scientific reports

Check for updates

OPEN

Deep learning algorithm for predicting rapid progression of abdominal aortic aneurysm by integrating CT images and clinical features

Se Jin Oh[1,2,11,12]✉, Jae-ik Shin[3,4,11], Eun Na Kim[5,11], Ariani Widiastini[6], Yiyu Hong[2], Insuk Sohn[2], Kwang Nam Jin[7], Joon Seo Lim[8], Ji Seong Kim[9], Hong-Jae Choi[1], You Jung Ok[1], Jae-Sung Choi[1] & Jae Woong Choi[10,12]✉

Abdominal aortic aneurysm (AAA) progression carries a significant rupture risk, demanding accurate prediction models beyond traditional methods that rely on limited clinical parameters and often overlook complex factor interplay. We aimed to enhance prediction by developing and validating an end-to-end multi-modal deep learning (DL) model that integrates features extracted using ResNet from computed tomography (CT) images, geometric features derived from radiomics based on CT annotations, and clinical features obtained from clinical records. This retrospective study utilized data from 561 AAA patients sourced from Boramae Medical Center and Seoul National University Hospital, including 14,252 annotated CT axial images alongside detailed clinical information. Patients were categorized into rapid or slow progression groups based on an annual growth rate threshold of 2.5 mm/year. The multi-modal DL model that incorporated CT images, clinical features, and geometric features demonstrated superior predictive performance for rapid progression, achieving an area under the receiver operating characteristic curve (AUC) of 0.807 and an accuracy of 0.758. This significantly outperformed traditional machine learning models utilizing only clinical data (AUC: 0.716) or only geometric features (AUC: 0.715). The improvement in AUC was statistically significant according to DeLong's test. This study underscores the value of AI-driven, multi-modal approaches for enhancing patient-specific AAA risk stratification, potentially enabling more precise monitoring and optimized timing for clinical interventions.

Keywords Abdominal aortic aneurysm, Deep learning, Multi-modal model, CT imaging, Digital health

Abdominal aortic aneurysm (AAA) is a life-threatening degenerative vascular disease characterized by the pathological dilation of the aortic wall. AAA is commonly defined as an increase in aortic diameter of more than 50% of its normal size or a measurement exceeding 30 mm[1]. If left untreated, AAA continues to expand and may

[1]Department of Thoracic and Cardiovascular Surgery, SMG-SNU Boramae Medical Center, Seoul National University College of Medicine, 20, Boramae-ro 5-gil, Dongjak-gu, Seoul 07061, Republic of Korea. [2]Healthcare AI Research Institute, Seoul National University Hospital, Seoul, Republic of Korea. [3]Arontier Co., Ltd, Seoul, Republic of Korea. [4]Department of Radiation Oncology, Yonsei University College of Medicine, Seoul, Republic of Korea. [5]Department of Pathology, Seoul National University Hospital, Seoul National University College of Medicine, Seoul, Republic of Korea. [6]Seoul National University College of Medicine, Seoul, Republic of Korea. [7]Department of Radiology, SMG-SNU Boramae Medical Center, Seoul National University College of Medicine, Seoul, Republic of Korea. [8]Clinical Research Center, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea. [9]Department of Thoracic and Cardiovascular Surgery, Chungnam National University Hospital, Chungnam National University College of Medicine, Daejeon, Republic of Korea. [10]Department of Thoracic and Cardiovascular Surgery, Seoul National University Hospital, Seoul National University College of Medicine, 101, Daehak-ro, Jongno-gu, Seoul 03080, Republic of Korea. [11]Jae-ik Shin, Jae-ik Shin and Eun Na Kim contributed equally to this work and share the first authorship. [12]Se Jin Oh and Jae Woong Choi contributed equally to this work and share the corresponding authorship. ✉email: wpwnn@snu.ac.kr; cjw01@snu.ac.kr

eventually rupture, significantly contributing to cardiovascular mortality; accordingly, AAA ranks among the top ten causes of death worldwide[2,3]. The incidence of AAA is particularly high in men over 60 years of age, with a prevalence ranging from 4% to 8% in this population[4]. As global life expectancy increases, the burden of AAA is also rising, with approximately 200,000 new cases diagnosed and 40,000 surgical interventions performed annually in the United States alone[5].

Despite its critical significance, the precise pathophysiological mechanisms underlying AAA progression remain poorly understood, and no pharmacological therapy has been proven to halt or slow disease progression[6]. Consequently, AAA management predominantly relies on periodic imaging surveillance to monitor aneurysm growth, with surgical intervention recommended only when the aneurysm reaches a critical size threshold[7]. However, AAA progression is highly variable, with some aneurysms exhibiting rapid expansion and others remaining stable over prolonged periods[8–11]. This unpredictability complicates clinical decision-making, particularly regarding the optimal timing for surgical intervention, underscoring the need for more refined, data-driven approaches to AAA monitoring and treatment.

Traditional methods for predicting AAA progression have primarily focused on biomechanical modeling and imaging-based assessments of wall stress[12,13], as well as physiological measurements[14]. However, these approaches are often constrained by technical limitations such as discrepancies in image resolution, estimation errors in wall stress, and challenges in standardizing measurements across diverse patient populations. Recent advancements in artificial intelligence (AI) and deep learning (DL) have introduced innovative methodologies for analyzing complex imaging data and extracting predictive features that may not be apparent through conventional assessments. Notably, DL algorithms applied to three-dimensional (3D) computed tomography (CT) angiography have demonstrated superior performance in predicting rupture risk in intracranial aneurysms, highlighting the potential of AI-based models in vascular disease prognosis[15,16]. In the context of AAA research, DL applications have primarily been utilized for aneurysm segmentation and screening[17], with relatively fewer studies focusing on predictive modeling of disease progression[18,19]. Given the multifactorial nature of AAA progression—driven by a combination of proteolytic activity, oxidative stress, and chronic inflammation—integrating multimodal data, including clinical features and CT imaging, into a comprehensive DL model may enhance predictive accuracy[20,21].

In this study, we introduce a deep learning model designed to predict the rapid progression of AAA by integrating CT imaging data with clinical features. By leveraging structural information from imaging and contextual insights from clinical features, our approach aims to develop a robust, data-driven framework for individualized risk assessment. Using a large cohort of patients with longitudinal imaging and clinical follow-up data, we seek to contribute to the advancement of AI-based surveillance strategies for AAA. Ultimately, this predictive model has the potential to facilitate more accurate risk stratification, optimize the timing of interventions, and reduce the incidence of AAA rupture, thereby improving patient outcomes.

Materials and methods

Study populations

The study was approved by the Institutional Review Board (IRB) of SMG-SNU Boramae Medical Center (BMC; IRB No. 20-2022-103) and Seoul National University Hospital (SNUH; IRB No. 2309-027-1463). Informed consent was waived by the Institutional Review Boards of both SMG-SNU Boramae Medical Center and Seoul National University Hospital, as the study was retrospective in nature and all clinical and imaging data were anonymized prior to analysis.

All methods were performed in accordance with the relevant guidelines and regulations, including the Declaration of Helsinki. Patients diagnosed with AAA through contrast-enhanced CT scans at these institutions between January 2010 and May 2023 were included based on the following criteria: (a) AAA diameter of ≥ 28 mm and (b) at least one follow-up CT scan obtained at an interval of ≥ 4 months. The exclusion criteria were as follows: (a) initial AAA diameter of ≥ 48 mm, (b) suprarenal or juxtarenal AAA, (c) ruptured aneurysm, (d) dissecting aneurysm, and (e) suspected mycotic aneurysm.

Classification of rapid and slow progression groups

The cohort was divided into two groups: rapid progression and slow progression, based on the growth rate of AAA. The growth rate was defined as the mean expansion rate of the aorta over the entire follow-up period and was calculated using the following formula:

$$Growth\ Rate\ (mm/year) = \frac{D_{final} - D_{initial}\ (mm)}{\text{period time}\ (\text{years})}$$

According to the National Health Service AAA Screening Program in the UK, the rapid progression group was defined by a mean expansion rate of 1.9 mm/year for initial sizes ranging from 2.8 to 3.9 mm, 2.7 mm/year for sizes between 3.0 and 4.5 mm, and 3.5 mm/year for sizes between 4.6 and 8.5 mm[22,23]. Since our cohort included patients with initial sizes ranging from 2.7 to 4.7 mm, a cut-off growth rate of 2.5 mm/year was used to classify patients into rapid vs. slow progression groups[24].

Finally, the BMC dataset and the SNUH dataset comprised features extracted from different modalities, such as clinical features and CT images. The rapid progression group was labeled as positive (label = 1), while the slow progression group was labeled as negative (label = 0).

