortic arch or descending aorta (type B). Intramural hematoma was defined as AD in our study because of its similarity in management and prognosis with typical AD.[8] All cases of AD were reviewed by our research team to ensure quality. Non-AD cases were collected from the ED during the same period. Patients with suspected AD without CT diagnoses were excluded from the study.

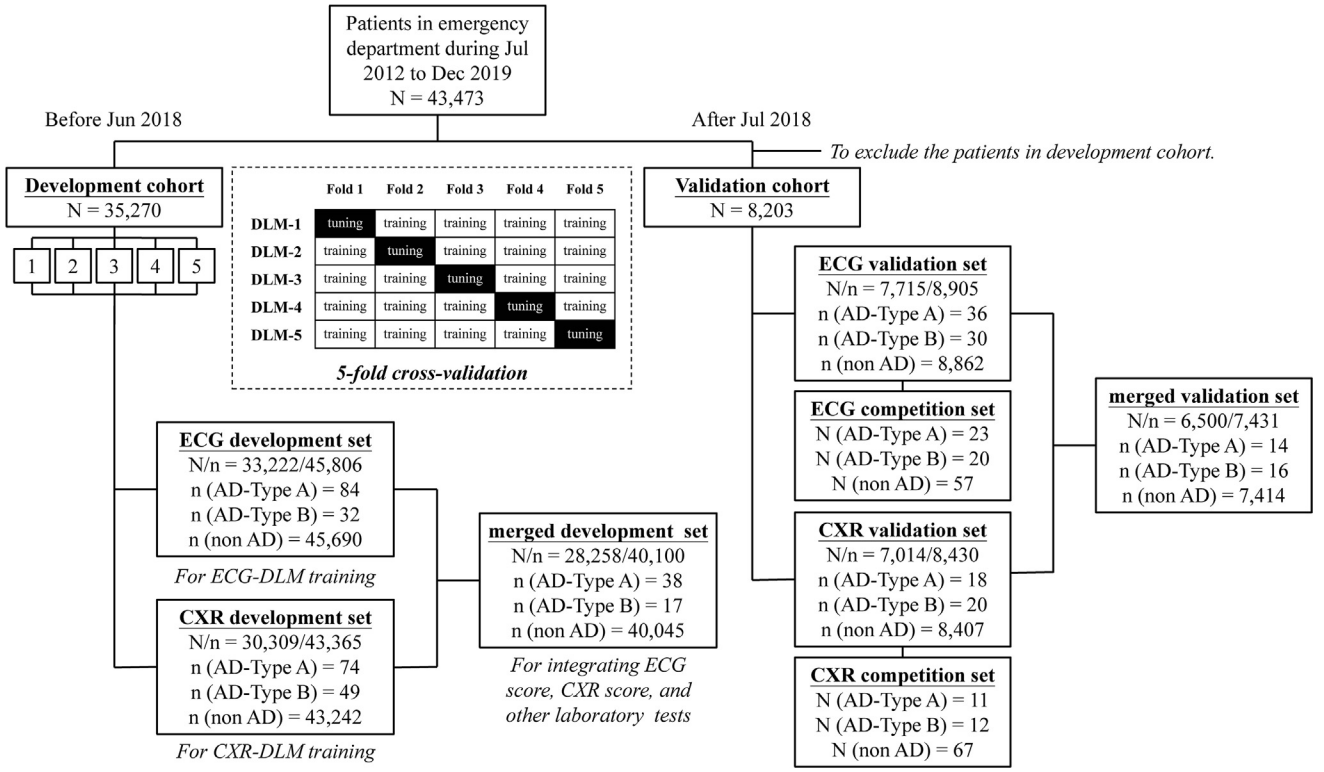

Figure 1. Development and validation cohorts' generation and summary of study process. Each ECG dataset record corresponded to an ECG, and each CXR dataset record corresponded to a CXR. Records in merged validation set were their intersection. AD, aortic dissection; CXR, chest X-ray; DLM, deep-learning model; ECG, electrocardiogram; N, number of patients; n, number of records.

Retrospective development and validation datasets

Patients who presented to our ED from July 2012 to December 2019 were included in this study, and a total of 43,473 patients were allocated to several datasets. Figure 1 illustrates the overall data set generation process. The development cohort included 35,270 patients who presented to our ED before June 2018. The patients in the development cohort were divided into 5 subgroups for 5-fold cross-validation, and the details are given in the following section. There were no overlapping patients among the 5 subgroups. The 45,806 ED records with ECG tests obtained from 33,222 patients were allocated to an ECG development set, and there were 84 AD type A cases and 32 AD type B cases. A total of 43,365 CXRs from 30,309 patients were allocated to the CXR development set, and there were 74 AD type A cases and 49 AD type B cases. These 2 development sets were used to train the corresponding DLMs. To integrate the information from ECG, CXR, and other laboratory tests, we merged these 2 two data sets into a merged development set that included 40,100 records from 28,258 patients (38 AD type A cases and 17 AD type B cases). Each record in the merged development set underwent both ECG and CXR tests.

