



# Identifying Patients at Risk for Aortic Stenosis Through Learning from Multimodal Data

Tanveer Syeda-Mahmood<sup>(✉)</sup>, Yufan Guo, Mehdi Moradi, D. Beymer, D. Rajan, Yu Cao, Yaniv Gur, and Mohammadreza Negahdar

IBM Almaden Research Center, San Jose, CA, USA  
stf@us.ibm.com

**Abstract.** In this paper we present a new method of uncovering patients with aortic valve diseases in large electronic health record systems through learning with multimodal data. The method automatically extracts clinically-relevant valvular disease features from five multimodal sources of information including structured diagnosis, echocardiogram reports, and echocardiogram imaging studies. It combines these partial evidence features in a random forests learning framework to predict patients likely to have the disease. Results of a retrospective clinical study from a 1000 patient dataset are presented that indicate that over 25% new patients with moderate to severe aortic stenosis can be automatically discovered by our method that were previously missed from the records.

## 1 Introduction

With the growth of big data through large electronic health records (EHR), there is an opportunity to leverage medical image analysis in combination with other modality data in EHR to impact the quality of care to patients in a significant way. In this paper, we present one such clinical study in uncovering patients likely to have aortic stenosis. Aortic stenosis (AS) is a common heart disease that can result in sudden death. It can be diagnosed through the Doppler patterns in echocardiogram studies as shown in Fig. 1b. Although the disease can be treated through surgery or transcatheter aortic valve replacements (AVR), it often goes untreated for several reasons. The absence of chest pain and other symptoms may make the disease asymptomatic and not a candidate for detection in echocardiographer's instructions. This together with echocardiographer's skill errors can cause a

---

The original version of this chapter was revised: Correct name of the author "Yufan Guo" has now been updated. The correction to this chapter is available at [https://doi.org/10.1007/978-3-319-46726-9\\_73](https://doi.org/10.1007/978-3-319-46726-9_73)

---

**Electronic supplementary material** The online version of this chapter ([https://doi.org/10.1007/978-3-319-46726-9\\_28](https://doi.org/10.1007/978-3-319-46726-9_28)) contains supplementary material, which is available to authorized users.

Echocardiogram Report

Pt. Name: X00000X Echo Date: 10/2007  
 Echo Time: Facility Code: SEF  
 MR #: X000000X Sonographer: X0X  
 Patient Location:

Age: Gender: O HT:0.00 WT:0.00 BSA:0.00  
 Referring MD: X00000X  
 Indication:

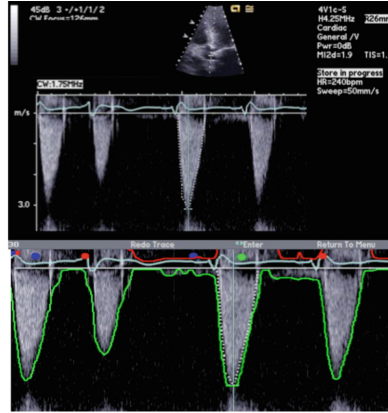
---

M-Mode Dimensions	Diastole	Systole	Normal Range
RVID			< 3.0 cm diast.
LVID	6.73	6.96	< 6.0 cm diast.
AO	3.83		< 4.0 cm diast.
IVS	1.21	3.68	< 1.2 cm diast.
LVPW	1.26	1.56	< 1.1 cm diast.
EPSS			< 0.6 cm
FE		11.6	28 - 53 %
LV EF		24.4	> 55 %

---

2D	Diastole	Systole	Normal Range
R07AW			0.2 to 0.7 cm diast.
IVS			0.6 to 1.1 cm diast.
LV			3.5 to 6.0 cm diast.
LVPW			0.6 to 1.1 cm diast.
Ro Root			2.1 to 3.4 cm diast.
LA A/P			2.3 to 3.8 cm syst.
LA Area A/C			8.8 to 23.4 cm <sup>2</sup> syst.
LA Volume			28 to 48 cc

(a)



(b)

**Fig. 1.** Illustration of missed diagnosis from echocardiogram (a) reports and (b) images

Doppler pattern depicting the disease to be missed entirely. Figure 1b (top) shows one such case where the echocardiographer missed the evidence for moderate aortic stenosis in the Doppler spectrum. When the relevant measurements are made by the echocardiographer and inserted into the study screens, they may still fail to make it into the overall report. Finally, even if the pattern is detected and makes it into the echocardiogram report, pure data entry errors in EHR can leave out the evidence of the disease from a patient record. With thousands of echocardiography studies taken annually, manual peer review is costly and rarely performed, with the result that many patients are going untreated.