Overview of multi-modal model construction

An overview of the construction of the multi-modal model is illustrated in Fig. 1. We developed this model to predict the rapid progression of AAA by extracting features from three distinct modalities across both datasets.

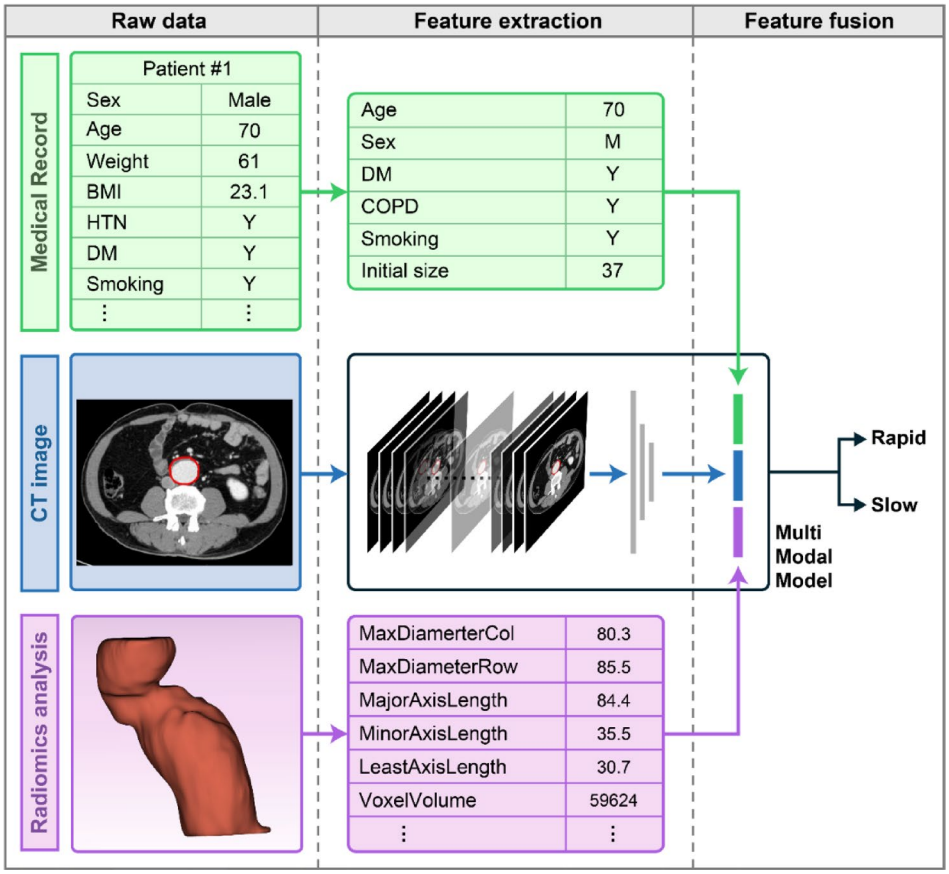

Fig. 1. An overview of the multi-modal model construction. A deep learning model was constructed with data obtained from CT images, medical records, and radiomics analysis. All features were integrated into the model to predict the rapid progression of AAA.

The first modality encompasses significant clinical features that contribute to AAA progression, selected from various clinically relevant risk factors derived from patients' medical records. The second modality employs a deep learning model that extracts features from post-processed CT images centered on annotated regions of the abdominal aorta. The third modality focuses on extracting geometric features from the 3D structure of the annotated abdominal aorta. By analyzing and integrating the characteristics of data from each modality, we aim to enhance the accuracy of AAA progression predictions.

Clinical variable extraction

The selection of clinical variables was based on established evidence from major guidelines and cohort studies[25–28], as well as data-driven analysis within our cohort. In total, we screened approximately 20 clinical variables and retained only those demonstrating consistent significance in our analysis or strong support in prior literature. We extracted the following clinical variables from medical records: chronic obstructive pulmonary disease (COPD), diabetes mellitus (DM), smoking status, sex, age, and initial size of AAA. COPD and DM were binarized, with the presence of these conditions labeled as 1 (positive). For sex, males were labeled as 1 and females as 2. Age and initial size were treated as continuous variables. The initial size refers to the diameter of the AAA as measured on the initial CT scan. Smoking status was categorized into three groups: never, current, and past (within the last 30 days). Individual clinical features were normalized using the Z-score to align with the BMC data[10,29].

CT image data acquisition

CT images were obtained using a 64-slice CT scanner (Brilliance 64, Philips Healthcare, Amsterdam, Netherlands), a 128-slice CT scanner (Ingenuity, Philips Healthcare) at BMC, and a dual-source CT scanner (SOMATOM Definition, Siemens Healthineers, Erlangen, Germany) or a 256-slice CT scanner (iCT, Philips Healthcare). The protocol for performing CT scans involved contrast-enhanced imaging in the craniocaudal direction. The scan parameters were as follows: (a) collimation of 64×0.625 mm, 64×0.6 mm, 32×0.6 mm, or 128×0.625 mm; (b) tube voltage of 100 kVp or 120 kVp; (c) tube current of 120 mAs, 200 mAs, or a range of 104 to 620 mA; and (d) rotation speed of 270 to 500 milliseconds. The tube voltage and current settings were adjusted based on the patient's body habitus. Transverse datasets were reconstructed with slice thicknesses ranging from 2.0 mm to 10.0 mm. The resultant CT axial images from BMC exhibited a mean pixel spacing of 0.68 mm and

a mean slice thickness of 3.4 mm. In contrast, CT images from SNUH demonstrated a mean pixel spacing of 0.69 mm and a mean slice thickness of 3.4 mm. Both datasets had an image matrix of 512 pixels.

Segmentation of the abdominal aorta

We used the 3D Slicer program[30] to select the corresponding axial CT images that included the abdominal aorta. Figure 2 illustrates an example of the annotation process. The craniocaudal extent of the abdominal aorta was defined as extending from the renal arteries to just before the iliac bifurcation. The outer margin of the abdominal aorta on the CT images was annotated manually and saved as binary masks. The CT window level and width were adjusted to 40 Hounsfield Units (HU) and 350 HU, respectively, to assist the radiologist in distinguishing the contrast-enhanced aorta from other soft tissues.

Geometric feature extraction from radiomics

Geometric features were extracted from CT images annotated with the aorta using the PyRadiomics package[31]. The 3D structure of the aorta was constructed by integrating all annotations along with the pixel spacing and slice thickness of the CT images. Fourteen 3D shape-based features were derived from this structure, including Elongation, Flatness, Least Axis Length, Major Axis Length, Maximum 2D Diameter (column), Maximum 2D Diameter (row), Maximum 2D Diameter (slice), Maximum 3D Diameter, Mesh Volume, Minor Axis Length, Sphericity, Surface Area, Surface Volume Ratio, and Voxel Volume. Although both originate from the same imaging source, CT image features capture pixel-level intensity patterns, while geometric features reflect the patient's anatomical characteristics derived from expert-defined annotations. Therefore, they represent distinct modalities that provide complementary and non-redundant information for the prediction task. The details of the feature definitions are presented in Table S1. The individual geometric features were normalized using Z-score normalization for the BMC dataset, and the same scaling was applied to the SNUH data.

Machine learning model using clinical and geometric features

To compare DL models utilizing a combination of features described in the following section, machine learning (ML)-based classification models were developed using the XGBoost package[13,32]. These models were based on the extreme gradient boosting algorithm, and hyperparameters were optimized to maximize the area under the curve (AUC) using the Optuna package[14,33]. Two models were trained: one incorporating clinical features and

