Artificial Intelligence Predicts Hospitalization for Acute Heart Failure Exacerbation in Patients Undergoing Myocardial Perfusion Imaging

Attila Feher*1,2, Bryan Bednarski*3, Robert J. Miller3,4, Aakash Shanbhag3,5, Mark Lemley3, Leonidas Miras6, Albert J. Sinusas1,2, Edward J. Miller1,2, and Piotr J. Slomka3

1Section of Cardiovascular Disease, Department of Internal Medicine, Yale School of Medicine, New Haven, Connecticut; 2Department of Radiology and Biomedical Imaging, Yale School of Medicine, New Haven, Connecticut; 3Departments of Medicine (Division of Artificial Intelligence in Medicine), Biomedical Sciences, and Imaging, Cedars-Sinai Medical Center, Los Angeles, California; 4Department of Cardiac Sciences, University of Calgary, Calgary, Alberta, Canada; 5Signal and Image Processing Institute, Ming Hsieh Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, California; and 6Division of Cardiology, Bridgeport Hospital, Yale University School of Medicine, Bridgeport, Connecticut

Heart failure (HF) is a leading cause of morbidity and mortality in the United States and worldwide, with a high associated economic burden. This study aimed to assess whether artificial intelligence models incorporating clinical, stress test, and imaging parameters could predict hospitalization for acute HF exacerbation in patients undergoing SPECT/CT myocardial perfusion imaging. Methods: The HF risk prediction model was developed using data from 4,766 patients who underwent SPECT/CT at a single center (internal cohort). The algorithm used clinical risk factors, stress variables, SPECT imaging parameters, and fully automated deep learning–generated calcium scores from attenuation CT scans. The model was trained and validated using repeated hold-out (10-fold cross-validation). External validation was conducted on a separate cohort of 2,912 patients. During a median follow-up of 1.9 y, 297 patients (6%) in the internal cohort were admitted for HF exacerbation. Results: The final model demonstrated a higher area under the receiver-operating-characteristic curve (0.87 ± 0.03) for predicting HF admissions than did stress left ventricular ejection fraction (0.73 ± 0.05, P < 0.0001) or a model developed using only clinical parameters (0.81 ± 0.04, P < 0.0001). These findings were confirmed in the external validation cohort (area under the receiver-operating-characteristic curve: 0.80 ± 0.04 for final model, 0.70 ± 0.06 for stress left ventricular ejection fraction, 0.72 ± 0.05 for clinical model; P < 0.001 for all). Conclusion: Integrating SPECT myocardial perfusion imaging into an artificial intelligence–based risk assessment algorithm improves the prediction of HF hospitalization. The proposed method could enable early interventions to prevent HF hospitalizations, leading to improved patient care and better outcomes.

Key Words: heart failure; congestive heart failure; SPECT; myocardial perfusion imaging; machine learning

J Nucl Med 2024; 00:1–7
DOI: 10.2967/jnumed.123.266761

Heart failure (HF) is a major cause of morbidity and mortality in the United States and worldwide. The total prevalence of HF is projected to increase by almost 50% from 2012 to 2030, affecting more than 8 million adults (1). HF is associated with a high socioeconomical burden, with frequent emergency room visits and impatient hospitalizations for HF exacerbation. Recent data show that hospitalizations for HF exacerbations dramatically increased from 2008 to 2018 for both HF with reduced ejection fraction (HFrEF) and HF with preserved ejection fraction (HFpEF) (1). Despite significant advancements in therapies for the treatment of HFrEF, quality of life and life expectancy for those affected by HF remain poor, with estimated 5-y survival of 50% after the diagnosis of HF is established (2). Identifying patients who are at risk for HF exacerbation enables opportunities for the implementation of prevention strategies.

Ischemic cardiomyopathy in the setting of obstructive coronary artery disease (CAD) is the most common primary etiology of HFrEF, being responsible for 40%–70% of cases (2). SPECT myocardial perfusion imaging (MPI) is the most frequently used imaging modality for the diagnosis of CAD. In the last few decades, SPECT MPI has undergone major advances with the advent of cadmium zinc telluride solid-state detector technology, specialized collimators, and software-based resolution recovery, resulting in improved performance when compared with conventional SPECT technology.

