

ARTICLE OPEN



Artificial intelligence for the diagnosis of heart failure

Dong-Ju Choi^{1,3}, Jin Joo Park^{1,3}, Taqdir Ali² and Sungyoung Lee²

The diagnosis of heart failure can be difficult, even for heart failure specialists. Artificial Intelligence-Clinical Decision Support System (AI-CDSS) has the potential to assist physicians in heart failure diagnosis. The aim of this work was to evaluate the diagnostic accuracy of an AI-CDSS for heart failure. AI-CDSS for cardiology was developed with a hybrid (expert-driven and machine-learning-driven) approach of knowledge acquisition to evolve the knowledge base with heart failure diagnosis. A retrospective cohort of 1198 patients with and without heart failure was used for the development of AI-CDSS (training dataset, $n = 600$) and to test the performance (test dataset, $n = 598$). A prospective clinical pilot study of 97 patients with dyspnea was used to assess the diagnostic accuracy of AI-CDSS compared with that of non-heart failure specialists. The concordance rate between AI-CDSS and heart failure specialists was evaluated. In retrospective cohort, the concordance rate was 98.3% in the test dataset. The concordance rate for patients with heart failure with reduced ejection fraction, heart failure with mid-range ejection fraction, heart failure with preserved ejection fraction, and no heart failure was 100%, 100%, 99.6%, and 91.7%, respectively. In a prospective pilot study of 97 patients presenting with dyspnea to the outpatient clinic, 44% had heart failure. The concordance rate between AI-CDSS and heart failure specialists was 98%, whereas that between non-heart failure specialists and heart failure specialists was 76%. In conclusion, AI-CDSS showed a high diagnostic accuracy for heart failure. Therefore, AI-CDSS may be useful for the diagnosis of heart failure, especially when heart failure specialists are not available.

npj Digital Medicine (2020)3:54; <https://doi.org/10.1038/s41746-020-0261-3>

INTRODUCTION

The prevalence of heart failure (HF) has been increasing^{1,2}. HF is associated with high morbidity and mortality³. Because HF is a complex syndrome that can result from structural and functional cardiac disorder, rather than a single disease entity, its correct diagnosis can be challenging even for HF specialists. Currently, HF is classified according to ejection fraction, i.e., HF with reduced ejection fraction (HF_rEF), HF with mid-range ejection fraction (HF_{mr}EF), and HF with preserved ejection fraction (HF_pEF)⁴. A correct diagnosis is mandatory before proper treatment can be initiated^{4,5}. Furthermore, present-day physicians are challenged by rapidly changing scientific evidences, new drugs, and the complexity of guidelines for HF management, especially in outpatient clinic. With enormous advancements in information and communication technologies, such as easy storage, acquisition, and recovery of big data and knowledge, artificial intelligence (AI) has been gaining an important role in cardiology⁶.

Two types of AI decision systems are available: a white-box-based and a black-box-based one. A white-box AI-based decision system involves correlations and transparency among rules for the analysis of accumulated data, and is mainly constructed using supervised algorithms such as decision tree algorithm⁷. On the contrary, a black-box-based AI has opaque algorithms and its process and reasoning applied in providing the respective conclusions are difficult to clarify. IBM Watson for Oncology (WFO) is black-box AI-based decision systems. WFO demonstrated a concordance rate of 93% for the treatment recommendation in breast cancer. However, WFO cannot disclose the recommendation processes for the final clinical decision⁸.

Clinical Decision Support System (CDSS) is a health information technology that assists physicians in clinical decision making. The concept of computer-based clinical decision has been developed for informatics six decades ago⁹. In spite of the enthusiasm for

evolving CDSS which is assisted with the potential of AI, the realities and complexities of real clinical practice limit the rapid evolution of CDSS. An effective CDSS requires CDSS to match the individual patient's characteristics to the clinical knowledge base, provides patient-centric assessments and recommendations, and finally presents recommendations in white-box manner to the physicians for their final decision⁸.

We conducted a study assessing the level of agreement with respect to the HF diagnosis to identify the three types of HF, i.e. HF_rEF, HF_{mr}EF, and HF_pEF, between HF specialists and AI-CDSS at a tertiary center in Korea. First, we created an AI-CDSS using a hybrid approach of expert-driven knowledge acquisition and ML-driven rule generation. Second, we evaluated the diagnosis concordance (degree of agreement) of AI-CDSS in a test set of patients with and without HF as a pilot clinical study. Third, we prospectively tested the diagnostic performance of AI-CDSS in consecutive patients presenting with dyspnea to the outpatient clinic.