The validation cohort included patients who presented to our ED from July 2018 to December 2019, which excluded patients in the development cohort. A total of 8203 patients were included in the validation cohort. The 8905 ED records with ECG tests from 7715 patients were allocated to an ECG validation set, and there were 36 AD type A cases and 30 AD type B cases. The 8430 records with CXRs from 7014 patients were allocated to a CXR validation set, and there were 18 AD type A cases and 20 AD type B cases. A merged validation set included 7431 records and was allocated to both ECG tests and CXRs from 6500 patients (14 AD type A cases and 16 AD type B cases).

Prospective evaluation

A prospective cohort study, including all patients visiting the ED between February 1 and July 31, 2021, was conducted to evaluate the DLMs for AD diagnosis in clinical practice. The methods of prospective cohort study are described in the Supplemental Appendix S1.

Implementation of the DLM

The ECG-DLM was developed based on the ECG12Net architecture developed previously[17] for detection of AD. The DLM was trained using an ECG development set with 2 categories: AD and non-AD. The standard input format of ECG12Net has a length of 1024 numeric sequences. In the training stage, we randomly cropped 1024 sequences as input. In the inference stage, 9 overlapping lengths of 1024 sequence intervals sampled from the original 5000 lengths were used to generate nine probabilities for each ECG. The CXR-DLM training process was revised from a previous study,[18] which was based on an architecture called 121-layer DenseNet.[19] We resized our CXR images to allow the short side to be 256 pixels, without changing the aspect ratio. In the training stage, we randomly cropped a 224 × 224 pixel image as input and applied a random lateral inversion with 50% probability. In

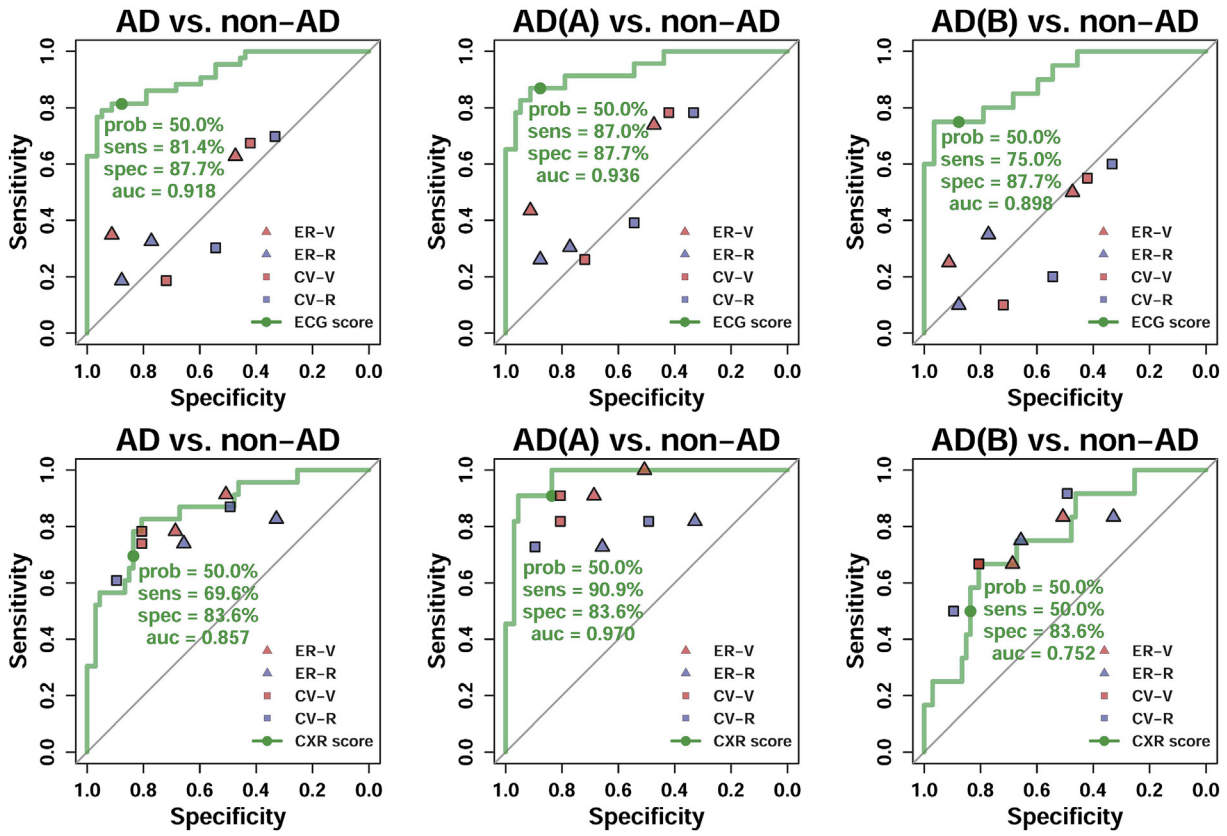

Figure 2. Performance of recognition of aortic dissection in the human-machine competitions. The top and bottom panels show ECG competition and CXR competition sets, respectively. The **red** and **blue** points represent visiting staffs and residents, respectively. The **triangular** and **square marks** represent the emergency physicians and cardiologists, respectively. AD, aortic dissection; AUC, area under the curve; CV-R, Resident in Cardiology; CV-V, Visiting Staff in Cardiology; CXR, chest X-ray; ECG, electrocardiogram; ER-R, Resident in Emergency Department; ER-V, Visiting Staff in Emergency Department; prob, probability; sens, sensitivity; spec, specificity.