The goal of this work is to develop an automated method for retrospectively predicting patients likely to have aortic stenosis by combining medical image analysis of Doppler patterns with textual content analysis of imaging and reports in a multimodal learning framework. Specifically, we extract evidence of aortic stenosis from 5 sources, namely, (a) billable diagnosis, (b) significant problems from EHR, (c) echocardiogram reports, (d) measurements shown on echocardiography video frames, and (e) CW Doppler patterns in echocardiography videos. Disease concepts are identified in echocardiogram reports using a concept extraction algorithm to detect UMLS concept vocabularies and their relevant associated measurements. Measurements captured by echocardiographers are reliably extracted through selective image processing and optical character recognition in tabular regions on echocardiogram video frames. Finally, diagnostically relevant measurements for aortic stenosis are automatically extracted from Doppler envelopes using a three step process of relevant Doppler frame identification, envelope tracing and measurement extraction. The frame identification involved classification of convolutional neural network (CNN)-based learned features from Doppler regions. The envelop extraction was made robust by incorporating echocardiographer's tracings. Finally the disease-specific features extracted from each multimodal source of information are combined using a random forest learning formulation to predict patients that are likely to have aortic valve disease.

## 2 Related Work

To our knowledge, this is the first work on identifying patients at risk that combines medical text and image analysis of echocardiogram studies. While previous studies have argued for the use of multimodal information for cohort identification [7], the primary information leveraged was either structured or textual data. The work reported here, however, overlaps three inter-disciplinary fields of text analysis, optical character recognition (OCR), and medical image analysis each of which is rich in literature. Several algorithms for extraction of clinical concepts from text have been reported in [9]. However, measurements must be extracted in addition to disease name mentions for aortic stenosis detection, which has not been addressed previously. Similarly, while there is considerable work in OCR in general, extracting clinical measurements from text screens of echocardiogram studies has not been well-addressed with the reported methods relying on manual creation of templates for various manufacturer’s echo screens [8]. Finally, reliable extraction of Doppler envelopes has proved to be notoriously challenging particularly in the presence of electrocardiogram (ECG) fluctuations during arrhythmia and overlay artifacts in Doppler spectra [3,5,9]. Lastly, the automatic selection of Doppler frames depicting aortic valves has not been previously reported in literature.

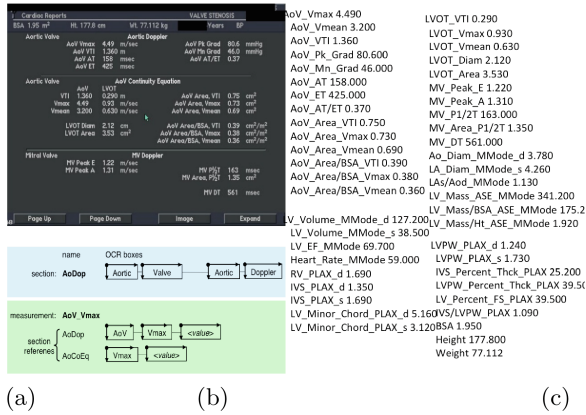
## 3 Disease Evidence Extraction from Multimodal Data

**Disease Extraction from Reports.** To extract evidence of aortic stenosis from echocardiogram reports, we generated a large knowledge graph of over 5.6 million concept terms by combining over 70 reference vocabularies such as SNOMED CT, ICD9, ICD10, RadLex, RxNorm, and LOINC and used its concept nodes as vocabulary phrases. The occurrences of clinical concepts within sentences of the clinical reports uses the longest common suffix (LCF) algorithm as described in [9]. To detect evidence of stenosis, we find tuples of  $\langle D_i, S_j, A_k, V_l \rangle$ , where  $D_i$  are disease name indicators (e.g. “aortic valve disorders”, “aortic valve stenosis”, etc.),  $S_j$  are specific symptoms associated with the disease such as “chest pain”,  $A_k$  are anatomical abnormalities such as “thickened”, “calcified”, and  $V_l$  are qualifiers such as “mild, moderate, severe”. These detections are done within neighboring sentences in selected paragraphs where the aortic valve is described in echocardiogram reports.