RESULTS

Development of cardiovascular AI-CDSS

Using the training dataset of 600 patients with and without HF, the AI-CDSS was created using predefined steps including expert-driven knowledge acquisition, machine learning (ML)-driven rule generation, and hybridization of both types of knowledge.

Expert-driven knowledge acquisition. In the knowledge modeling phase, the clinical recommendations of diagnosis were first transformed into mind maps and then transformed to a decision tree. The decision tree was evaluated and modified by the physicians until a consensus was achieved. The final decision tree was termed as R-CKM (Supplementary Fig. 1) and included 14

¹Division of Cardiology, Department of Internal Medicine, Seoul National University Bundang Hospital, Seongnam, Republic of Korea. ²Department of Computer Science and Engineering, Kyung Hee University, Yongin, Republic of Korea. ³These authors contributed equally: Dong-Ju Choi, Jin Joo Park. ✉email: djchoi@snuh.org; sylee@oslab.khu.ac.kr

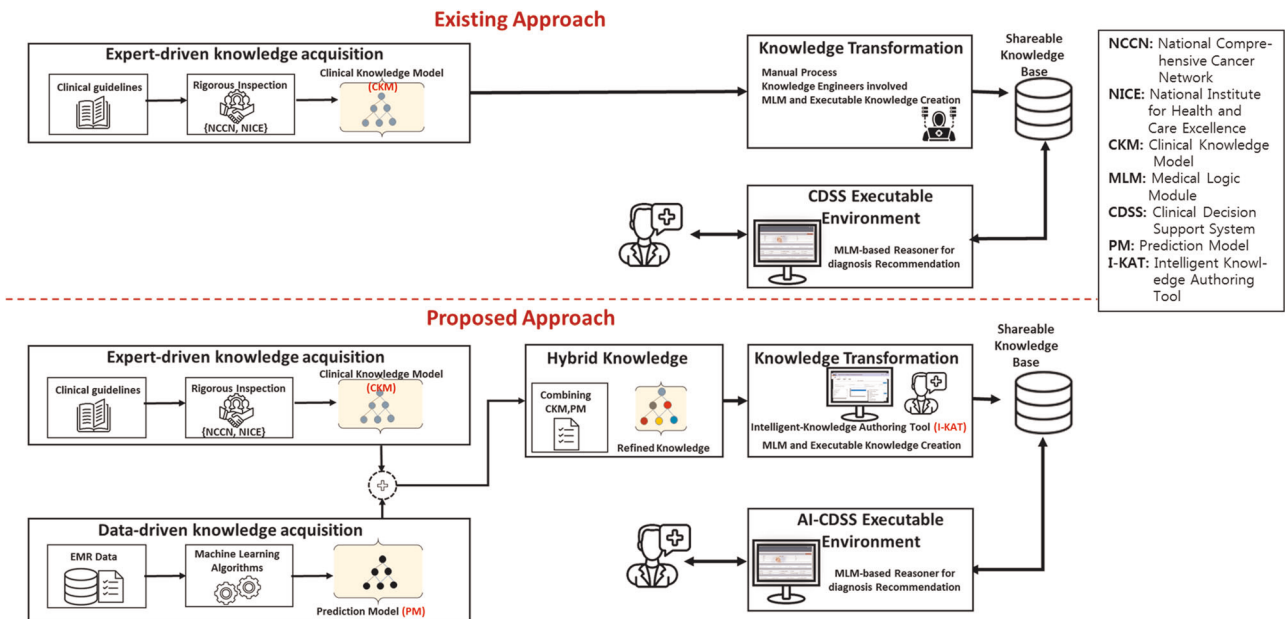


Fig. 1 Comparison of existing CDSSs and our proposed artificial intelligence-CDSS. CDSS Clinical Decision Support System, CKM clinical knowledge model, I-KAT Intelligent Knowledge Authoring Tool, NCCN National Comprehensive Cancer Network, NICE National Institute for Health and Care Excellence, PM prediction model.

contributing factors (Supplementary Table 1) and 4 possible outcomes: HF_rEF, HF_mrEF, HF_pEF, and no-HF.