Artificial intelligence (AI) has been previously used to improve diagnostic accuracy for the prediction of obstructive CAD in patients undergoing SPECT MPI (3,4) and has been applied to predict adverse cardiovascular events in this patient population (5,6). In addition, several clinical models have been developed to predict incident HF in the general population (7,8); however, to our knowledge no AI model has yet been developed to predict HF exacerbations incorporating data obtained during SPECT MPI. Therefore, we set out to evaluate whether AI models incorporating clinical, stress test, SPECT imaging parameters, and fully automated deep learning coronary artery calcium (CAC) scores using CT attenuation correction (CTAC) scans can predict hospitalization due to HF in patients undergoing SPECT/CT MPI.
MATERIALS AND METHODS

Study Population
The REFINE SPECT registry (9) is a multicenter observational cohort study including patients undergoing SPECT MPI for known or suspected CAD using cadmium zinc telluride solid-state detector systems. Two distinct sites within REFINE SPECT, for which HF outcomes were available, were used for model development and hold-out validation. Yale University patients (n = 4,766) were used for development and internal 10-fold cross validation. University of Calgary patients (n = 2,912) were withheld from all model training and used as an external testing set. Patients without CTAC were excluded. The study protocol complied with the Declaration of Helsinki, and sites obtained either written informed consent or a waiver of consent to the use of the deidentified data. The study was approved by the institutional review board at all sites, with the overall study approved by the institutional review board at Cedars-Sinai Medical Center.

Clinical Data
In this retrospective study, we collected demographic data about the participants’ age, sex, body mass index, family history of CAD, smoking status, and whether they had hypertension, dyslipidemia, diabetes, peripheral artery disease, a prior diagnosis of HFpEF or HFrEF, a history of previous myocardial infarction, and prior coronary artery bypass graft surgery. Resting blood pressure and heart rate were acquired before exercise or before stressor administration. Peak stress heart rate and blood pressure, as well as clinical and electrocardiogram response to stress, were collected at the time of clinical reporting and were included in the model without distinction of whether it was recorded during exercise or pharmacologic stress. Heart rate response was defined as the difference between peak stress (exercise or pharmacologic) and resting heart rates. The primary endpoint was hospitalization for HF exacerbation determined by review of electronic medical records (9). HF hospitalizations with a left ventricular ejection fraction (LVEF) of less than 40% were categorized as HFrEF, whereas HF hospitalizations with an LVEF of 40% or more at the time of HF hospitalization were categorized as HFpEF based on prior definitions (10).

Image Acquisition and Protocol
Imaging protocols and data collection details for the REFINE SPECT population have been described previously (9), with additional details in the supplemental materials (supplemental materials are available at http://jnm.snmjournals.org) (11–15). All patients were imaged with a 570c or 530 solid-state SPECT camera system (GE Healthcare). Supine stress and rest imaging data were primarily used for all patients; if missing, prone data were used instead.

Deep Learning Calcium Scoring
The pretrained convLSTM model was used to infer automatic, AI-generated CAC scores from CTAC images for the internal (Yale, n = 4,766) and external (Calgary, n = 2,912) populations, which were combined with other parameters used for this analysis. Details regarding CAC scoring and model training are provided in the supplemental materials (5,16–18), and demographic information for the training cohort is provided in Supplemental Table 1. In the training dataset, during calcification scoring by the expert readers, patients with stents, coronary artery bypass grafting, pacemaker wires, or other artifacts were marked, and this information was included at the time of training. As such, the AI training set thus could learn how to avoid these artifacts.