the inference stage, a 10-crop evaluation was used to generate 10 probabilities for each CXR.[20] The optimal details are provided in the Supplemental Appendix S1.

We used 5-fold cross-validation, in which 4 images were used as the training subset and 1 as the tuning subset in each fold for totally trained 5 ECG DLMs and 5 CXR DLMs. Models M1 to M5 were derived from 5-fold cross-validation, and model M0 was the mean prediction based on M1 to M5. The generation of the ECG and CXR scores is summarized in the Supplemental Appendix S1 and Supplemental Figure S1. The ECG and CXR scores were the average of standardized probabilities with a cutoff point of 0.5 for AD predicted via 5 DLMs in each training subset using 5-fold cross-validation. We used logistic regression to integrate the probabilities predicted from ECGs and CXRs in each tuning subset and generated the AI score, and we further enhanced the accuracy using laboratory data.

Statistical analysis and model performance assessment

The details of the human-machine competition and the statistical analyses are given in the Supplemental Appendix S1, and we used a significance level of $P < 0.05$ throughout the analysis. In summary, the primary analysis was to evaluate the performance of DLMs in diagnosis of AD via area under the curve (AUC), sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). The secondary analysis aimed to analyze the effects of patient characteristics on performance of DLM. Therefore, we used AUC ranking and logistic regression analysis of each characteristic for detecting AD (Supplemental Appendix S1).

Results

Primary analysis

Patient characteristics are given in the Supplemental Tables S1-S3 and Supplemental Appendix S2. The human-machine competition for AD recognition is summarized in Figure 2. The AUCs of the ECG-CXR score for predicting all AD, AD type A, and AD type B were 0.918/0.857, 0.936/0.970, and 0.898/0.752, respectively. The detection of AD type A exhibited better performance than AD type B in both the ECG score (sensitivity/specificity: 87.0%/87.7%) and CXR score (sensitivity/specificity: 90.9%/83.6%), which were better than those of all clinicians. However, the CXR score was not superior to all clinicians in detecting AD type B. Generally, both ECG and CXR scores had the highest global performance of AD recognition in the human-machine