ML-driven rule generation. We used five machine learning algorithms i.e. Decision Tree (DT), Random Forest, Chi-squared Automatic Interaction Detection (CHAID), J48, and Classification and Regression Tree (CART). All algorithms selected only few features such as left ventricular ejection fraction (LVEF), left atrial volume index (LAVI), and left ventricular mass index (LVMI) as highly contributing factors (Supplementary Table 2). To boost the model performance, the auto-feature selection method was used and LVEF, electrocardiography, LVMI, and LAVI were selected as the most significant features (Supplementary Fig. 2).

The five algorithms showed different accuracy (Supplementary Table 3). We also calculated the rank of each algorithm based on the accuracy, number of rules extracted, and number of attributes involved, using the rank formula developed in our previous work¹⁰. Finally, CART algorithm was selected to create ML-driven knowledge, because it showed the highest accuracy and rank of 88.5 and 0.5736, respectively. The CART algorithm mainly focused on features of LVEF, LAVI, and tricuspid regurgitation velocity (Supplementary Fig. 3). The algorithm correctly predicted HF_mrEF and HF_rEF with 100% accuracy, whereas HF_pEF and no-HF were predicted with 78.9% and 80.5% accuracy, respectively.

Hybrid knowledge. The merging of the CKM from the expert-driven knowledge and the PM from the ML-driven knowledge approach led to the final hybrid knowledge in form of R-CKM (Fig. 1, Supplementary Materials). Sometimes, physician may miss some of the attributes or path of attributes during development of CKM, and the ML generated PM finds the missing attributes or paths. For instance, the CKM is starting with the “Sign & Symptoms” as shown in (Supplementary Fig. 3), while the PM starts checking from “LVEF” as shown in (Supplementary Fig. 4). Therefore, the hybridization algorithm recognizes that the CKM is missing a path of “Not Available” values between “Sign & Symptoms” and “LVEF” attributes. When we added this new path into CKM, the number of knowledge base rules increased drastically. The addition of new

path into R-CKM increases the coverage of patient cases to generate right recommendations and increase the accuracy.

Validation of AI-CDSS

Study population. The test dataset included 598 patients (490 patients with HF, 108 patients without HF). Patients with HF were older (73.1 ± 13.8 years vs. 64.8 ± 13.8 years, $P < 0.001$), more likely to be male (52% vs. 37%, $P = 0.005$), and had higher N-terminal pro-brain natriuretic peptide levels ($10,075 \pm 11,778$ pg/L vs. 82 ± 68 pg/L, $P < 0.001$). Concerning the echocardiographic parameters, patients with HF had lower LVEF ($45.5 \pm 17.4\%$ vs. $64.1 \pm 6.5\%$, $P < 0.001$), higher LAVI (53.9 ± 21.1 ml/m² vs. 31.2 ± 8.5 ml/m², $P < 0.001$), and higher E/e' (18.6 ± 9.8 vs. 9.8 ± 3.5 , $P < 0.001$) (Table 1). Among patients with HF, 199 (40.6%), 63 (12.9%), and 228 (46.5%) were classified as having HF_rEF, HF_mrEF, and HF_pEF, respectively.

Diagnostic accuracy. The results of comparative analysis are shown in Fig. 2. The concordance rate was 100% in HF_rEF and HF_mrEF for all three approaches. With respect to HF_pEF, the concordance rate was 82%, 79%, and 99.5% for expert-driven, ML-driven, and hybrid CDSS, respectively. Similar findings were observed for no-HF. The overall diagnostic accuracy was 90%, 88.5%, and 98.3% for expert-driven, ML-driven, and hybrid CDSS, respectively, showing a remarkable increase in accuracy by 8% with the hybrid approach, i.e., AI-CDSS.

The expert-driven approach had a sensitivity and a specificity of 0.96 and 0.71, respectively (Supplementary Table 4), whereas the ML-driven approach had a sensitivity and a specificity of 0.72 and 0.94, respectively (Supplementary Table 5). Strikingly, the hybrid approach had a sensitivity and specificity of 0.94 and 0.99, respectively (Supplementary Table 6).

Subgroup analysis. We divided the patients according to echocardiographic parameters. Set A included all echocardiography parameters, whereas set B included only LVEF, LAVI, and LVMI. The concordance rate was lower in set B than in set A (Supplementary Fig. 4). In our study, the age of the included patients ranged from 20 to 92 years. Age did not affect the accuracy of the system (Supplementary Table 7).

Table 1. Characteristics of the study population retrospective patients ($n = 598$).