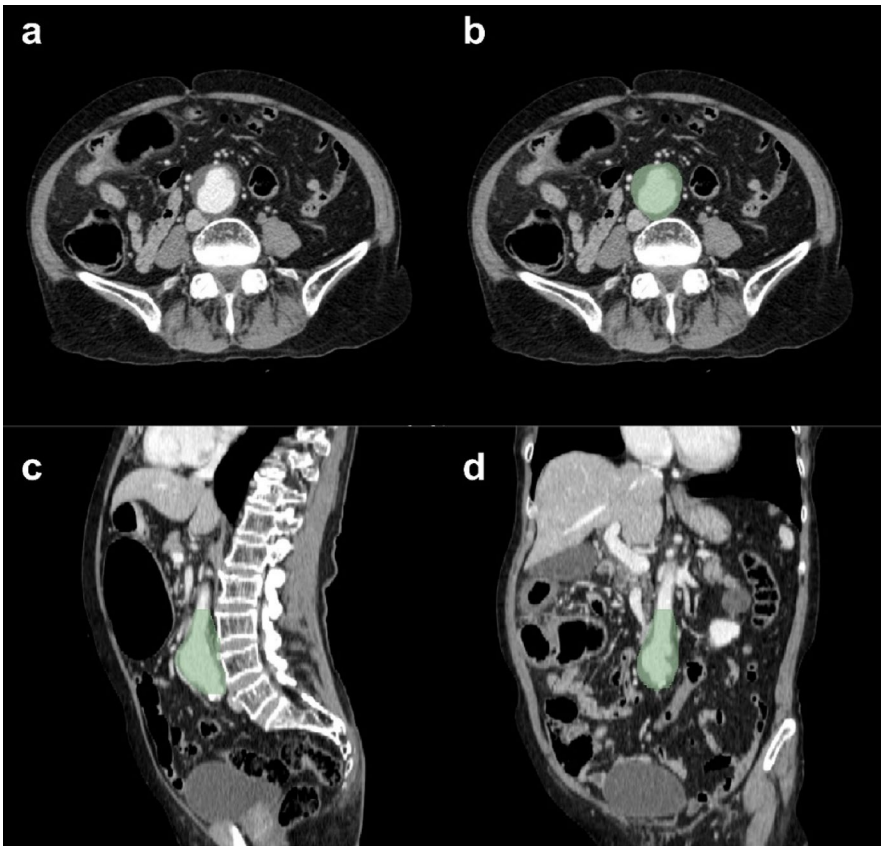

Fig. 2. AAA annotation on CT image using 3D slicer program. (**a**) CT window level and window width applied to the CT image in the axial view. (**b**) Aorta annotation in the axial view. (**c**) Aorta annotation before the renal arteries, verified in the sagittal view. (**d**) Aorta annotation from the aortic bifurcation, verified in the sagittal view.

the other utilizing geometric features from the SNUH dataset. Model evaluation was conducted using the BMC dataset.

Preprocessing of CT images for DL models

To isolate the abdominal aorta in the CT images, pixels outside the annotated region were replaced with a blank value (Air, -1000 HU). Axial CT slices covering the entire volume of the abdominal aorta -from the aortic bifurcation to the renal arteries- were used as input. Each slice was cropped to a matrix size of 128 × 128 pixels centered on the annotation. We confirmed that the entire annotated aortic region was consistently contained within this window across all patients. This preprocessing step focused on preserving local texture patterns, as spatial resolution (including pixel spacing and slice thickness) had already been accounted for in the separately extracted geometric features. The same CT window level and window width used for annotation were applied. Subsequently, the images were resized to 224 × 224 pixels to match the input requirements of the DL model (Fig. 3).

Construction of DL models using CT images

The DL framework PyTorch[15,34] was utilized to construct a deep learning model, as illustrated in Fig. 1. A modified ResNet18 convolutional neural network (CNN)[16,35] was used to extract features from 2D axial CT images containing annotations. The input layer of the model was adapted for grayscale images instead of RGB color images, as in the original ResNet18. The convolutional layers of the network extracted a feature vector of 512 dimensions from each CT image. Three fully connected layers were added to compress these 512 features into 8 features. Finally, the output layer generated the probability of rapid progression of AAA after applying a sigmoid activation function. The Binary Cross-Entropy Loss function was used to minimize the difference between the predicted probabilities and the patient group labels (Rapid = 1.0, Slow = 0.0). We applied class weights to the loss function, with weights proportional to the inverse label frequency in the training data. The model was optimized using the AdamW optimizer with a learning rate of 0.0001[17,36].

The model was trained using CT images from the SNUH dataset and tested on the BMC dataset. To enhance the model's robustness against inter-patient anatomical variability and variations in CT acquisition conditions, we applied image augmentation techniques during training. These included Gaussian blur, median blur, color jitter, horizontal flip, vertical flip, random 90-degree rotations, and random shift-scale rotations, implemented using Albumentations[18,37], which are commonly used in deep learning–based medical image analysis to improve generalizability and prevent overfitting[18,37].

Model evaluation was conducted on a per-patient basis. To obtain a patient-level prediction, we applied a soft voting strategy[38] by averaging the rapid progression probabilities predicted from multiple CT slices for each patient. The aggregated probability was then compared with the patient's label for model evaluation. The training and testing processes of the deep learning model are illustrated in Fig. 1.

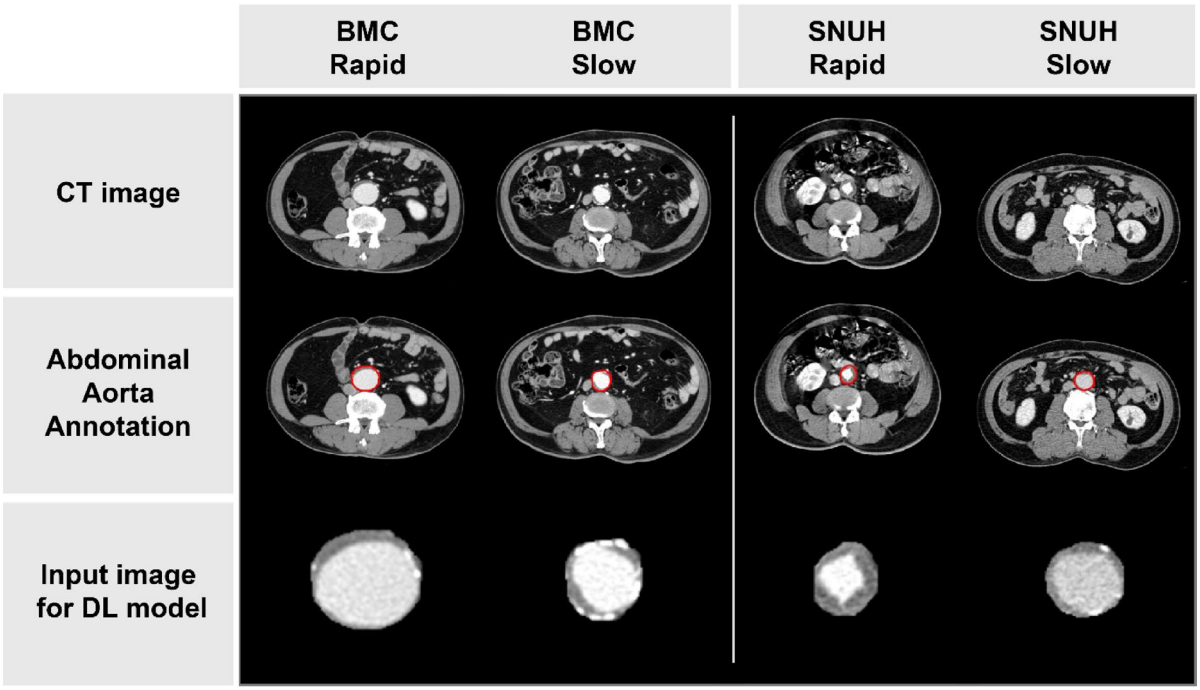

Fig. 3. CT image preprocessing for the deep learning model. The first row consists of CT angiography images obtained from a CT scanner. The second row displays the aorta annotation on the CT image, which is outlined in red. The CT image within the annotated region has been cropped and resized for use as the input image for the deep learning model.

DL-based multi-modal models

We developed a novel DL model by integrating data obtained from CT images, medical records, and radiomics analysis. The multi-modal model was built upon an existing model for CT images. Specifically, the input to the output layer consisted of 8 CT image features, 6 clinical features, and 14 geometric features corresponding to the CT images. The concatenation of these multi-modal features in the output layer is illustrated in Fig. 1. Consequently, a total of 28 features were used to determine the probability of rapid progression of AAA in the output layer. Three models were trained: one utilizing CT images and clinical features, another using CT images and geometric features, and a third combining CT images, clinical features, and geometric features. In training our multimodal DL model, we applied the same image augmentation techniques described in the previous section to the CT image modality. However, no augmentation was applied to the clinical or geometrical features. This approach follows the conventional paradigm in multimodal learning, where augmentation is primarily applied to image-based modalities while preserving the integrity of structured tabular inputs, such as clinical data and derived geometrical measurements[72,73]. The testing procedure was consistent with that employed for the CT image-based model discussed in the previous section.

Statistical analysis

To compare the BMC and SNUH cohorts, p-values were calculated for each feature individually using Pearson's chi-squared test, Welch's t-test, and the Wilcoxon rank-sum test[19,39–41], as implemented in the Statsmodels package[22,42]. These tests were applied to the features based on their distribution[22,43]. A p-value of less than 0.05 was considered to indicate a statistically significant difference in the distribution of the datasets.