Next, we selected key measurement names indicating aortic stenosis as per AHA guidelines, namely, peak velocity, mean pressure gradient, and aortic valve area. Using their values ranges and units, as per guidelines, we developed a measurement name-value pair detector. As the spoken utterances of these names vary in echocardiograms, we did a n-gram analysis of a corpus of over 50,000 reports in our data collection to identify all such significant variants of the measurement names. To detect occurrences of measurement names and their associated values within the context of a detected sentence, we analyze the pattern of their occurrences in a sentence using part-of-speech (POS) tagging, and dependency graph parsing [4]. For each root concept (e.g. ‘gradient’), a chain of its modifiers

**Table 1.** False discovery rate (FDR) of disease (AS) and measurement (peak velocity and mean gradient) detection.

	FDR	False positives
AS	2/191	Indication/Hx: EVAL FOR MS/MR, <b>AS/AI</b> <i>De-Identified AS SMOKER</i>
Peak velocity	1/364	aortic stenosis is present. The <b>aortic valve peak velocity</b> is 2.6 ↔ <b>9 m/s</b> , the peak gradient is 28.9 mmHg,
Mean gradient	0/410	-


**Fig. 2.** Illustration of measurement extraction from echocardiography screens.

(in the form of nouns or adjectives, e.g. ‘mean trans aortic’) were automatically identified from a sentence using the Stanford POS tagger [4]. By analyzing thousands of sentences containing the occurrences of measurement vocabulary terms in connection with measurement values and units, we formed regular expression patterns, such a pattern “< A >< B >< C >” where “A” is any disease indicating phrase A: {aorta, aortic, AV, AS}, “B” is any measurement term {gradient, velocity, area}, and C is no negation terms of the kind {no, not, without, neither, none}. Once the pattern was matched, we looked for numeric values following the measurement names in the same sentence that were juxtaposed with names of relevant units. An example of aortic stenosis measurement extraction is illustrated below in bold.

**Aortic Valve:** The **aortic valve** is thickened and calcified. Severe **aortic stenosis** is present. The **aortic valve peak velocity** is **6.18 m/s**, the peak gradient is 152.8 mmHg, and the **mean gradient** is **84.9 mmHg**. The **aortic valve area** is estimated to be **0.28 cm<sup>2</sup>**.

In general, the text-based aortic stenosis detection is fairly stable with very few false positives as indicated in Table 1. Only 3 errors were observed after a thorough analysis of the detected cases, as listed in the third column.

**Table 2.** Accuracy of Doppler envelop extraction and measurement calculation.

Measurement made	Images tested	Error
$V_{max}$	1054	$0.29 \pm 0.78$ m/s
$M_g$	785	$0.08 \pm 10.05$ mmHg

+

**Extracting Echocardiographer Measurements.** The evidence for aortic stenosis can be extracted from the measurements made by the echocardiographer captured as text-only screens such as the one shown in Fig. 2a. To extract the measurements, we select the frames depicting the measurements and apply relevant tabular template to identify the semantic names of the measurements. An optical character recognition algorithm is then used to extract text. Unlike the approach in [8], we use a different OCR engine (DataCap) and learn the document layout templates of device manufacturer’s screens automatically. The template learning is focused per anatomical region and exploits the invariance in topological layout of the measurement name value pairs in the tabular regions. Once the templates are learned, they are matched to any given text only screen to read off the expected measurement names. Following the approach in [8], we process the images within the text regions through an image enhancement process to increase the robustness of OCR. Figure 2c shows the text extracted from measurement screen of Fig. 2a using our video text detection algorithm. The OCR-based measurement extraction module was tested on 114 text-only frames across 114 patients, and a total of 1719 measurements were verified. For this validation set, our system extracted 99.7% of the measurements correctly, with the remaining errors caused by the numeric values being split by the OCR engine.

**Disease Extraction from Doppler Image Analysis.** In Doppler echocardiography images, the clinically relevant region is known to be within the Doppler spectrum, contained in a rectangular region of interest as shown in Fig. 1b. To ensure the measurement extraction is attempted on relevant frames depicting the aortic valve, we developed a classifier using features derived from the region depicting Doppler patterns in images. This image region was fed to a pre-trained convolutional neural network (CNN) consisting of 5 convolution layers, two fully connected layers and a SoftMax layer with 1000 output nodes [2]. The CNN is being used as a feature generator here as has been reported in other literature [6]. Even though the CNN was trained in another imaging domain, the earlier layers of the neural network capture generic features such as edges which are also applicable in our domain. For our task of feature generation, we harvest a feature vector of size 4096 at the output of the first fully connected layer of the network and classify the images using a support vector machine (SVM) classifier. To train the SVM, we created an expert reviewed dataset of 496 CW Doppler patterns, each labeled with one of the four valve types. A set consisting of 100 of these images was randomly isolated as a test set. The SVM was optimized for kernel type and slack and kernel variables on the remaining 396 images using five-fold cross validation. Using the CNN derived features, the SVM achieved an accuracy of 92% across all valves with all aortic valve CW Doppler frames being

labeled correctly. The tricuspid stenosis valve pattern accounted for nearly half the errors as it is similar to the aortic stenosis valve pattern.