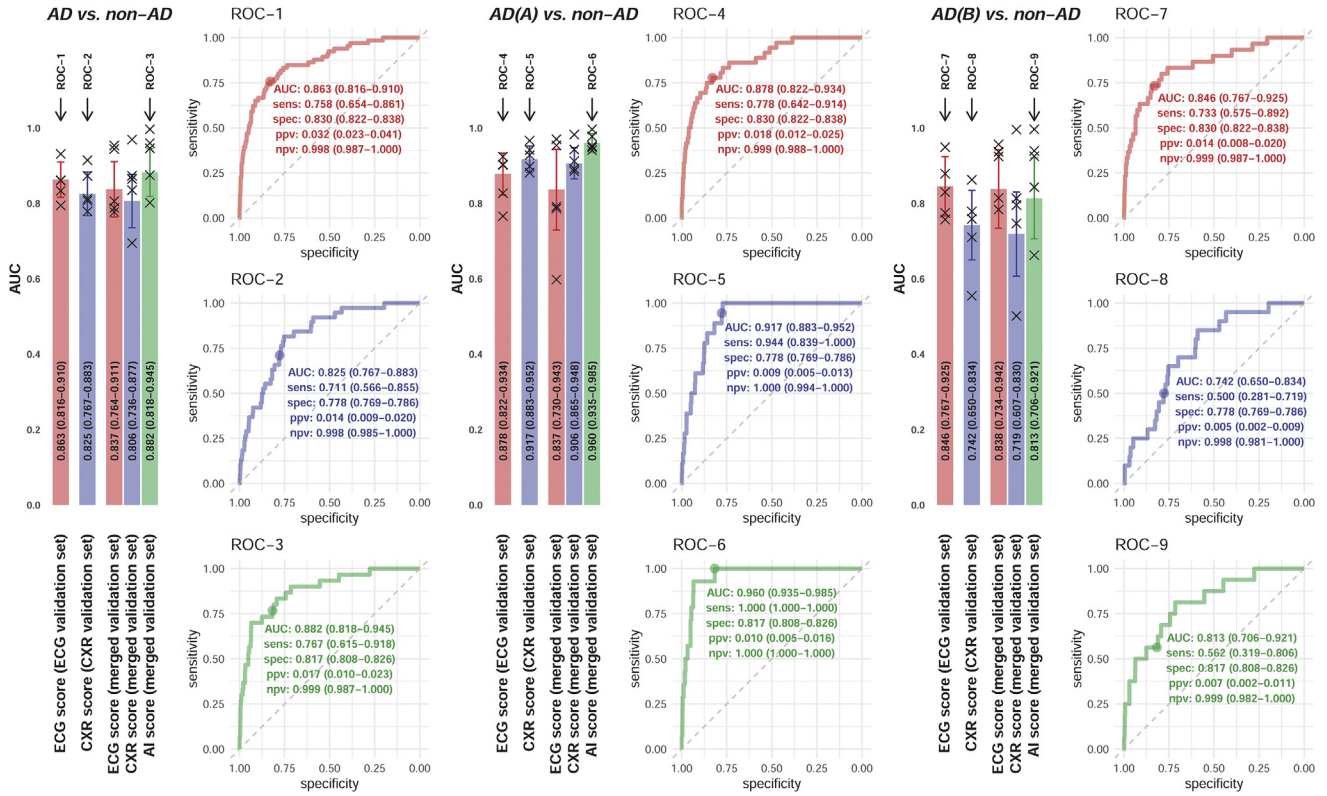

Figure 3. Performance of artificial-intelligence (AI) system. The area under the curves (AUCs) of models M1 to M5 (derived from 5-fold cross-validation) are shown as a **black x** in each tuning subset, and the AUCs of overall model M0 (the mean prediction based on M1 to M5) are shown as bar heights with corresponding 95% confidence intervals (CIs). The left, middle, and right panels show the performance in detecting all aortic dissection (AD), AD type A, and AD type B, respectively. We selected the conditions with highest AUC to present the receiver operating characteristic (ROC) curve as shown on the right. CXR, chest X-ray; ECG, electrocardiogram; npv, negative predictive value; ppv, positive predictive value; sens, sensitivity; spec, specificity.

competition (Supplemental Fig. S2). The AUCs of models M1 to M5 (derived from 5-fold cross-validation) in the corresponding tuning sets and the model M0 in the validation sets, representing AUCs of the DLM (mean prediction based on M1 to M5), are summarized in Figure 3. The AUCs of the AI score (AUC of green bars = 0.882, 0.960, and 0.813 for all AD, AD type A, and AD type B, respectively), the combination of ECG and CXR scores, were higher than those in the separate ECG score (AUC of red bars = 0.837, 0.837, and 0.838 in all AD, AD type A, and AD type B, respectively) and CXR score (AUC of blue bars = 0.806, 0.906, and 0.719 in all AD, AD type A, and AD type B, respectively), especially in AD type A, with a sensitivity of 100.0% and a specificity of 81.7%. In the differentiation of AD type A and AD type B, the ECG, CXR, and AI scores all exhibited limited performance (Supplemental Fig. S3). We also analyzed the effect of each ECG lead on AD prediction (Supplemental Fig. S4). Lead V5 had the best AUC of 0.729 for detection of AD, which also exhibited the highest AUC of 0.754 in AD type A.

Patient characteristics for diagnosis of AD

The univariate and multivariate analyses of patient characteristics for diagnosis of AD are shown in Supplemental Figure S5. After multivariable adjustment, the D-dimer, hemoglobin, percentage of lymphocytes, and percentage of basophils all exhibited significant predictive roles. Notably, the D-dimer played the most critical role and was even a better indicator than the combination of all 4 indicators for the prediction of AD (Fig. 4A). Therefore, we used the D-dimer test to enhance the DLM for diagnosis of AD. Figure 4B shows the AUCs of ECG, CXR, D-dimer, and the combination models for AD prediction. ECGs alone could predict all types of AD, whereas CXRs alone exhibited poor performance in prediction of AD type B. Application of the D-dimer improved the performance of the ECG, CXR, and AI scores in predicting all types of AD. The AUCs of the combination of the AI score and D-dimer were 0.910, 0.963, and 0.850 for the prediction of all AD, AD type A, and AD type B, respectively. Detailed results of DLMs enhanced by the D-dimer in each validation set are shown in Supplemental Fig. S6.

AI models for diagnosis of AD

Supplemental Figure S7 summarizes the performance of DLMs in subsets of the validation cohort. In patients who underwent the D-dimer test, the PPVs of combined ECG and CXR were 3.5%, 2.5%, and 1.0% for all AD, AD type A, and AD type B in the merged validation set, respectively. The additional D-dimer combined with ECG and CXR enhanced the PPVs to 6.2%, 4.4%, and 2.0% for all AD, AD type A,