Machine Learning
Gradient-boosted decision tree models (XGBoost, Python version 1.3.3) were trained for the binary classification of HF at follow-up. Tenfold cross validation was performed with the Yale University cohort for internal model development and testing. In each fold, a unique data split (80% training, 10% validation, 10% testing) was used to train and test a new model such that across all folds all patients were used in testing exactly once. A nested grid search within each fold was used for hyperparameter search and selection. A final model was built using a split (90% training, 10% validation) using the Yale University cohort with the best hyperparameters (on average from all folds) to maximize the final training sample size.

For external validation, the final model was developed exclusively in the Yale University cohort and was applied directly to the previously unseen University of Calgary data.

Model Explainability
Feature importance analysis was performed using 2 methods. First, the in-built XGBoost feature importance methods were used to compare the absolute information gain from all input variables to the trained model. After performing cross validation, we selected our best model and used the 10 most important features to build an additional model for comparison. Next, Shapley additive explanations were used to analyze testwise feature influence on model performance (19).

Data Preprocessing
A complete list of variables is provided in Supplemental Table 2. Input variables with more than 20% missing data in the internal cohort (10 total) were dropped from the analysis to reduce missing-data bias. Remaining missing variables were accommodated by XGBoost’s in-built method for learning per-variable missing data; no further imputation methods were needed. The complete feature set included 30 clinical, 14 stress, 26 perfusion imaging, and 2 calcium features, totaling 72 variables. Models were created to appropriately evaluate 5 multivariable models: clinical only, clinical plus stress, clinical plus stress plus calcium, clinical plus stress plus nuclear imaging, and clinical plus stress plus calcium plus nuclear imaging. Three univariable models were also evaluated for summed stress score, stress total perfusion deficit, and stress LVEF. The AI model is inherently multivariable because all variables are considered simultaneously in the prediction; therefore, the measure of importance is provided in terms of relative importance of the variable in the AI model.

Statistical Analysis
Details regarding statistical analysis methods are provided in the supplemental materials (20).

RESULTS

Patient Characteristics
The final development and internal validation population comprised 4,766 patients with a median age of 64 y (interquartile range [IQR], 56–73 y). Of these, 2,647 (56%) were men, and 298 (6.3%) had a prior history of HF (134 patients with HFrEF and 164 patients with HFpEF). There were 2,912 patients included in the external testing group, with a median age of 67 y (IQR, 59–75 y). Of these, 1,565 (54%) were men, and 127 patients (4.4%) had a prior HF history (56 patients with HFrEF and 71 patients with HFpEF). The characteristics of patients in the internal and external testing groups are shown in Table 1.

Feature Importance
Among the clinical, imaging, and stress parameters, prior HFrEF and HFrEF history, exercise duration, heart rate response to stress, left ventricular mass, and stress end-systolic shape index had the highest variable importance for the final AI model across all folds in cross-validation (Fig. 1) and in external validation
TABLE 1
Internal and External Cohort Patient Characteristics

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Internal cohort (n = 4,766)</th>
<th>External cohort (n = 2,912)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (y)</td>
<td>64 (56–73)</td>
<td>67 (59–75)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Men</td>
<td>2,647 (56%)</td>
<td>1,565 (54%)</td>
<td>0.13</td>
</tr>
<tr>
<td>Body mass index (kg/m²)</td>
<td>29 (25–34)</td>
<td>29 (25–33)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Diabetes</td>
<td>1,236 (26%)</td>
<td>757 (26%)</td>
<td>&gt;0.9</td>
</tr>
<tr>
<td>Hypertension</td>
<td>3,055 (64%)</td>
<td>1,734 (60%)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Dyslipidemia</td>
<td>2,546 (53%)</td>
<td>1,478 (51%)</td>
<td>0.025</td>
</tr>
<tr>
<td>Smoking</td>
<td>950 (20%)</td>
<td>190 (6.5%)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>History of myocardial infarction</td>
<td>493 (10%)</td>
<td>245 (8.4%)</td>
<td>0.006</td>
</tr>
<tr>
<td>History of CABG</td>
<td>334 (7.0%)</td>
<td>143 (4.9%)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>History of PCI</td>
<td>750 (16%)</td>
<td>420 (14%)</td>
<td>0.13</td>
</tr>
<tr>
<td>History of HF</td>
<td>298 (6.3%)</td>
<td>127 (4.4%)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>History of HFrEF</td>
<td>134 (2.8%)</td>
<td>56 (1.9%)</td>
<td>0.018</td>
</tr>
<tr>
<td>History of HfPEF</td>
<td>164 (3.4%)</td>
<td>71 (2.4%)</td>
<td>0.016</td>
</tr>
<tr>
<td>Exercise stress</td>
<td>1,733 (36%)</td>
<td>1,848 (63%)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>HF hospitalization on follow-up</td>
<td>297 (6.2%)</td>
<td>173 (5.9%)</td>
<td>0.6</td>
</tr>
<tr>
<td>HFrEF hospitalization</td>
<td>103 (2.3%)</td>
<td>68 (2.4%)</td>
<td>0.7</td>
</tr>
<tr>
<td>HfPEF hospitalization</td>
<td>194 (4.2%)</td>
<td>105 (3.7%)</td>
<td>0.3</td>
</tr>
</tbody>
</table>