We compared two ML-based models that utilized features from two modalities with four DL-based models that incorporated features from three modalities. The performance of each model was evaluated individually using accuracy, the receiver operating characteristic (ROC) curve, and the area under the receiver operating characteristic curve (AUC)[29,44] employing the Scikit-learn package[30,45]. An AUC value closer to 1.0 indicates superior model performance. DeLong's test[46] was utilized to assess the improvement resulting from the addition of extra modalities, using the pROC R package[47]. A p-value of less than 0.05 indicates moderate evidence of a statistically significant difference between the AUC values.

Results

Cohort dataset and clinical characteristics

A total of 561 patients diagnosed with AAA were included in this study, consisting of 236 patients from BMC and 325 patients from SNUH. Both CT imaging data and corresponding medical records were analyzed. The dataset comprised 14,252 annotated CT images, with 9,077 images from the SNUH cohort utilized for training and 5,175 images from the BMC cohort employed for testing. Within the BMC dataset, the rapid progression group included 81 patients (1,732 CT images), while the slow progression group consisted of 155 patients (3,443 CT images). In the SNUH dataset, 114 patients (3,306 CT images) were classified as rapid progressors, and 211 patients (5,771 CT images) were categorized as slow progressors. The follow-up period was defined as the interval between the initial and final CT scans. The median follow-up duration was 4.4 years in the SNUH cohort (range: 0.6–16.3) and 3.8 years in the BMC cohort (range: 0.3–16.0).

The proportion of male patients was significantly lower in the BMC cohort compared to the SNUH cohort (74.5% vs. 86.5%; $p < 0.001$). Patients in the BMC cohort were generally older than those in the SNUH cohort (73.1 ± 8.9 years vs. 68.2 ± 8.1 years ; $p < 0.001$). The proportions of COPD (17.8% vs. 24.9%; $p = 0.056$) and DM (21.6% vs. 27.4%; $p = 0.144$) were lower in the BMC cohort than in the SNUH cohort, albeit without statistical significance. Smoking history was significantly different between the two cohorts ($p < 0.001$), with the proportion of current smokers being higher in the BMC cohort (23.7% vs. 17.8%) and that of never-smokers being lower in the SNUH cohort (45.3% vs. 48.9%). The initial AAA size of the BMC cohort was smaller than that of the SNUH cohort, albeit without statistical significance (34.4 ± 5.8 mm vs. 35.4 ± 5.8 mm; $p = 0.072$). Details of the cohorts and clinical characteristics are presented in Tables 1 and S2.

Geometric feature analysis

The characteristics of the geometric features for both institutions are summarized in Table 2. The SNUH dataset exhibited higher mean values and greater variability across nearly all geometric features compared to the BMC dataset. Among these features, the Maximum 2D Diameter Slice was found to be correlated with aortic length in the superior-to-inferior direction. The aortic length was determined by multiplying the number of CT slices by the slice thickness, and the results were comparable between the two datasets, as annotations consistently extended from the aortic bifurcation to the renal arteries. Additionally, Minor Axis Length and Least Axis Length were associated with ellipsoid diameter along the posterior-to-anterior and proximal-to-distal directions, respectively. These geometric features reflect the anatomical orientation and positioning of AAA within the retroperitoneal space, which influences measurements along the superior-inferior and posterior-anterior axes.

Comparison of the distributions of clinical and geometric features between datasets

The SNUH dataset included a larger number of patients than the BMC dataset. Consequently, model training was conducted using the SNUH dataset, while model testing was performed with the BMC dataset. To evaluate the performance of the single-center cohort model, it is essential to compare it with a cohort from another center. When two datasets differ significantly, the model will require greater generalization to effectively test data from another center's cohort. The differences between the two datasets were assessed through statistical analysis. To analyze the statistical differences, a Pearson's chi-squared test was employed for categorical clinical features, while a Wilcoxon rank-sum test was used for continuous clinical features. Welch's t-test was applied to the "Sphericity" and "SurfaceVolumeRatio" geometric features, as these were confirmed to be normally

Variables	BMC, *n* (%) (*n* = 236)	SNUH, *n* (%) (*n* = 325)	*P* value
Number of images	5175	9077	
Sex, male	176 (74.5)	281 (86.5)	< 0.001*
Age, years (mean ± SD)	73.1 ± 8.85	68.2 ± 8.06	< 0.001**
COPD	42 (17.8)	81 (24.9)	0.056*
DM	51 (21.6)	89 (27.4)	0.144*
Smoking status			< 0.001*
Never	107 (45.3)	159 (48.9)	
Current	56 (23.7)	58 (17.8)	
Past	73 (30.9)	108 (33.2)	
Initial size of AAA, mm (mean ± SD)	34.4 ± 5.83	35.4 ± 5.75	0.072**

Table 1. Comparison of demographic and clinical variables. AAA: Abdominal aortic aneurysm; BMC: Boramae Medical Center; COPD: chronic obstructive pulmonary disease; DM: diabetes mellitus; SD: standard deviation; SNUH: Seoul National University Hospital. * Pearson's chi-squared test between the BMC and SNUH datasets. ** Wilcoxon rank-sum test between the BMC and SNUH datasets.

Features	BMC	SNUH	*P* value
Maximum 2D diameter (Column)	80.4 ± 15.9	98.3 ± 27.9	< 0.001*
Maximum 2D diameter (Row)	85.5 ± 15.5	99.1 ± 26.4	< 0.001*
Maximum 2D diameter (Slice)	42.7 ± 9.8	42.3 ± 9.6	0.690*
Maximum 3D diameter	89.4 ± 15.4	104.2 ± 26.9	< 0.001*
Major axis length	84.4 ± 17.7	102.3 ± 31.6	< 0.001*
Minor axis length	35.5 ± 7.2	36 ± 6.5	0.311*
Least axis length	30.7 ± 5.7	31.2 ± 4.9	0.216*
Elongation	0.44 ± 0.12	0.38 ± 0.13	< 0.001*
Flatness	0.38 ± 0.11	0.33 ± 0.11	< 0.001*
Sphericity	0.7 ± 0.05	0.69 ± 0.05	< 0.001**
Surface volume ratio	0.18 ± 0.03	0.18 ± 0.02	0.008**
Voxel volume	59,624 ± 27,891	69,245 ± 26,999	< 0.001*
Mesh volume	59,527 ± 27,871	69,142 ± 26,987	< 0.001*
Surface area	10,306 ± 3131	11,784 ± 3277	< 0.001*

Table 2. Comparison of geometric features. Values are mean ± standard deviation. BMC: Boramae Medical Center; SNUH: Seoul National University Hospital. * Wilcoxon rank-sum test between BMC and SNUH. ** Welch's t-test between BMC and SNUH.

distributed by the Shapiro-Wilk test[33,48]. The Wilcoxon rank-sum test was also utilized for the examination of other geometric features.

There were no significant differences between the two datasets regarding COPD, DM, initial AAA size, Maximum 2D Diameter Slice, Minor Axis Length, and Least Axis Length. In contrast, other features were significantly different between the two datasets. The observed differences in geometric features can be attributed to the distribution of initial sizes within the clinical features.

Verification of cut-off values in patient data distribution

Prior to applying the cut-off value of 2.5 mm/year established in previous studies[22,23], we evaluated our dataset using a range of cut-off values from 2.0 to 3.0 mm/year to assess their impact on model performance. Adjusting the cut-off values resulted in changes in group assignments for a subset of patients, who were reclassified from the slow progression group to the rapid progression group in both the BMC and SNUH datasets. Based on these reclassified datasets, two types of machine learning models were independently trained and tested: one utilizing clinical features and the other utilizing geometric features. The resulting variations in ROC AUC and accuracy, highlighting the sensitivity of model performance to the selected cut-off, are comprehensively illustrated in Fig. 4.

Among the cut-off values, 2.5 and 2.7 mm/year had ROC AUC values exceeding 0.7 in both data models, and we chose the smallest value (2.5 mm/year) as the cut-off value for this study. Furthermore, the appropriateness of the 2.5 mm/year cut-off was confirmed in this study.