	No heart failure	Heart failure				P value*
		All	HFrEF	HFmrEF	HFpEF	
Age (years)	64.8 ± 13.8	73.1 ± 13.8	70.3 ± 14.6	74.7 ± 14.1	75.2 ± 10.6	<0.001
Male (%)	37	52	54.3	52.4	50.0	0.005
HF symptoms, signs (%)	81.5	89.8	94.0	84.1	87.7	0.015
Clinical history (%)	14.8	51.6	66.3	55.6	37.7	<0.001
Physical exam (%)	9.3	51.4	60.8	49.2	43.9	<0.001
Abnormal ECG (%)	46.0	93	99.5	96.8	86.2	<0.001
NT-pro-BNP (pg/L)	82.4 ± 68.0	10075 ± 11778	15665 ± 12604	8634 ± 9666	5595 ± 9306	<0.001
Echocardiography						
LVEF (%)	64.1 ± 6.5	45.5 ± 17.4	27.1 ± 7.5	45.3 ± 2.6	61.6 ± 6.5	<0.001
LAVI (mL/m ²)	31.2 ± 8.5	53.9 ± 21.1	60.5 ± 18.6	52.6 ± 27.5	48.0 ± 19.4	<0.001
LVMI (mg/m ²)	83.4 ± 18.3	127.3 ± 44.7	151.0 ± 41.5	129.0 ± 50.7	106.0 ± 33.9	<0.001
E/e	9.8 ± 3.5	18.6 ± 9.8	22.9 ± 10.34	17.4 ± 8.6	15.6 ± 8.3	<0.001
Septal e' (cm/s)	6.9 ± 2.4	5.0 ± 2.1	4.2 ± 1.7	5.1 ± 2.3	5.6 ± 2.2	<0.001
TRV (m/s)	2.6 ± 1.5	2.9 ± 0.7	3.0 ± 0.7	2.8 ± 0.5	2.8 ± 0.7	0.001
GLS ($n = 324$) (%)	16.4 ± 3.9	10.8 ± 5.0	7.1 ± 2.7	10.4 ± 2.8	14.6 ± 4.3	<0.001

*P value between no heart failure and heart failure.

ECG electrocardiography, GLS global longitudinal strain, HF heart failure, LAVI left atrial volume index, LVEF left ventricular ejection fraction, NT-proBNP N-terminal pro-B-type natriuretic peptide, TRV tricuspid regurgitation velocity.