**Extraction of Doppler Patterns.** Our method of extracting Doppler spectrum uses similar pre-processing steps of region of interest detection, ECG extraction, and periodicity detection as described in [9], but adds a major enhancement exploiting the tracings of echocardiographers as shown in Fig. 3. To extract echocardiographer’s envelope annotation, we exclude the calculated Doppler velocity profile from the ROI and apply Otsu’s thresholding algorithm on the remaining image to highlight the manual delineation which is connected to the baseline. Then, we add the extracted annotation to the filled up largest region, as shown in Fig. 3 and trace the boundary pixels. The Doppler envelop extraction was tested on over 7000 images during training, and the results of the various stages of processing are indicated in Table 2.

**Measurement Extraction from Doppler Patterns.** Using the AHA guidelines, the maximum jet velocity ( $V_{max}$ ) is defined as the peak velocity in the negative direction for the Doppler pattern for aortic stenosis. Since the Doppler pattern envelope traces are available, the pixel value of the negative peak in the Doppler spectra can be easily noted. To convert the imaging-based measurement to a physical velocity value, we analyze the text calibration markers on the vertical axis in the ROI using OCR engine to read off the velocity value. The maximum value of velocity during systole within each cycle is a candidate for the  $V_{max}$ . The second measurement indicative of aortic stenosis is mean pressure gradient (MPG). MPG is calculated from velocity information following the estimation reported in [1] as  $M_g \approx \sum_V \frac{4V^2}{N}$  where N is the number of pixels within the QT interval of ECG, and V is the velocity.

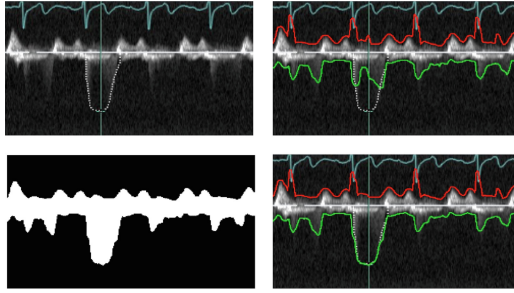
**Disease Prediction using Multimodal Learning.** Collecting all the measurements derived from each modality processing, we form a feature vector as follows.

$$F_p = \{V_{1b}, V_{2s}, V_{3t}, V_{4t}, V_{5t}, V_{6o}, V_{7o}, V_{8i}, V_{9i}\} \quad (1)$$

where the ‘b’ is for billable diagnosis, ‘s’ for significant problems, ‘t’ for textual reports, ‘o’ for video text, and ‘i’ for image analysis features. The first 3 features are binary while the rest are actual measurements made in the respective modalities. To train the predictor, we use a set of patients with known aortic stenosis (confirmed diagnosis in EHR), and learn the correlation between feature values and the disease label (aortic stenosis) using a random forests learner. The random forests were constructed with 100 trees, with each tree having a minimum node size of 10, and maximum depth of 10.

## 4 Clinical Study Results

We conducted a retrospective clinical study on a large patient data set acquired from a nearby hospital. The experimental context was to evaluate if there were missed diagnosis of aortic stenosis in their records when in fact evidence could be



**Fig. 3.** Illustration of Doppler envelop extraction using echocardiographer annotations.

**Table 3.** Comparative performance of rule-based baseline and random forest with features extracted from structured information, reports, images, and OCR text. min(I,O) refers to the fusion of image and OCR features by taking the minimum of the two for each individual feature/parameter.