(CABG = coronary artery bypass grafting; PCI = percutaneous coronary intervention.
Categoric variables are shown as numbers and percentages; continuous variables are shown as median values and IQR.)

(Supplemental Fig. 1). Shapley additive explanations for the final AI model in the external testing population are shown in Supplemental Figure 2.

Internal Testing
In the internal testing group, HF hospitalization occurred in 297 patients during a median follow-up of 1.9 y (IQR, 1.1–2.8 y; 103 hospitalizations for HFrEF and 194 hospitalizations for HfPEF exacerbation; time to HF hospitalization, 1.0 y [IQR, 0.3–1.8 y]). The prediction performance for HF hospitalization of the final AI model (area under the receiver-operating-characteristic curve [AUC], 0.87 ± 0.03) was significantly higher than stress total perfusion deficit (AUC, 0.69 ± 0.04; P < 0.0001), stress LVEF (AUC, 0.73 ± 0.05; P < 0.0001), or clinical AI model (AUC, 0.81 ± 0.04; P < 0.0001) (Fig. 2). Additionally, a higher AUC was observed with the final AI model than with summed stress score (AUC, 0.67 ± 0.05; P < 0.001), the model with clinical and stress parameters (AUC, 0.83 ± 0.03; P < 0.001), or the model with clinical and stress parameters with coronary calcifications (AUC, 0.84 ± 0.03; P < 0.001) (Supplemental Fig. 3). Figures 3 and 4 demonstrate individualized AI risk prediction in 2 individual patients with or without subsequent HF exacerbation.

External Validation
In the external group, HF hospitalization occurred in 173 patients during a median follow-up of 2.7 y (IQR, 1.6–4.0 y; 68 hospitalizations for HFrEF and 105 hospitalizations for HfPEF; time to HF hospitalization, 0.9 y [IQR, 0.3–1.8 y]). The prediction performance for HF hospitalization of the final AI model (AUC, 0.80 ± 0.04) was significantly higher than for stress total perfusion deficit (AUC, 0.66 ± 0.06; P < 0.0001), stress LVEF (AUC, 0.70 ± 0.06; P < 0.0001), or the clinical AI model (AUC, 0.72 ± 0.05; P < 0.0001) (Fig. 5). Additionally, the AUC for the final AI model was higher than for summed stress score (AUC, 0.66 ± 0.06; P < 0.0001), the AI model with clinical and stress parameters (AUC, 0.75 ± 0.04; P < 0.001), or the AI model with clinical and stress parameters with coronary calcifications (AUC, 0.77 ± 0.05; P < 0.001) (Supplemental Fig. 4).

Reduced-Features Comparison
An additional model was evaluated that used only the 10 most important features from the final AI model. For 10-fold cross validation in the internal cohort, the final AI model was significantly better than the top-10-features model (AUC, 0.87 [IQR, 0.84–0.89] vs. 0.85 [IQR, 0.82–0.88]; P < 0.01). In external validation, the final AI model also had higher performance, but the difference was not statistically significant (AUC, 0.80 [IQR, 0.76–0.85] vs. 0.79 [IQR, 0.75–0.84]; P = 0.14) (Supplemental Fig. 5).