Diagnostic Accuracy of AI-CDSS

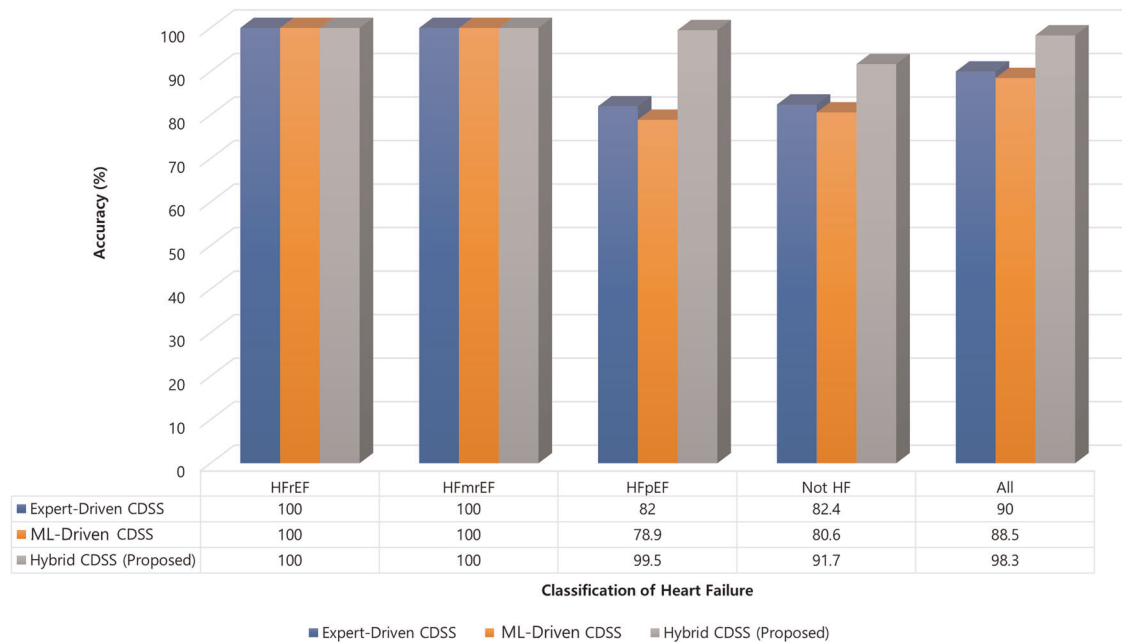


Fig. 2 Comparative analysis of the diagnostic accuracy of different approaches in the retrospective cohort. CDSS Clinical Decision Support System, HFmrEF heart failure with mid-range ejection fraction, HFpEF heart failure with preserved ejection fraction, HFrEF heart failure with reduced ejection fraction.

Accuracy of AI-CDSS in a prospective cohort of patients with dyspnea

A total of 100 consecutive patients who presented with dyspnea to the outpatient clinic were enrolled. Of these, the data of three patients were not complete; thus, the data of 97 patients were used in the final analysis. Of the 97 patients, 43 (44%) had HF. In this prospective cohort, the concordance rate of the non-HF specialists was 76%, whereas that of AI-CDSS was 98% (Fig. 3).

Especially, the diagnosis of HFmrEF and HFpEF was low among the non-HF specialist, whereas the diagnosis of no-HF was comparably high.

DISCUSSION

Correct diagnosis of HF can be challenging for physicians, even for HF specialists. In this study, we first created AI-CDSS by using the

Diagnostic accuracy of physicians and AI-CDSS

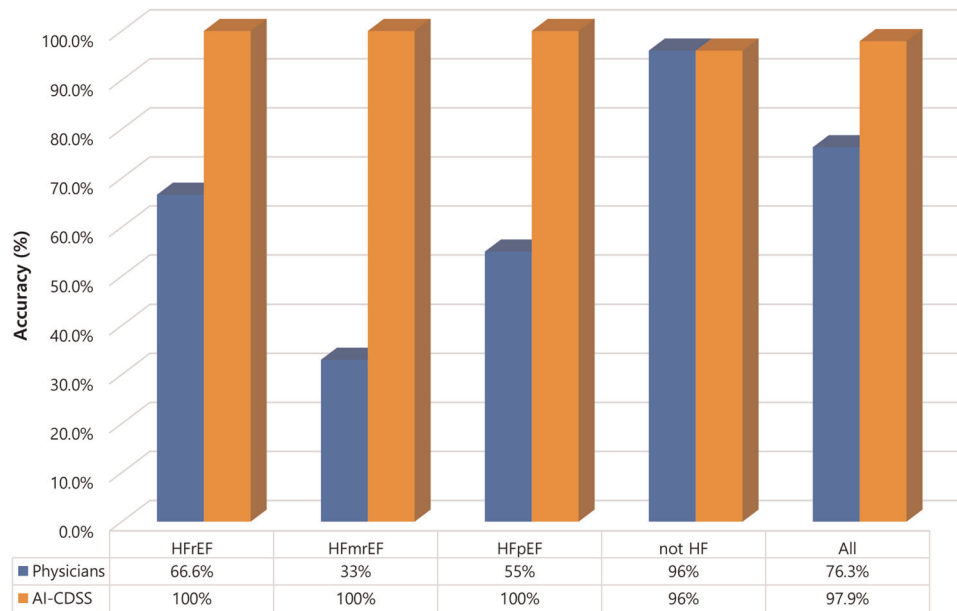


Fig. 3 Comparative analysis of the diagnostic accuracy of physicians and AI-CDSS in the prospective cohort. Abbreviations are as in Fig. 2.

data of 1198 patients with and without HF and showed that AI-CDSS had very high diagnostic accuracy in these patients. In a prospective cohort of patients presenting with dyspnea to the outpatient clinic, AI-CDSS consistently showed a remarkably high diagnostic accuracy. By contrast, non-HF specialists showed a relatively low diagnostic accuracy for HF. Therefore, AI-CDSS may be useful for the diagnosis of HF, especially when HF specialists are not available.

CDSS has been applied in clinical diagnosis¹¹, preventive care¹², and chronic disease management¹³, among others. It provides a proficient decision-making service to improve the quality of healthcare, but has several complexities and limitations⁸. Generally, AI-CDSS acquires knowledge from structured and unstructured data using ML and natural language processing techniques^{14,15}. The amalgamation of ML-driven rule generation with expert-driven knowledge acquisition enhances the system accuracy¹⁰. Therefore, we chose the hybrid approach for the knowledge acquisition of AI-CDSS, which includes three distinct steps: expert-driven knowledge acquisition, ML-driven rule generation, and hybridization of both types of knowledge.