	Features					Performance			
	Structured	Report	Image	OCR	min(I,O)	Precision	Recall	F-score	Accuracy
Baseline	x	x	x	x		0.84	1.00	0.93	0.92
	x					1.00	0.53	0.70	0.81
		x				0.96	0.55	0.70	0.81
			x			0.80	0.50	0.62	0.75
				x		0.94	0.50	0.66	0.79
Random					x	0.94	0.56	0.70	0.81
Forest			x	x		0.78	0.59	0.67	0.77
		x			x	0.93	0.73	0.82	0.87
		x	x	x		0.82	0.71	0.77	0.83
	<b>x</b>	<b>x</b>			<b>x</b>	<b>0.96</b>	<b>0.89</b>	<b>0.92</b>	<b>0.94</b>
	x	x	x	x		0.87	0.89	0.88	0.90

found from the underlying clinical data. Specifically, we restricted the analysis to patients for which all 4 modalities of information were available, namely, billable diagnosis, significant problems, and echocardiogram reports and imaging studies giving rise to a total of 991 patients with 1,226 reports and 121,811 Doppler images. These studies were independently validated clinically and 395 patients were found to have aortic stenosis serving as the ground truth.

A 10 fold cross-validation was done by randomly splitting the data into 10 folds, 9 for training and 1 for testing. Table 3 shows the precision, recall, F-score, and overall accuracy of the baseline and random forests with different combinations of features, including a fusion of image and OCR features – referred to as min(I,O). Selecting the minimum of these two values gave a more conservative estimate of the severity of the disease. Out of the 395 patients manually identified by experts, 99 were newly discovered patients from our multimodal analysis giving rise to over 25 % new discoveries.

**Comparison Against Baseline.** Our baseline was a rule-based model, which returned all patients with at least one piece of evidence from any of the five



sources. Here the evidence was either the presence of disease mentions or exceeding the normal ranges for  $V_{max}$  and  $M_g$  according to the AHA guidelines. The best-performing model was a random forest with features from all the different sources, achieving 96 % precision that is 12 % higher than the baseline. Combining features using random forests compensates for potential errors in individual modality detections, making its precision higher than the baseline method. The higher precision will reduce unnecessarily flagging of patients which would have otherwise have lowered the confidence in such prediction system for practical uses.

## 5 Conclusions

In this paper we have presented a new use of medical image analysis in combination with textual and other multimodal data analysis for purposes of identifying patient cohorts at risk for serious diseases such as aortic stenosis. While the textual detection method can be easily generalized for other diseases, future work will focus on developing disease detectors in imaging modalities to augment the decision making.

## References

1. Baumgartner, H., Hung, J., Bermejo, J., Chambers, J.B., Evangelista, A., Griffin, B.P., Iung, B., Otto, C.M., Pellikka, P.A., Quiñones, M.: Echocardiographic assessment of valve stenosis: EAE/ASE recommendations for clinical practice. *Eur. Heart J. Cardiovasc. Imaging* **10**(1), 1–25 (2009)
2. Chatfield, K., Simonyan, K., Vedaldi, A., Zisserman, A.: Return of the devil in the details: delving deep into convolutional nets. In: *British Machine Vision Conference* (2014)
3. Chen, T., Kumar, R., Troianowski, G., Syeda-Mahmood, T., Beymer, D., Brannon, K.: PSAR: predictive space aggregated regression and its application in valvular heart disease classification. In: *ISBI*, pp. 1122–1125 (2013)
4. Manning, C.D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S.J., McClosky, D.: The Stanford CoreNLP natural language processing toolkit. In: *Association for Computational Linguistics (ACL) System Demonstrations*, pp. 55–60 (2014)
5. Shechner, O., Sheinovitz, M., Feinberg, M., Greenspan, H.: Image analysis of doppler echocardiography for patients with atrial fibrillation. In: *ISBI*, pp. 488–491 (2004)
6. Shin, H.C., Roth, H.R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., Summers, R.M.: Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Trans. Med. Imaging* **35**(5), 1285–1298 (2016)
7. Shivade, C., Raghavan, P., Fosler-Lussier, E., Embi, P.J., Elhadad, N., Johnson, S.B., Lai, A.M.: A review of approaches to identifying patient phenotype cohorts using electronic health records. *J. Am. Med. Inform. Assoc.* **21**(2), 221–230 (2014)
8. Syeda-Mahmood, T., Beymer, D., Amir, A.: Disease-specific extraction of text from cardiac echo videos for decision support. In: *ICDAR*, pp. 1290–1294 (2009)
9. Syeda-Mahmood, T., Kumar, R., Compas, C.: Learning the correlation between images and disease labels using ambiguous learning. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) *MICCAI 2015*. LNCS, vol. 9350, pp. 185–193. Springer, Heidelberg (2015). [https://doi.org/10.1007/978-3-319-24571-3\\_23](https://doi.org/10.1007/978-3-319-24571-3_23)