DISCUSSION
This study represents the first evidence, to our knowledge, demonstrating that integrating SPECT MPI into an AI-based risk assessment algorithm significantly improves the prediction of hospitalizations due to HF, when compared with relying solely on standard clinical parameters. The results show that including results from SPECT/CT MPI, such as LVEF, myocardial perfusion, stress test parameters, and coronary calcifications, significantly improves the identification of individuals at the highest risk of HF-related hospitalization. As SPECT MPI is the most frequently used imaging modality for the diagnosis of CAD, this integration not only enhances the accuracy of predicting HF-related hospitalizations in these patients but also highlights the potential...
of combining advanced imaging techniques with AI in HF risk assessment.

HF is a widespread pandemic associated with significant morbidity and mortality and high economic burden (1). Prediction of HF exacerbation has been previously demonstrated using serum biomarkers, such as troponin and brain natriuretic peptide (21). Early studies using transthoracic echocardiography demonstrated that reduced LVEF or diastolic dysfunction are predictive of subsequent HF admission (22,23). Interestingly, in our study LVEF was not among the most important variables, as might be explained by the relatively higher rate of HFpEF hospitalizations in our cohort and the inclusion of stress left ventricular end-systolic volume and prior HFrEF diagnosis in our model. The 5 highest-listed variables have been associated with increased risk for HF hospitalization (24–26). Importantly, previous studies focused mainly on individual variables and did not integrate clinical and imaging variables together to enhance HF prediction. Our study builds on previous research and suggests that an AI algorithm can identify, among patients undergoing SPECT/CT MPI, those at the highest risk of HF exacerbation.

CAC scoring by noncontrast CT can accurately estimate CAC burden within the coronary arteries, which can serve as a surrogate for CAD (27,28). During hybrid SPECT/CT and PET MPI, the low-dose CTAC may be used for quantification of CAC, with excellent correlation with standard Agatston scores (17,29–31). To our knowledge, this study was the first to incorporate CTAC calcium scores for the prediction of HF hospitalization.

Multiple AI models have been developed recently for the prediction of HF hospitalization (7,32,33). A prospective study including leg bioimpedance, age, sex and self-reported myocardial infarction provided a highly accurate AI prediction model for HF exacerbations (7). A recent study integrating echocardiographic and clinical parameters in an AI model also showed improved prediction for HF exacerbation in comparison to the Framingham HF risk model (AUC, 0.75 vs. 0.67; \( P < 0.001 \)) in patients with atrial fibrillation (32). AI models have also been shown to improve the prediction of readmission for HF in comparison to standard logistic regression (33). Our study builds on this line of prior research and, to our knowledge, was the first to incorporate SPECT MPI findings in an AI model for the prediction of HF hospitalization.

The proposed AI-based approach has substantial clinical implications, as it can allow health care providers to identify high-risk individuals among those who undergo SPECT MPI (one of the most commonly performed imaging tests). We propose that this tool could offer additional information about HF risk in individuals who have already undergone SPECT MPI without the need for further testing. The identification of patients at high risk for HF exacerbation can facilitate early interventions, such as the initiation of cardioprotective medications or the intensification of diuresis. This could also include invasive hemodynamic monitoring (34) and targeted prescribing of medications shown to reduce HF hospitalizations in high-risk populations (35). The use of an AI-based risk assessment algorithm allows for a personalized approach to managing patients who are at risk for HF by guiding providers in developing individualized management plans including closer monitoring. By identifying high-risk patients for HF hospitalization, an AI-based accurate risk assessment can facilitate directing health care resources toward targeted interventions.
Ultimately, this could reduce health care costs by reducing HF hospitalizations.