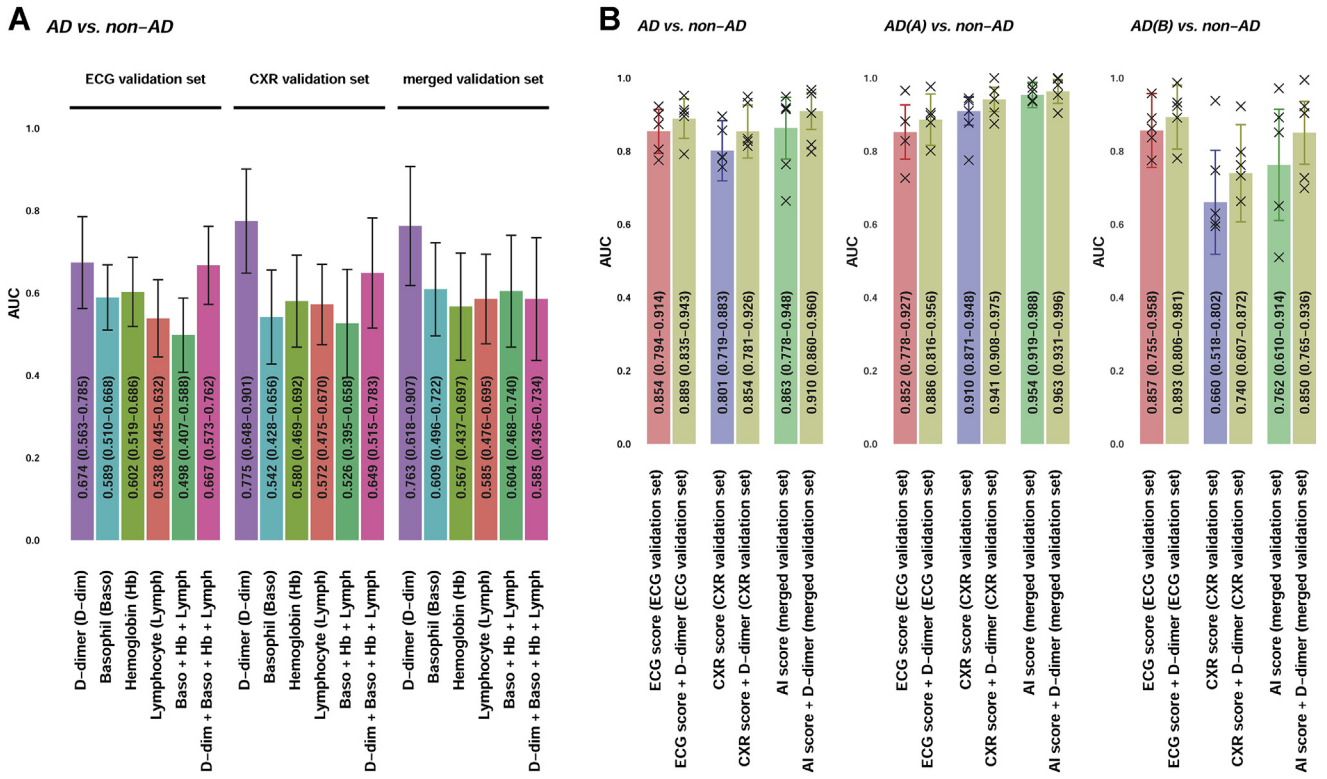

Figure 4. Performance of patient characteristics and artificial intelligence (AI) system in detecting aortic dissection (AD). (**A**) Receiver operating characteristic (ROC) curve analysis of blood-cell–related indicators and D-dimer in validation cohort. The combination scores are calculated by logistic regression. The area under the curves (AUCs) in overall development samples are shown as bar heights with corresponding 95% confidence intervals (CIs). Patients in this analysis received all related laboratory tests (n = 2510/2293/2061 in the ECG/CXR/merged validation set). (**B**) AUCs of models M1 to M5 (derived from 5-fold cross-validation) are shown as a **black x** in each tuning subset, and the AUCs of overall model M0 (the mean prediction based on M1 to M5) are shown as bar heights with corresponding 95% CIs. The left, middle, and right panels show the performance in detecting all AD, AD type A, and AD type B, respectively. CXR, chest X-ray; ECG, electrocardiogram.

and AD type B, respectively. We also stratified the patients in the validation set based on common chief complaints during triage. The calculations of sensitivities, specificities, PPVs, and NPVs of DLMs in different subgroups are given in Supplemental Figure S7 and Supplemental Tables S4-S12. Among the complaints of coma, chest pain, and abdominal pain, chest pain was the most common presentation in patients with AD.[8] In patients with AD, common ECG manifestations were nonspecific ST-T changes and left ventricular hypertrophy, whereas the CXR results exhibited prominent aortic knob and widened mediastinum (Supplemental Figs. S8-S12). A comparison of chest pain patients with all patients revealed that the AUCs of combined ECG and CXR were similar, whereas the PPVs were a factor of 2 to 4 higher in patients with chest pain. Furthermore, the combination of D-dimer, ECG, and CXR in patients with chest pain performed better than in all patients, with AUCs of 0.943 (95% confidence interval [CI], 0.898-0.989) for all AD, 0.974 (95% CI, 0.948-1.000) for AD type A, and 0.892 (95% CI, 0.787-0.996) for AD type B, with a sensitivity of 90.0%, a specificity of 93.5%, and a PPV of 15.5% in AD type A.

