SIIM CAIMI25
Conference on Artificial Intelligence in Medical Imaging
OCT 20-21 | UNIVERSITY OF CALIFORNIA, IRVINE

Pre-Imaging Predictors of Cardiac MR Image Quality Using Large Language Model-Based Clinical Data Extraction

Masha Bondarenko, University of California, San Francisco; Hong Yu, MD, PhD; Ali Nowroozi, MD; Jae Ho Sohn, MD, MS

Introduction/Background
Cardiac magnetic resonance (CMR) image quality is essential for accurate cardiovascular diagnosis, yet suboptimal scans are common and can compromise clinical decision-making. Large language models (LLMs) offer a scalable approach for extracting structured clinical information from free-text notes to expedite data collection. This study aimed to develop a model that uses LLM-extracted clinical features to predict poor CMR image quality prior to imaging.

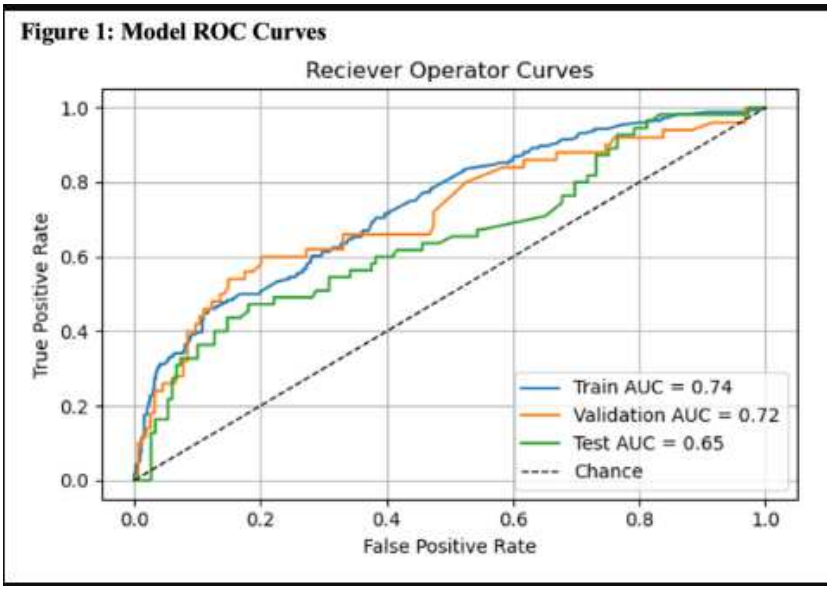
Methods/Intervention
1,020 adult CMR exams performed at UCSF from 2014 to 2024 were analyzed. A HIPAA-compliant LLM was used to classify image quality from radiology reports into four categories: Fair, Suboptimal, Severely Limited, and Non-diagnostic. Categories were binarized into Poor (Severely Limited/Non-diagnostic) and Good (Fair/Suboptimal) and verified by a radiologist. Up to ten pre-imaging clinical notes per patient were also processed by the LLM to extract clinical features, grouped into radiologist-defined clinical conditions, and verified by the radiologist. A logistic regression model with L1 regularization was trained on 42 variables using demographics, clinical history, and scan parameters, and evaluated via internal testing.

Results/Outcome
Among 1,020 reports from unique patients (mean age 52.7 ± 18.8, male = 618/1022), there was significant agreement between the radiologist and the LLM ($\kappa = 0.99$). Statistically significant predictors of poor image quality included 3T scanner use (adjusted odds-ratio (OR, 95% CI) = 2.87 [2.00, 4.12], $p< 0.001$), outpatient status (0.36 [0.25, 0.52], $p< 0.001$), non-smoker status (OR 0.68 [0.49, 0.92], $p=0.01$), and histories of valve disease (OR 0.51 [0.36, 0.73], $p=0.0002$), noncompliance (OR=1.89 [1.26, 2.85], $p=0.00$), infection (OR 1.53 [1.11, 2.10], $p=0.008$), pulmonary embolism (OR 2.53 [1.193, 5.37], $p=0.015$), and claustrophobia (OR= 2.52 [1.18, 5.35], $p=0.01$). The final model achieved an AUC of 0.74 (training), 0.72 (validation), and 0.65 (test).

Conclusion
LLM-extracted pre-scan clinical features enable prediction of poor CMR image quality before imaging. This approach can guide proactive interventions to improve diagnostic yield.

Statement of Impact
This study demonstrates the feasibility of using LLMs for automated, pre-scan risk stratification, enabling proactive interventions to improve image quality and diagnostic yield in CMR.

Figure 1: Model ROC Curves



ROC curve for final model.

Table 1: Multivariate Results on Significant Variables

Variables	Logistic Regression Coefficient	Adjusted Odds Ratio	95% CI	P>|z|
constant	-1.0439	0.3521	(0.2195, 0.5647)	<0.0001*
3T	1.0557	2.8739	(2.0037, 4.1222)	<0.0001*
Outpatient	-1.002	0.3671	(0.2563, 0.5259)	<0.0001*
Valve Disease	-0.6629	0.5153	(0.3638, 0.7301)	0.0002*
Noncompliance	0.6405	1.8975	(1.2613, 2.8547)	0.0021*
Infection	0.4256	1.5306	(1.1155, 2.1001)	0.0084*
Pulmonary Embolism	0.9296	2.5335	(1.1939, 5.376)	0.0154*
Never Smoker	-0.3842	0.681	(0.4987, 0.9298)	0.0156*
Claustrophobia	0.9259	2.5242	(1.1898, 5.3553)	0.0158*
Heart Failure	0.3171	1.3732	(0.9644, 1.9552)	0.0786
Ventricular Tachycardia	0.3933	1.4818	(0.9383, 2.3404)	0.0917
Arrhythmia	-0.2234	0.7998	(0.5771, 1.1085)	0.1797
Pulmonary Hypertension	0.2265	1.2542	(0.8384, 1.8762)	0.2704
Infiltrative	-0.2614	0.77	(0.4802, 1.2346)	0.2779
Spanish Primary Language	0.4218	1.5247	(0.6925, 3.3572)	0.2949
Chronic Kidney Disease	0.2333	1.2628	(0.7959, 2.0035)	0.3217
Mixed Race	0.3321	1.3938	(0.6439, 3.0172)	0.3994

** Highlighted rows correspond to statistically significant variables at an alpha level of 0.05.*

Odds ratios for selected variables.

Table 2: Performance of the L1-regularized logistic regression model.

Dataset Splits	Accuracy	Sensitivity	Specificity	Precision	F1 Score	AUC
Train	0.758	0.227	0.972	0.769	0.351	0.737
Validation	0.775	0.120	0.987	0.750	0.207	0.716
Test	0.750	0.164	0.966	0.643	0.261	0.650

Model statistics.

Keywords

Cardiac MRI; Large language models; Image Quality