Although our study included a considerable number of patients, it was a retrospective study that comes with inherent limitations. In this patient population, no data were available to test our model against standard HF prediction tools such as echocardiographic parameters or brain natriuretic peptide, and we did not compare the performance of our AI model against estimated HF hospitalization risk based on expert physician opinion. It is important to note that our study used high-spatial-resolution cadmium zinc telluride solid-state detector systems, whereas conventional SPECT systems, which are more frequently used, have limited spatial and temporal resolution. As a result, the generalizability of our findings may be limited.

As would be expected, there were significant differences between the internal and external testing populations. Although this leads to lower performance in the external testing population, it provides a more realistic and generalizable estimate of performance in new centers. Further studies are needed to evaluate whether therapeutic interventions guided by our AI predictions can effectively modify the risk of future HF hospitalizations in patients undergoing SPECT/CT MPI.

CONCLUSION

Our study provides evidence that AI can effectively predict hospitalization for acute HF exacerbation in patients undergoing SPECT/CT MPI. Through integration of clinical, stress, nuclear imaging, and CTAC CAC scoring parameters, our model outperformed traditional measures such as stress LVEF, stress total perfusion deficit, and summed stress score. These findings have significant implications for improving risk stratification and facilitating early interventions to prevent HF hospitalizations, with the ultimate goal of enhancing patient care and outcomes.

DISCLOSURE

This research was supported in part by grants R01HL089765 and R35HL161195 from the National Heart, Lung, and Blood Institute/National Institutes of Health (principal investigator, Piotr Slomka). The content is solely the responsibility of the authors and does not necessarily represent

FIGURE 3. Individualized AI risk prediction in patient with subsequent HF exacerbation. (A) For this 47-y-old man without prior history of HF, AI model using only clinical variables scored patient as low risk for HF exacerbation (top), whereas final AI model score including all clinical, stress, and imaging variables classified patient as high risk (bottom). (B) MPI demonstrated apical scar, dynamic images demonstrated severely reduced LVEF, and attenuation CT showed severe diffuse coronary calcifications (estimated calcium score, 1,221). Patient experienced acute HF exacerbation 346 d after MPI. BMI = body mass index; BP = blood pressure; ECG = electrocardiograph; PCI = percutaneous coronary intervention.

FIGURE 4. Individualized AI risk prediction in patient without subsequent HF exacerbation. (A) For this 73-y-old man without prior history of HF, AI model using only clinical variables scored patient as high risk for HF hospitalization (top), whereas final AI model classified patient as low risk (bottom). (B) MPI showed normal perfusion, dynamic images demonstrated normal LVEF, and attenuation CT showed no coronary calcifications. Patient survived 2 y without hospitalization for HF exacerbation. BMI = body mass index; BP = blood pressure; ECG = electrocardiograph; LV = left ventricle.
FIGURE 5. Receiver-operating-characteristic curves for HF hospitalization prediction (external cohort). TPD = total perfusion deficit.

the official views of the National Institutes of Health. Piotr Slomka participates in software royalties for QPS software at Cedars-Sinai Medical Center and received research grant support from Siemens Medical Systems. Robert Miller has received grant support and consulting fees from Pfizer. Edward Miller has received grant support from and is a consultant for GE Healthcare. No other potential conflict of interest relevant to this article was reported.

KEY POINTS

QUESTION: This study aimed to assess whether AI models incorporating clinical risk factors, stress variables, SPECT imaging parameters, and fully automated deep learning–generated calcium scores from attenuation CT scans could predict hospitalization for acute HF exacerbation in patients undergoing SPECT/CT MPI.

PERTINENT FINDINGS: The final AI-based HF risk prediction model developed in the internal cohort (4,766 patients) demonstrated a higher AUC for predicting HF admissions than did stress LV EF or a model developed using only clinical parameters. These findings were confirmed in the external validation cohort (2,912 patients).

IMPLICATIONS FOR PATIENT CARE: The proposed AI-based risk assessment algorithm has significant implications for improving risk stratification and facilitating early interventions to prevent HF hospitalizations, with the ultimate goal of enhancing patient care and outcomes.

REFERENCES