Prediction of rapid AAA progression using a multi-modal DL model

The ROC curve and AUC for testing all models on the BMC dataset are shown in Fig. 5. The baseline DL-based model, which used CT images, was trained on the SNUH dataset and tested on the BMC dataset, achieving an

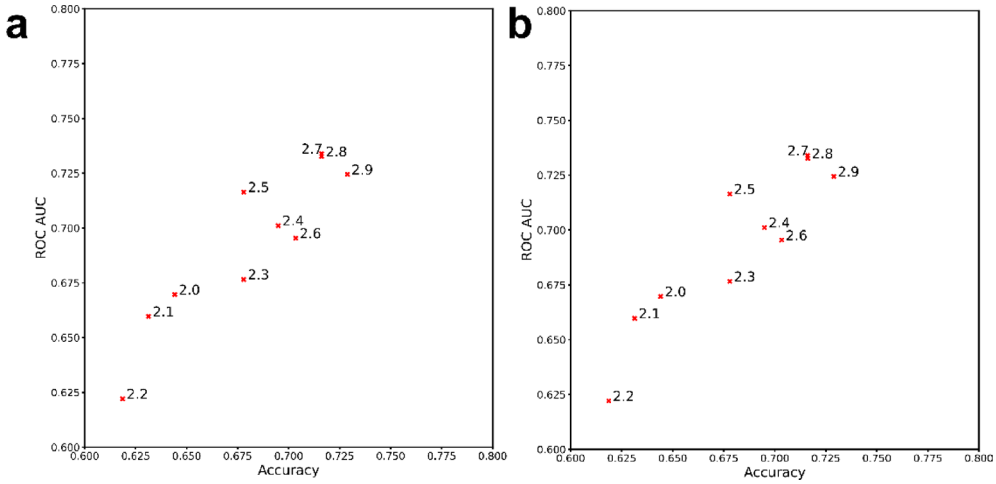

Fig. 4. Accuracy and ROC AUC at various cut-off values in the tests of (**a**) the clinical feature-based ML model and (**b**) the geometric feature-based ML model. The cut-off values of 2.5 and 2.7 mm/year demonstrated better accuracy and ROC AUC compared to other values in both models.

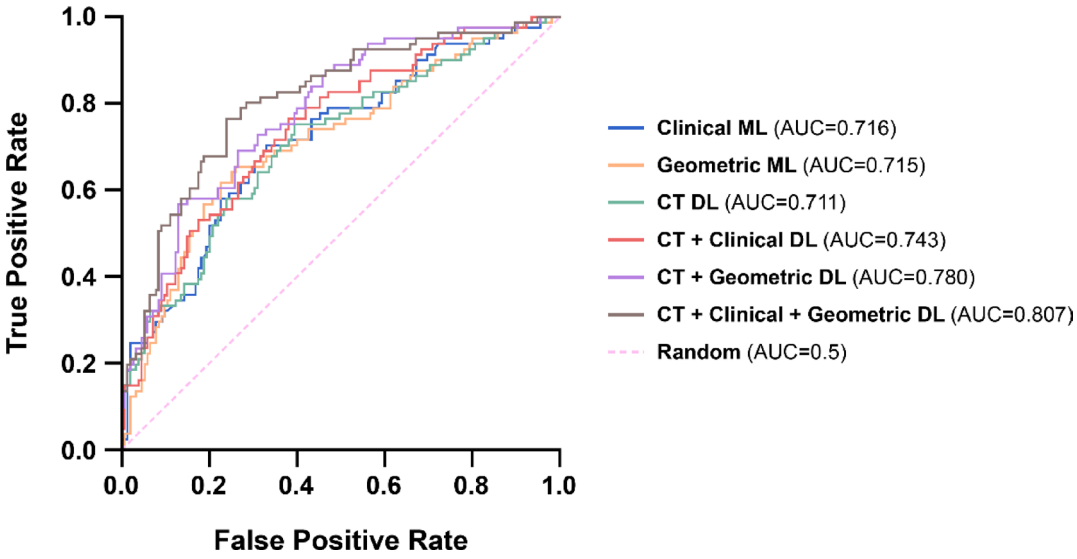

Fig. 5. ROC curves and AUC results. A comparative analysis was conducted between ML-based models employing clinical or geometric features and DL-based models that integrated combinations of CT image features with clinical or geometric features. The DL-based model that incorporated all features achieved the highest ROC AUC value of 0.807.

AUC of 0.711. This performance was comparable to that of ML-based models using either clinical features or geometric features, with AUCs of 0.716 and 0.715, respectively. DL models that incorporated clinical features or geometric features showed improved performance, with AUCs of 0.743 and 0.780, respectively, compared to the baseline DL model. The addition of features significantly enhanced performance when compared to the ML-based models using clinical or geometric features (AUCs of 0.716 and 0.715). Our multi-modal DL model, which utilized all features, achieved the highest performance among all models, with an AUC of 0.807. Additionally, the accuracy of the models is detailed in Table 3. The trend in accuracy among the models was similar to that of the AUC. The corresponding confusion matrices for each model are presented in Supplementary Table S3 to further illustrate the classification performance.

The results of DeLong's test, which assessed the statistical significance between our best model and other models, are presented in Table 4. The multi-modal DL model, which incorporated CT images, clinical features, and geometric features, demonstrated a significant improvement in the AUC compared to the DL model that utilized CT images only ($p = 0.001$), the DL model that utilized CT images and clinical features ($p = 0.021$), the ML model that used clinical features only, and the ML model that utilized geometric features only ($p = 0.007$, $p < 0.001$). In contrast, there were no significant differences between the multi-modal DL model and the DL model that used CT images and geometric features.

Model Metrics	Clinical ML	Geo ML	Clinical + Geo ML	CT DL	CT + Clinical DL	CT + Geo DL	CT + Clinical + Geo DL
AUC	0.716	0.715	0.723	0.711	0.743	0.780	0.807
Accuracy	0.678	0.708	0.674	0.653	0.665	0.763	0.758

Table 3. Performance evaluation of the ML and DL models. AUC: area under the receiver operating characteristic curve; Clinical: Clinical feature; CT: computed tomography; DL: deep learning; Geo: Geometric feature; ML: machine learning.

Model 1	AUC 1	Model 2	AUC 2	*p*-value
CT + Clinical + Geo, DL	0.807	Clinical, ML	0.716	0.007
		Geo, ML	0.715	< 0.001
		CT, DL	0.711	0.001
		CT + Clinical, DL	0.743	0.021
		CT + Geo, DL	0.780	0.08

Table 4. Comparison of AUC values across different models. Abbreviations: AUC: area under the receiver operating characteristic curve; Clinical: Clinical feature; CT: computed tomography; DL: deep learning; Geo: Geometric feature; ML, machine learning.

Modality	Accuracy	AUC
Clinic + CT + Geo	0.758	0.807
CT + Geo	0.733	0.801
CT	0.669	0.663
Clinic + Geo	0.742	0.814
Geo	0.737	0.799
Clinic	0.538	0.561
Clinic + CT	0.606	0.634
Nothing	0.636	0.500

Table 5. Impact of modality exclusion on model performance (BMC test set). AUC: area under the receiver operating characteristic curve; Clinical: Clinical feature; CT: computed tomography; Geo: Geometric feature.

Modality	Shapley value (Accuracy)	Shapley value (AUC)
Clinic	-0.032	0.021
Geo	0.131	0.224
CT	0.029	0.066
Imbalance metric	0.109	0.135

Table 6. Quantitative modality contribution analysis using Shapley values. AUC: area under the receiver operating characteristic curve; Clinical: Clinical feature; CT: computed tomography; Geo: Geometric feature.

To investigate the dependency on each modality, we performed an ablation study on our final multi-modal model using clinical, CT image, and geometric features from the BMC dataset. In this experiment, each modality was individually ablated by replacing its input with the mean value across the dataset, and model performance was evaluated using the remaining modalities. The resulting changes in accuracy and AUC are summarized in Table 5. We also computed the Shapley value for each modality to quantify its contribution to model performance[49,50] and calculated the imbalance metric for the tri-modal case based on these values[51], as shown in Table 6. For model selection, we used AUC to determine the optimal epoch. Among the three modalities, geometric features demonstrated the highest contribution to AUC, followed by CT image features, while clinical features contributed the least. Although clinical features are well-established and medically interpretable, their relatively lower contribution reflects the average behavior observed in the test dataset and does not diminish their potential value in specific cases. Overall, the results indicate a slight inter-modality imbalance in feature contribution.

Discussion

In the present study, we developed a multi-modal deep learning model that integrates clinical, geometric, and imaging-derived features to enhance the prediction of AAA progression. The model extracts features from CT scans using a CNN, representing a single imaging modality. Additionally, geometric features, including 2D and 3D representations of the abdominal aorta, were derived from CT annotations, providing structural insights into aneurysm morphology. The third modality consists of clinical features, which are widely recognized as key factors influencing AAA progression, extracted from electronic medical records. An end-to-end multi-modal model was subsequently trained, integrating all three data sources within a unified neural network architecture. This comprehensive approach enables the model to learn complex interactions among anatomical, structural, and clinical parameters, offering a more robust predictive framework than single-modality models. The proposed multi-modal deep learning model demonstrated superior performance compared to traditional machine learning models, which are typically limited to single-feature data. Our model achieved an AUC of 0.807 and an accuracy of 0.758, highlighting its potential clinical utility for AAA risk stratification and surveillance.