In expert-driven knowledge acquisition, we first built CKM by transforming expert-driven knowledge into a mind map and decision tree¹⁰. Finally, the decision tree was validated with the PM of ML-driven knowledge.

In ML-driven knowledge acquisition, we created a PM with an available big dataset. Various ML algorithms can be used to analyze and extract the hidden patterns in the form of knowledge models. In our study, we used white box AI and causal machine algorithms such as decision tree, random forest, CHAID (Chi-squared Automatic Interaction Detector), J48, and CART/CRT. White-box model AI are selected for their transparency, which enables easily determining all attributes for classification and verification of new patient data¹⁶. They also increase the physicians' satisfaction level, because the rationale for a decision is also provided to the physician using the features contributing to the final decision. In addition, the computational complexity of the decision tree constructing algorithms (white box) is relatively low. By contrast, black box algorithms have no transparency in

knowledge modeling owing to difficulty in interpreting the inner working layers of the models.

In the hybridization of the both knowledge types, we validated the PM from the ML-driven approach against the mind map of CKM from the expert-driven approach to produce the final hybrid knowledge in form of R-CKM.

Finally, we developed a web-based application in the form of cardiovascular AI-CDSS for use of physicians in real clinical practice. For this purpose, the R-CKM knowledge was transformed into MLM for knowledge shareability and computer-executable format using I-KAT, which had been developed by our team⁷. The resultant knowledge can be easily shared and integrated into various formats of HF diagnosis systems, because the resultant knowledge was built with consolidation of the standard data model vMR (Virtual Medical Record) and the standard terminology SNOMED CT (Systematized Nomenclature of Medicine—Clinical Terms).

Because HF is a syndrome with various clinical features, its diagnosis can be very challenging even for HF specialists. In patients with HF, pulmonary congestion can develop because of congestion in the left heart, causing dyspnea. However, dyspnea as a symptom can also arise from lung disease, anemia, and mental disorders¹⁷. Leg swelling is a typical sign of congestion in the right heart. However, it also has many differential diagnoses, including kidney disease, adverse effect of drugs, and chronic venous insufficiency, among others¹⁸. In clinical practice, many patients are diagnosed as having HF even if they do not have HF, and vice versa. A correct diagnosis of HF is crucial because patients with HF have a grave prognosis that is comparable to that of oncologic malignancies¹⁹, and there exist therapy that can improve survival in patients with HF^{4,6}. Consequently, misdiagnosis of HF can hinder the chance of improving the outcomes. AI-CDSS is a tool that helps in making better medical decisions, thereby reducing clinical errors and improving the quality of life. It has the potential to generate alerts and reminders, diagnostic assistance, therapy critiquing and planning, and image recognition and interpretation.

Currently, HF is classified according to LVEF into HFrEF, HFmrEF, and HFpEF. With respect to HFrEF, a decrease in LVEF may alert physicians to the possible diagnosis of HF. By contrast, for HFpEF >50%, the normal systolic function may “blind” the physicians and HFpEF may remain undiagnosed. We showed that AI-CDSS showed acceptably high concordance for diagnosing HF regardless of type, whereas non-HF specialists misdiagnosed HFpEF in almost half of the patients.

In medicine, IBM WFO demonstrated high concordance with oncologists in treatment recommendations¹⁴. In the field of cardiology, our study presents the clinical feasibility of AI for diagnosing HF.

There are several limitations in this study. The intervention of the physicians is crucial in knowledge creation and validation. However, the level of expertise varies from physician to physician, so that the CKM developed by physicians in a hospital may differ from that developed in another hospital. Similarly, because the attributes in the PM depend on the patient data used, they may also differ from variables recommended in the guidelines. Therefore, further studies are necessary to validate the AI-CDSS in other study populations.

In conclusions, AI-CDSS showed high diagnostic accuracy for HF, independent of HF types. Therefore, AI-CDSS may be useful for the diagnosis of HF, especially when HF specialists are not available.