The prospective evaluation of DLMs in patients with or without chest pain is shown in Supplemental Figure S13. As shown in Figure 5, AI identified 8 patients (AD type A = 4, AD type B = 1, and non-AD = 3, with PPV for AD = 62.5%) in 2249 patients with chest pain and 25 patients (AD type A = 3 and non-AD = 22, with PPV for AD = 12.0%) in 23,636 patients without chest pain. The sensitivity/specificity of AD type A were 80.0%/99.9% and 75.0%/99.9% in patients with and without chest pain, respectively. AD type B was more frequently unidentified than AD type A. The performance of the DLM in the prospective cohort exhibited a lower sensitivity but a higher specificity and PPV than in the validation cohort owing to the series test-study design.

Discussion

In our study, we developed a series of DLMs with ECG and CXR analyses at the ED to detect AD. The integrated model of ECG and CXR results reached an AUC of 0.960 for detecting AD type A, with a sensitivity of 100.0% and a specificity of 81.7%. Applying our models to patients with chest pain or adding AI-enabled D-dimer analysis could further improve the accuracy of AD diagnosis. Importantly, our AI models provided an excellent diagnostic support system to exclude AD type A in patients with acute chest pain.

The ECG presentation of ST-segment elevation in patients with AD would prompt physicians to focus on acute coronary syndrome and would be a disaster once diagnosis of patients with AD is delayed and they are treated with antiplatelet and

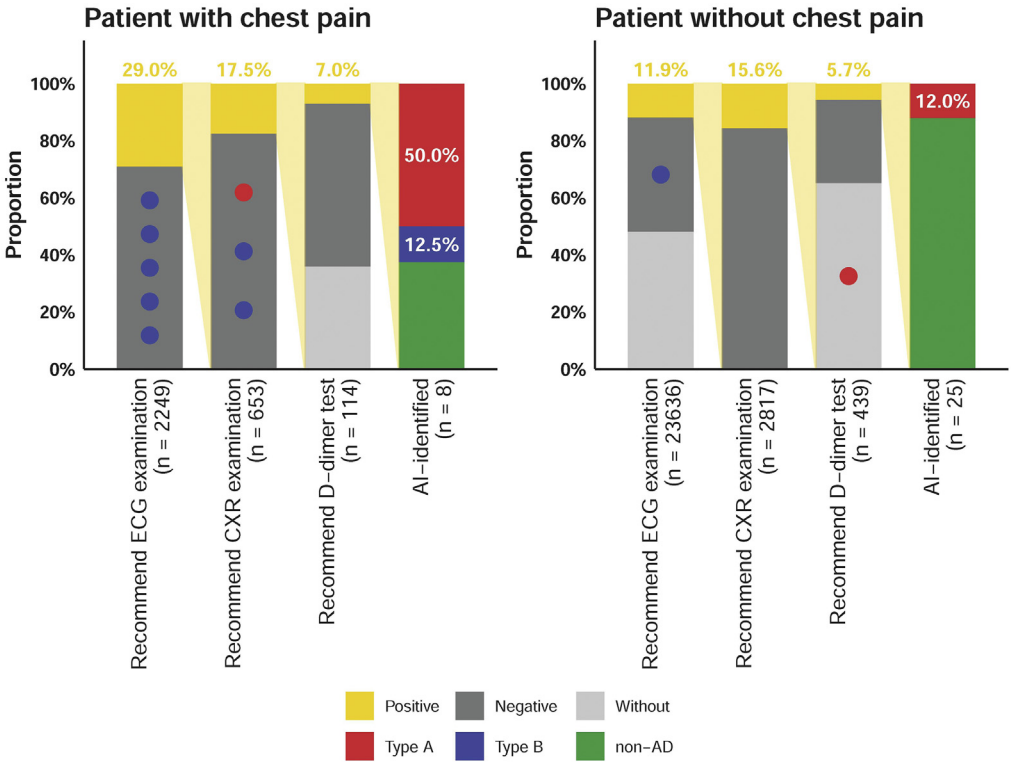

Figure 5. Prospective evaluation of artificial intelligence (AI) system. Results of the deep-learning model for artic dissection (AD) identification in patients with and without chest pain. Each point on the bar represents an AI-unidentified AD case. CXR, chest X-ray; ECG, electrocardiogram.

anticoagulant agents. In the presented ECG model, AD type A and AD type B seemed to be evaluated by the subtle morphologic changes of ST segments and T waves in nonspecific leads (Supplemental Figs. S8-S11). Interestingly, the presentation of ST-segment elevation in the ECG was considered a non-AD feature in the model, and the same ECG could still be correctly recognized as an AD by other predominant features (Supplemental Fig. S8). Although the mechanisms of these ECG features of AD are unclear, these findings may provide valuable information for future research.