Although the model's precision and F1-score were relatively low, it consistently showed strong discriminative performance, with an AUROC above 0.80. In early disease prediction scenarios, such as identifying patients at risk for rapid AAA progression, sensitivity and overall discrimination are often prioritized over precision, particularly when the clinical cost of missing high-risk cases is high. As emphasized in recent literature[52], models with modest precision can still offer substantial value in clinical settings if they help identify individuals who may benefit from additional imaging, monitoring, or follow-up care. Therefore, our model holds promise as a practical decision-support tool for guiding AAA surveillance strategies, even when precision metrics are relatively low.

Due to its asymptomatic nature, AAA is often detected incidentally during imaging studies conducted for unrelated medical purposes. Large-scale screening studies have revealed that the prevalence of AAA is significantly higher in older adults, particularly those over 65 years of age, and is more common in men[53]. Furthermore, established risk factors such as smoking, hypertension, peripheral arterial disease, and a family history of AAA are strongly associated with its development and progression[54]. Management of AAA involves regular surveillance and timely surgical intervention, which may include either open surgical repair or endovascular aneurysm repair, depending on the patient's condition and the size of the aneurysm[7]. Current guidelines recommend follow-up imaging using CT scans at 12-month intervals for aneurysms measuring between 4.0 and 4.9 cm, with shorter intervals of six months for aneurysms exceeding 5.0 cm[7,55]. However, accurately predicting AAA progression and rupture remains a challenge. While some studies indicate that larger aneurysms are associated with more rapid growth and a higher risk of rupture[56,57], there have also been reports of sudden ruptures occurring in smaller aneurysms, underscoring the variability in disease progression[58]. This unpredictability presents significant challenges for clinicians, highlighting the need for more advanced and patient-specific predictive tools to enhance clinical decision-making.

Recent advancements in AI offer promising solutions to address these challenges. DL, a subset of AI, has demonstrated exceptional performance in extracting and analyzing complex imaging features, often surpassing human experts in image recognition tasks[59–61]. Previous efforts to predict AAA progression have primarily utilized either DL[34,62] or traditional ML models independently[63–66]. However, studies that integrate both approaches, such as those applied to breast cancer risk prediction, have shown improved accuracy compared to conventional clinical guidelines[20]. Building upon these findings, this study introduces a novel 3D DL model that incorporates clinical features, geometric features, and radiomics derived from CT imaging to enhance the prediction of AAA progression.

Previous studies on the prediction of AAA progression have primarily relied on traditional ML models, such as logistic regression and support vector machines, with aortic diameter measured via ultrasound as the primary input feature[63]. While ultrasound remains a cost-effective and accessible screening tool, its limited spatial resolution restricts the ability to detect subtle structural changes in the aorta[24]. In contrast, CT imaging provides higher spatial resolution and three-dimensional visualization, making it the preferred modality for assessing the geometric and morphological changes associated with AAA. Our study incorporated advanced CT-based features, such as aortic tortuosity and thrombus volume, both of which have been identified as critical predictors of disease progression.

One of the early indicators of AAA progression is aortic tortuosity, a characteristic best observed as a three-dimensional structure of the aorta using CT imaging modalities. Chandrashekar et al.[24] demonstrated that geometric features such as curvature, undulation, and convex hull volume can quantify tortuosity, which is associated with AAA progression. However, their study primarily focused on patients with large aneurysms, as current screening guidelines prioritize older individuals with significantly dilated aortas. In contrast, our study enrolled patients with varying aortic diameters and ages, including those undergoing CT scans for unrelated conditions, thereby broadening the understanding of AAA progression across a more diverse population.

Thrombus formation within the aorta, another critical feature, has been associated with the growth and rupture of AAAs. Hirata et al.[67] quantified thrombus areas on CT scans and demonstrated their correlation with disease progression. While traditional ML models, such as logistic regression, support vector machines, random forests, and boosting algorithms, have been used to analyze thrombus features and other clinical data, they often fail to fully exploit the intricate relationships among multiple data modalities[66,68,69]. In the present study, baseline models optimized using the XGBoost method[13,32] were compared against our multi-modal DL model, which integrated clinical data with imaging-derived features.

The integration of radiomics into our approach enabled a comprehensive analysis of aortic morphology and structure by leveraging features derived from intensity, shape, and texture metrics. Radiomic features have been shown to correlate strongly with the progression of AAA. Wang et al.[70] demonstrated the utility of radiomics in identifying key morphological characteristics, such as aortic diameter, volume, and texture,

which were also incorporated into our model; importantly, Wang et al. enhanced this approach by manually annotating the aorta to differentiate between the lumen and wall regions, facilitating the extraction of precise geometric features critical for understanding vascular geometry. This detailed segmentation process allowed us to construct 3D geometric features of the abdominal aorta, which were subsequently used to develop shape-based radiomics models. These models outperformed conventional approaches that rely solely on clinical or imaging data by providing a more accurate characterization of vascular geometry, a fundamental factor in AAA progression. However, the manual nature of this annotation process presents a limitation for large-scale clinical implementation. Future advancements in automated segmentation techniques, such as those reported by Hirata et al.[67], may offer promising solutions to overcome this bottleneck, paving the way for the routine integration of such methods into clinical workflows.

The multi-modal deep learning framework introduced in this study represents a significant advancement over traditional machine learning approaches. While conventional models, such as logistic regression, random forests, and boosting algorithms, have proven effective in analyzing individual data types, they may fall short of capturing the complex interactions between modalities. Prior studies, including Lipkova et al.[71] and Steyaert et al.[72], have emphasized the importance of multi-modal data that combines clinical features with imaging-derived features in enhancing clinical predictions. Similarly, Yi et al.[73] demonstrated that hybrid models, which integrate radiomics and CNN-based analyses, achieve superior performance in predicting aortic dissection. Building upon these findings, our study utilized multi-modal data, including image-derived features, clinical features, and geometric features, within a single end-to-end deep learning architecture to provide robust predictions for the rapid progression of AAA.

A key focus of this study was the utilization of CNNs to extract features from CT images, thereby overcoming the inherent limitations associated with manual feature selection. Our study builds upon this foundation by employing an end-to-end deep learning framework that is capable of extracting and synthesizing data from both CT images and medical records, thus providing a comprehensive approach to AAA prediction. Moreover, comparisons with earlier studies utilizing CNN-based methods, such as those conducted by Golla et al.[18], underscore the importance of training models on heterogeneous datasets to enhance clinical applicability.

Unlike previous studies that relied on uniform datasets, our study incorporated a diverse patient cohort characterized by variations in CT imaging protocols, resolutions, baseline aortic diameters, age distributions, and gender ratios. This real-world heterogeneity enhances the model's robustness and generalizability, making it more applicable to broader clinical populations. By integrating multi-source data, our approach addressed the limitations of earlier models that were trained using specific patient groups, thereby improving predictive accuracy and clinical decision-making in the management of AAA. In a related study, Yi et al.[73] combined radiomics features with CNN-based models to predict aortic dissection. Their hybrid approach demonstrated superior performance compared to single-modality models, which aligns with our findings. By incorporating multi-modal data—such as clinical features, radiomic features, and geometric characteristics—our model achieved robust and reliable predictions for AAA progression. Furthermore, this model represents a significant advancement over traditional ML methods, which have historically struggled to capture the complex interactions among diverse data modalities.

The implications of this study extend beyond the prediction of rapid AAA progression. By demonstrating the efficacy of multi-modal data integration, our findings underscore the transformative potential of AI in the management of vascular diseases. For example, similar hybrid models that integrate CNN-based texture analysis with radiomic features have shown promising results in predicting outcomes for aortic dissection, as illustrated by Yi et al.[73]. The successful application of these methods to AAA suggests their potential adaptability to other disease contexts, thereby providing actionable insights for personalized medicine. Furthermore, this study contributes to the growing body of evidence supporting the use of AI in bridging the gap between imaging data and clinical decision-making, offering a framework for the development of more advanced predictive tools in vascular health.