Although the presence of a widened mediastinum provides diagnostic clues for AD, CXRs are of limited value when dissection is confined to the ascending aorta, which accounts for 10% of all AD type A.[10,21] Our CXR model was trained by more than 40,000 CXRs and reached a physician level of accuracy in AD identification. The CXR model had a sensitivity of 94.4% in recognition of AD type A, whereas the sensitivity of AD type B was only 50.0%. AD type A might affect both the ascending and descending aortas and thus present a larger area involved than AD type B on the CXR, indicating more significant evidence for the model to recognize AD. This result demonstrates the prominent application of the CXR model in the detection of AD type A.

Compared with previous studies using machine-learning models to screen for AD in inpatients,[22,23] our study focused on patients visiting the ED with high clinical relevance. Instead of using results from a series of blood tests in previous studies, we applied DLMs mainly on ECG and CXR to identify the features of AD, which provided much direct and critical information. In addition, we analyzed the laboratory indicators in our validation cohort and found that combining patient characteristics, especially D-dimer, with ECG and CXR enhanced the diagnostic power compared with using patient characteristics or ECG or CXR alone. Our integrated DLMs achieved AUCs of 0.910 and 0.963 in detecting all AD and AD type A, respectively, which are similar to previous studies with an AUC of 0.857 in detecting all AD.[22] Additional studies are necessary to evaluate the model performance of our integrated model in collaboration with the previous machine-learning data from inpatient blood sampling and patient characteristics.

Our AI model has numerous clinical applications. First, the use of chest CT increased rapidly despite steady patient volume and severity in previous studies, with a rate of 40 to 60 CT scans per 1000 ED visits.[24,25] The DLMs could identify AD among patients with chest pain with perfect sensitivity, thereby excluding those with low probability to avoid overuse of CT. Second, DLMs could be applied to distinguish AD involving coronary ostia from primary acute coronary diseases by ECG presentations, which is challenging in the ED. Third, by providing physician-level suggestions for shared decision making, the algorithm could assist in the diagnosis of AD, such in the case of a patient with chronic kidney disease requiring contrast-enhanced CT for definite diagnosis. Fourth, the algorithm could help physicians focus on selected patients, especially in nontertiary hospitals whose staff members have less familiarity with AD.[5] Finally, the ECG model could be used in ECG machines in ambulances to further shorten the time from onset of symptoms to definite diagnosis.

Limitations

Our study has several limitations. First, AI models were developed with patients recruited only from a tertiary hospital in Taiwan. Although the enrolled patients had similar characteristics to those of an international registry cohort,[3] and we applied several methods to reduce class imbalance and overfitting of DLMs, the generalization of DLMs should be evaluated carefully. The small sample size of patients with AD resulted in wide CIs in our results and restrained the actual performance of DLMs in clinical practice. Further large-scale prospective studies are warranted to confirm the performance of DLMs. Second, the lack of universal CT scanning is one of the limitations of our study, but the low incidence of AD in the general population and even in patients presenting with acute chest pain at the ED contributes to the extremely low false-negative rate. Third, the PPV of the algorithm was relatively low because of the low prevalence of AD. However, by using the model's high sensitivity, physicians could rapidly identify high-risk patients with AD and take appropriate action in time. Fourth, the performance of each ECG for AD recognition was not good enough, limiting its application in wearable devices. Finally, we could not completely understand how the algorithms make predictions, even with the help of visualization analysis. The black box in the DLMs might hinder the implementation of AI models.[26] To understand the AI decisions, we used class-activation mappings to display the focus of the algorithm on ECGs and CXRs (Supplemental Figs. S8-S12).[27,28] Future research on the decision-making process of the algorithm is required to increase the clinical practicability.

Conclusions

This is the first DLM using ECG and CXR features to detect AD, with an AUC of 0.960 for AD type A. With the assistance of DLMs, physicians could stratify patients with chest pain at risk of AD and reduce unnecessary examinations. Additional large-scale studies are warranted to confirm the performance of the DLMs.

Funding Sources

This study was supported by funding from the Ministry of Science and Technology, Taiwan (MOST 108-2314-B-016-001, MOST 109-2314-B-016-026 and MOST 110-2314-B-016-010-MY3 to C.L.), the National Science and Technology Development Fund Management Association, Taiwan (MOST 108-3111-Y-016-009, MOST 109-3111-Y-016-002 and MOST 110-3111-Y-016-005 to C.L.), and Cheng Hsin General Hospital, Taiwan (CHNDMC-109-19 and CHNDMC-110-15 to C.L.).

Disclosures

The authors have no conflicts of interest to disclose.

References

1. Bhuiya FA, Pitts SR, McCaig LF. Emergency department visits for chest pain and abdominal pain: United States, 1999-2008. NCHS Data Brief 2010:1-8.
2. Amsterdam EA, Kirk JD, Bluemke DA, et al. Testing of low-risk patients presenting to the emergency department with chest pain: a scientific statement from the American Heart Association. Circulation 2010;122:1756-76.
3. Evangelista A, Isselbacher EM, Bossone E, et al. Insights from the International Registry of Acute Aortic Dissection: a 20-year experience of collaborative clinical research. Circulation 2018;137:1846-60.
4. Salmasi MY, Al-Saadi N, Hartley P, et al. The risk of misdiagnosis in acute thoracic aortic dissection: a review of current guidelines. Heart 2020;106:885-91.
5. Harris KM, Strauss CE, Eagle KA, et al. Correlates of delayed recognition and treatment of acute type A aortic dissection: the International Registry of Acute Aortic Dissection (IRAD). Circulation 2011;124:1911-8.
6. Zhan S, Hong S, Shan-Shan L, et al. Misdiagnosis of aortic dissection: experience of 361 patients. J Clin Hypertens (Greenwich) 2012;14:256-60.
7. Hansen MS, Nogareda GJ, Hutchison SJ. Frequency of and inappropriate treatment of misdiagnosis of acute aortic dissection. Am J Cardiol 2007;99:852-6.
8. Erbel R, Aboyans V, Boileau C, et al. 2014 ESC Guidelines on the diagnosis and treatment of aortic diseases: document covering acute and chronic aortic diseases of the thoracic and abdominal aorta of the adult. The Task Force for the Diagnosis and Treatment of Aortic Diseases of the European Society of Cardiology (ESC). Eur Heart J 2014;35:2873-926.
9. Costin NI, Korach A, Loor G, et al. Patients with type a acute aortic dissection presenting with an abnormal electrocardiogram. Ann Thorac Surg 2018;105:92-9.
10. von Kodolitsch Y, Nienaber CA, Dieckmann C, et al. Chest radiography for the diagnosis of acute aortic syndrome. Am J Med 2004;116:73-7.
11. Funakoshi H, Mizobe M, Homma Y, Nakashima Y, Takahashi J, Shiga T. The diagnostic accuracy of the mediastinal width on supine anteroposterior chest radiographs with nontraumatic Stanford type A acute aortic dissection. J Gen Fam Med 2018;19:45-9.
12. Esteva A, Robicquet A, Ramsundar B, et al. A guide to deep learning in healthcare. Nat Med 2019;25:24-9.
13. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521:436-44.
14. Attia ZI, Friedman PA, Noseworthy PA, et al. Age and sex estimation using artificial intelligence from standard 12-lead ECGs. Circ Arrhythm Electrophysiol 2019;12:e007284.
15. Attia ZI, Noseworthy PA, Lopez-Jimenez F, et al. An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. Lancet 2019;394:861-7.
16. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts H. Artificial intelligence in radiology. Nat Rev Cancer 2018;18:500-10.
17. Lin CS, Lin C, Fang WH, et al. A deep-learning algorithm (ECG12Net) for detecting hypokalemia and hyperkalemia by electrocardiography: algorithm development. JMIR Med Inform 2020;8:e15931.
18. Rajpurkar P, Irvin J, Ball RL, et al. Deep learning for chest radiograph diagnosis: a retrospective comparison of the CheXNeXt algorithm to practicing radiologists. PLoS Med 2018;15:e1002686.
19. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017;2261-2269.

20. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Proceedings of the 25th International Conference on Neural Information Processing Systems-Vol 1. Lake Tahoe, Nevada: Curran Associates Inc., 2012:1097-105.

21. Kohl LP, Isselbacher EM, Majahalme N, et al. Comparison of outcomes in DeBakey type AI versus AII aortic dissection. Am J Cardiol 2018;122:689-95.

22. Huo D, Kou B, Zhou Z, Lv M. A machine learning model to classify aortic dissection patients in the early diagnosis phase. Sci Rep 2019;9:2701.

23. Liu L, Zhang C, Zhang G, et al. A study of aortic dissection screening method based on multiple machine learning models. J Thorac Dis 2020;12:605-14.

24. Lee J, Kirschner J, Pawa S, Wiener DE, Newman DH, Shah K. Computed tomography use in the adult emergency department of an academic urban hospital from 2001 to 2007. Ann Emerg Med 2010;56:591-6.

25. Broder J, Warshauer DM. Increasing utilization of computed tomography in the adult emergency department, 2000-2005. Emerg Radiol 2006;13:25-30.

26. Castelvecchi D. Can we open the black box of AI? Nature 2016;538:20-3.

27. Yang Z, Yang D, Dyer C, He X, Smola A, Hovy E. Hierarchical attention networks for document classification. Proceedings of the 2016 conference of the North American chapter of the Association for Computational Linguistics: Human Language Technologies: Association for Computational Linguistics. 2016;1480-1489.

28. Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A. Learning deep features for discriminative localization. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, Nevada, USA: IEEE, 2016.

Supplementary Material

To access the supplementary material accompanying this article, visit the online version of the *Canadian Journal of Cardiology* at www.onlinecjc.ca and at https://doi.org/10.1016/j.cjca.2021.09.028.