Our study has several limitations. First, the manual annotation of the aorta is labor-intensive, which restricts the scalability of the model. While our study utilized expert-annotated CT data, the future integration of automated annotation tools, as explored in previous studies[17,19,74,75], will be essential for streamlining clinical workflows. Second, our model primarily relied on contrast-enhanced CT imaging, which may not always be available in routine practice. The synthesis of contrast-enhanced images from non-contrast CT scans, as demonstrated by recent advancements[76], presents a promising solution to this limitation. Third, external validation on independent patient cohorts is necessary to confirm the generalizability of the findings and to assess performance in various clinical settings. Fourth, we used the average annual growth rate calculated between the earliest and latest CT scans for each patient to define aneurysm progression. This approach may fail to capture interim episodes of rapid expansion or regression, particularly in patients with long intervals between scans, potentially masking transient periods of marked growth. Nonetheless, post hoc review of such cases consistently revealed minimal true expansion over prolonged intervals, and most clinically relevant rapid progression events were promptly identified through subsequent referrals and interventions. To address this potential labeling bias, future prospective studies with standardized and more frequent imaging follow-up will be important. Finally, integrating AI-driven models into routine clinical workflows requires careful consideration of potential barriers, including automation errors, interpretability of results, and clinician acceptance. The development of explainable AI frameworks, which provide insights into the model's decision-making process, will be crucial in fostering trust among healthcare providers. Additionally, designing user-friendly interfaces for clinicians will ensure that these tools can be seamlessly incorporated into existing workflows without disrupting patient care.

In conclusion, this study presents a novel multi-modal DL model for predicting the rapid progression of AAA by integrating CT imaging with clinical data. The superior performance of our model highlights the significance of multi-modal approaches in predicting complex diseases, representing a significant advancement in AAA

management and personalized medicine. By addressing current limitations, such as the reliance on manual annotations and the need for external validation, future research can further enhance the clinical applicability of this model. Ultimately, this approach has the potential to substantially improve patient outcomes by facilitating earlier and more accurate predictions of AAA progression, thereby paving the way for personalized management of vascular diseases.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Received: 6 April 2025; Accepted: 26 September 2025
Published online: 03 November 2025

References

1. Li, X., Zhao, G., Zhang, J., Duan, Z. & Xin, S. Prevalence and trends of the abdominal aortic aneurysms epidemic in general Population - A Meta-Analysis. *PLoS ONE*. **8**, e81260 (2013).
2. Bengtsson, H. & Bergqvist, D. Ruptured abdominal aortic aneurysm: A population-based study. *J. Vasc Surg.* **18**, 74–80 (1993).
3. Kessler, V., Klopf, J., Eilenberg, W., Neumayer, C. & Brostjan, C. AAA revisited: A comprehensive review of risk Factors, Management, and hallmarks of pathogenesis. *Biomedicines* **10**, 94 (2022).
4. Baxter, B. T., Terrin, M. C. & Dalman, R. L. Medical management of small abdominal aortic aneurysms. *Circulation* **117**, 1883–1889 (2008).
5. Secretariat, M. A. Endovascular repair of abdominal aortic aneurysm: an evidence-based analysis. *Ont. Heal Technol. Assess. Ser.* **2**, 1–46 (2002).
6. Gao, J. et al. The mechanism and therapy of aortic aneurysms. *Signal. Transduct. Target. Ther.* **8**, 55 (2023).
7. Chaikof, E. L. et al. The society for vascular surgery practice guidelines on the care of patients with an abdominal aortic aneurysm. *J. Vasc Surg.* **67**, 2–77e2 (2018).
8. Hirose, Y., Hamada, S. & Takamiya, M. Predicting the growth of aortic aneurysms: A comparison of linear vs exponential models. *Angiology* **46**, 413–419 (1995).
9. Stonebridge, P. A. et al. Growth rate of infrarenal aortic aneurysms. *Eur. J. Vasc Endovasc Surg.* **11**, 70–73 (1996).
10. Sweeting, M. J. & Thompson, S. G. Making predictions from complex longitudinal Data, with application to planning monitoring intervals in a National screening programme. *J. R Stat. Soc. Ser. A: Stat. Soc.* **175**, 569–586 (2011).
11. Kurvers, H. et al. Discontinuous, staccato growth of abdominal aortic aneurysms. *J. Am. Coll. Surg.* **199**, 709–715 (2004).
12. Khosla, S. et al. Meta-analysis of peak wall stress in ruptured, symptomatic and intact abdominal aortic aneurysms. *J. Br. Surg.* **101**, 1350–1357 (2014).
13. Doyle, B. J. et al. Biomechanical assessment predicts aneurysm related events in patients with abdominal aortic aneurysm. *Eur. J. Vasc Endovascular Surg.* **60**, 365–373 (2020).
14. Martufi, G. et al. Multidimensional growth measurements of abdominal aortic aneurysms. *J. Vasc Surg.* **58**, 748–755 (2013).
15. Xie, Y. et al. Automatic risk prediction of intracranial aneurysm on CTA image with convolutional neural networks and radiomics analysis. *Front. Neurol.* **14**, 1126949 (2023).
16. Shi, Z. et al. A clinically applicable deep-learning model for detecting intracranial aneurysm in computed tomography angiography images. *Nat. Commun.* **11**, 6090 (2020).
17. Spinella, G. et al. Artificial intelligence application to screen abdominal aortic aneurysm using computed tomography angiography. *J. Digit. Imaging.* **36**, 2125–2137 (2023).
18. Golla, A. K. et al. Automated screening for abdominal aortic aneurysm in CT scans under clinical conditions using deep learning. *Diagnostics* **11**, 2131 (2021).
19. Mu, N. et al. Automatic segmentation of abdominal aortic aneurysms from CT angiography using a context-aware cascaded U-Net. *Comput. Biol. Med.* **158**, 106569 (2023).
20. Stahlschmidt, S. R., Ulfenborg, B. & Synnergren, J. Multimodal deep learning for biomedical data fusion: a review. *Brief. Bioinform.* **23**, bbab569 (2021).
21. Cho, M. J., Lee, M. R. & Park, J. G. Aortic aneurysms: current pathogenesis and therapeutic targets. *Exp. Mol. Med.* **55**, 2519–2530 (2023).
22. Oliver-Williams, C. et al. Safety of men with small and medium abdominal aortic aneurysms under surveillance in the National health service screening programme. *Circulation* **139**, 1371–1380 (2019).
23. Oliver-Williams, C. et al. Lessons learned about prevalence and growth rates of abdominal aortic aneurysms from a 25-year ultrasound population screening programme. *J. Br. Surg.* **105**, 68–74 (2018).
24. Chandrashekar, A. et al. Prediction of abdominal aortic aneurysm growth using geometric assessment of computerized tomography images acquired during the aneurysm surveillance period. *Ann. Surg.* **277**, e175–e183 (2023).
25. Kent, K. C. Abdominal aortic aneurysms. *N Engl. J. Med.* **371**, 2101–2108 (2014).
26. Nordon, I. M., Hinchliffe, R. J., Loftus, I. M. & Thompson, M. M. Pathophysiology and epidemiology of abdominal aortic aneurysms. *Nat. Rev. Cardiol.* **8**, 92–102 (2011).
27. Sweeting, M. J., Thompson, S. G., Brown, L. C. & Powell, J. T. collaborators, on behalf of the R. Meta-analysis of individual patient data to examine factors affecting growth and rupture of small abdominal aortic aneurysms. *Br. J. Surg.* **99**, 655–665 (2012).
28. Radak, D., Tanaskovic, S., Katsiki, N. & Isenovic, E. R. Protective role of diabetes mellitus on abdominal aortic aneurysm pathogenesis: myth or reality? *Curr. Vasc Pharmacol.* **14**, 196–200 (2016).
29. Patro, S. G. K. & Sahu, K. K. Normalization: A preprocessing stage. *ArXiv* https://doi.org/10.48550/arxiv.1503.06462 (2015).
30. Fedorov, A. et al. 3D slicer as an image computing platform for the quantitative imaging network. *Magn. Reson. Imaging.* **30**, 1323–1341 (2012).
31. van Griethuysen, J. J. M. et al. Computational radiomics system to Decode the radiographic phenotype. *Cancer Res.* **77**, e104–e107 (2017).
32. Krishnapuram, B. et al. XGBoost. *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.* 785-794. https://doi.org/10.1145/2939672.2939785 (2016).
33. Teredesai, A. et al. Optuna. *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.* 2623-2631. https://doi.org/10.1145/3292500.3330701 (2019).
34. Tsafrir, D. et al. PyTorch 2: faster machine learning through dynamic python bytecode transformation and graph compilation. *Proc. 29th ACM Int. Conf. Arch. Support Program. Lang. Oper. Syst.* **2**, 929–947. https://doi.org/10.1145/3620665.3640366 (2024).
35. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. *2016 IEEE Conf. Comput. Vis. Pattern Recognit. CVPR*. 770-778. https://doi.org/10.1109/cvpr.2016.90 (2016).
36. Loshchilov, I. & Hutter, F. Decoupled weight decay regularization. *ArXiv* https://doi.org/10.48550/arxiv.1711.05101 (2017).
37. Buslaev, A. et al. Albumentations: fast and flexible image augmentations. *Information* **11**, 125 (2020).

38. Silva, P. et al. COVID-19 detection in CT images with deep learning: A voting-based scheme and cross-datasets analysis. *Inf. Med. Unlocked*. **20**, 100427 (2020).
39. WELCH, B. L., THE GENERALIZATION OF 'STUDENT'S' & PROBLEM WHEN SEVERAL DIFFERENT POPULATION VARLANCES ARE INVOLVED. *Biometrika* **34**, 28–35 (1947).
40. Fay, M. P. & Proschan, M. A. Wilcoxon-Mann-Whitney or t-test? On assumptions for hypothesis tests and multiple interpretations of decision rules. *Stat. Surv.* **4**, 1–39 (2010).
41. Plackett, R. L. Karl pearson and the Chi-Squared test. *Int. Stat. Rev. / Revue Int. De Statistique*. **51**, 59 (1983).
42. Seabold, S., Perktold, J. & Statsmodels econometric and statistical modeling with python. 92–96. https://doi.org/10.25080/majora-92bf1922-011 (2010).
43. Lee, S. W. Methods for testing statistical differences between groups in medical research: statistical standard and guideline of life cycle committee. *Life Cycle*. **2**, e1 (2022).
44. Hand, D. J. & Till, R. J. A simple generalisation of the area under the ROC curve for multiple class classification problems. *Mach. Learn.* **45**, 171–186 (2001).
45. Pedregosa, F. et al. Scikit-learn: Machine learning in python. https://doi.org/10.48550/arxiv.1201.0490 (2012).
46. DeLong, E. R., DeLong, D. M. & Clarke-Pearson, D. L. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* **44**, 837–845 (1988).
47. Robin, X. et al. pROC: an open-source package for R and S+to analyze and compare ROC curves. *BMC Bioinform.* **12**, 77 (2011).
48. SHAPIRO, S. S. & WILK, M. B. An analysis of variance test for normality (complete samples)†. *Biometrika* **52**, 591–611 (1965).
49. Shapley, L. S. A value for ퟐ-person games. In *Classics in Game Theory* (ed Shubik, M.) 69–79 (Princeton Univ. Press, 1953).
50. Parcalabescu, L. & Frank, A. MM-SHAP: A Performance-agnostic metric for measuring multimodal contributions in vision and Language models & tasks. *ArXiv* https://doi.org/10.48550/arxiv.2212.08158 (2022).
51. Xu, S., Cui, C., Huang, M., Wang, H. & Hu, D. BalanceBenchmark: A survey for multimodal imbalance learning. *ArXiv* https://doi.org/10.48550/arxiv.2502.10816 (2025).
52. Kocak, B. et al. Evaluation metrics in medical imaging AI: fundamentals, pitfalls, misapplications, and recommendations. *Eur. J. Radiol. Artif. Intell.* **3**, 100030 (2025).
53. Lederle, F. A. et al. The aneurysm detection and management study screening program: validation cohort and final results. *Arch. Intern. Med.* **160**, 1425–1430 (2000).
54. Song, P. et al. The global and regional prevalence of abdominal aortic aneurysms: A systematic review and modeling analysis. *Ann. Surg.* **277**, 912–919 (2023).
55. Members, W. C. et al. 2022 ACC/AHA guideline for the diagnosis and management of aortic disease A report of the American heart Association/American college of cardiology joint committee on clinical practice guidelines. *J. Am. Coll. Cardiol.* **80**, e223–e393 (2022).
56. Aggarwal, S., Qamar, A., Sharma, V. & Sharma, A. Abdominal aortic aneurysm: A comprehensive review. *Exp. Clin. Cardiol.* **16**, 11–15 (2010).
57. Thompson, S. et al. Systematic review and meta-analysis of the growth and rupture rates of small abdominal aortic aneurysms: implications for surveillance intervals and their cost-effectiveness. *Heal Technol. Assess.* **17**, 1–118 (2013).
58. Nicholls, S. C., Gardner, J. B., Meissner, M. H. & Johansen, K. H. Rupture in small abdominal aortic aneurysms. *J. Vasc Surg.* **28**, 884–888 (1998).
59. Liu, X. et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *Lancet Digit. Heal.* **1**, e271–e297 (2019).
60. Aggarwal, R. et al. Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis. *Npj Digit. Med.* **4**, 65 (2021).
61. Zhou, S. et al. (ed, K.) A review of deep learning in medical imaging: imaging Traits, technology Trends, case studies with progress Highlights, and future promises. *Proc. IEEE* **109** 820–838 (2021).
62. Jiang, Z., Do, H. N., Choi, J., Lee, W. & Baek, S. A deep learning approach to predict abdominal aortic aneurysm expansion using longitudinal data. *Front. Phys.* **7**, 235 (2020).
63. Lee, R. et al. Applied machine learning for the prediction of growth of abdominal aortic aneurysm in humans. *EJVES Short. Rep.* **39**, 24–28 (2018).
64. ULLAH, N. et al. Machine learning algorithms for the prognostication of abdominal aortic aneurysm progression: a systematic review. *Minerva Surg.* **79**, 219–227 (2024).
65. Chung, T. K., Gueldner, P. H., Aloziem, O. U., Liang, N. L. & Vorp, D. A. An artificial intelligence based abdominal aortic aneurysm prognosis classifier to predict patient outcomes. *Sci. Rep.* **14**, 3390 (2024).
66. Liljeqvist, M. L. et al. Geometric and Biomechanical modeling aided by machine learning improves the prediction of growth and rupture of small abdominal aortic aneurysms. *Sci. Rep.* **11**, 18040 (2021).
67. Hirata, K. et al. Machine learning to predict the rapid growth of small abdominal aortic aneurysm. *J. Comput. Assist. Tomogr.* **44**, 37–42 (2020).
68. Kontopodis, N. et al. Prediction of abdominal aortic aneurysm growth by artificial intelligence taking into account clinical, biologic, morphologic, and Biomechanical variables. *Vascular* **31**, 409–416 (2023).
69. Summers, K. L., Kerut, E. K., To, F., Sheahan, C. M. & Sheahan, M. G. Machine learning-based prediction of abdominal aortic aneurysms for individualized patient care. *J. Vasc Surg.* **79**, 1057–1067e2 (2024).
70. Wang, Y. et al. Contrast-enhanced CT radiomics improves the prediction of abdominal aortic aneurysm progression. *Eur. Radiol.* **33**, 3444–3454 (2023).
71. Lipkova, J. et al. Artificial intelligence for multimodal data integration in oncology. *Cancer Cell.* **40**, 1095–1110 (2022).
72. Steyaert, S. et al. Multimodal data fusion for cancer biomarker discovery with deep learning. *Nat. Mach. Intell.* **5**, 351–362 (2023).
73. Yi, Y. et al. Advanced warning of aortic dissection on Non-Contrast CT: the combination of deep learning and morphological characteristics. *Front. Cardiovasc. Med.* **8**, 762958 (2022).
74. Caradu, C., Spampinato, B., Vrancianu, A. M., Bérard, X. & Ducasse, E. Fully automatic volume segmentation of infrarenal abdominal aortic aneurysm computed tomography images with deep learning approaches versus physician controlled manual segmentation. *J. Vasc Surg.* **74**, 246–256e6 (2021).
75. Brutti, F. et al. Deep learning to automatically segment and analyze abdominal aortic aneurysm from computed tomography angiography. *Cardiovasc. Eng. Technol.* **13**, 535–547 (2022).
76. Liu, J. et al. Virtual contrast enhancement for CT scans of abdomen and pelvis. *Comput. Méd Imaging Graph.* **100**, 102094 (2022).

Author contributions

SJO and JWC contributed to the conception and design of the study; AW, YJO, J-SC, JSK, and H-JC contributed to the data acquisition and interpretation of data; J-iS, YH, IS, and KNJ contributed the data analysis; SJO, J-iS, ENK, and JSL drafted the manuscript; All authors revised the manuscript. All authors gave final approval and agreed to be accountable for all aspects of the work, ensuring integrity and accuracy.

Funding
This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (RS-2023-00212983), and the SNUH Research Fund (grant #0320232070 [2023–2594]). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-22167-z.

Correspondence and requests for materials should be addressed to S.J.O. or J.W.C.