METHODS

Study population and data collection

Retrospective cohort. We included 1198 patients with and without HF from January 2016 to December 2017. We divided the patients into two datasets. The first 600 patients were used for the generation of AI-CDSS as a training dataset for ML, whereas the remaining 598 patients were used for the validation of the AI-CDSS as the test dataset. HF was defined as present when patients had symptoms (dyspnea, orthopnea) or signs of HF (rales, pitting edema, ascites) and met one of the following criteria: lung congestion, objective findings of left ventricular systolic dysfunction, or structural heart disease. Clinical information including demographics, symptoms, signs, medical history, laboratory examination, electrocardiography, and echocardiography was obtained. Control patients, i.e., those without HF, were randomly selected from the electronic medical records.

Prospective pilot cohort. For an additional validation of AI-CDSS, we enrolled 100 consecutive patients presenting with dyspnea to the outpatient clinic. The treating physicians performed history taking and physical examination, ordered diagnostic tests, and made a final diagnosis, i.e., HF or no-HF, according to their clinical judgment. The data were first recorded in an electronic clinical research form, and then automatically transferred to the AI-CDSS. A direct extraction of patients' data from electronic medical record was circumvented because of the information security regulation at our institution.

The study protocol was approved by the institutional review board of the Seoul National University Bundang Hospital. For the retrospective cohort, the requirement for written informed consent was waived by the institutional review board. Each patient in the prospective cohort provided informed consent before study enrollment. The study complied with the Declaration of Helsinki.

Echocardiography

All images were obtained using a standard ultrasound machine with a 2.5-MHz probe. Standard techniques were used to obtain M-mode, two-dimensional, and Doppler measurements in accordance with the American Society of Echocardiography guidelines²⁰. Tissue-Doppler-derived peak systolic, early, and late diastolic velocities of the septal mitral annulus were recorded. Left ventricular end-systolic and end-diastolic volumes were measured from apical four- and two-chamber views and LVEF was calculated using Simpson's biplane method.

Generation of cardiovascular AI-CDSS

The traditional CDSSs usually focus on the expert-driven approach with collaboration between physicians and knowledge engineers, where the

knowledge engineer is an expert in AI language who investigate the underlying problems, develop the main concepts, and efficiently represent the knowledge in the domain. The fundamental knowledge resource is the clinical practice guidelines and physicians' expertise. The AI-CDSS uses patient data as the second important resource of knowledge after processing with ML algorithm (ML-driven approach). Figure 1 shows the difference between the traditional CDSSs and the cardiovascular AI-CDSS. The existing CDSSs maintain the knowledge base by the knowledge engineers. In contrast, AI-CDSS focuses on the hybrid approach of expert-driven knowledge acquisition and ML-driven rule generation and overcomes the physicians' dependency on knowledge engineers. In AI-CDSS, the clinical knowledge model (CKM), a classical top-down decision tree, is generated by domain expert (physician) using guidelines and their experiences; it is called the Expert-Driven Knowledge. The second step is to create ML-based prediction model (PM) using several ML algorithms, which is called the machine learning (ML)-driven knowledge. The third step is to generate the refined-CKM (R-CKM) by the computer scientists using a quick, simple, and iterative agile software development. The R-CKM generation is composed of making prediction model of ML-driven knowledge and validation of expert-driven knowledge with respective ML-driven knowledge using several ML algorithms with training dataset of 600 patients' data; it is called the hybrid knowledge. Finally, the R-CKM knowledge is transformed into shareable and interoperable setting in the form of Health Level-7 (HL7) complaint standard knowledge representation, termed Medical Logic Module (MLM), using the Intelligent Knowledge Authoring Tool (I-KAT) developed by our group. The executable MLM in the shareable knowledge base is executed to generate decisions based on the patients' input to assist the physicians. More details of cardiovascular AI-CDSS are explained in the Supplementary methods.

Study variables

HF was defined when patients had signs or symptoms of HF and either lung congestion, objective findings of LV systolic dysfunction, or structural heart disease. The diagnosis of HF was confirmed by two independent HF specialists who had >10 years of clinical experience. The diagnosis by the experts was considered the gold standard.

According to the LVEF on echocardiography, patients were classified as having HFrEF (LVEF < 40%), HFmrEF (40% ≤ LVEF < 50%), and HFpEF (LVEF ≥ 50%).

The diagnostic accuracy of AI-CDSS was measured using experts' diagnosis as the gold standard. Concordance was defined as present when experts and AI-CDSS had the same diagnosis, i.e., both HF or both no-HF. Discordance was defined to exist when there was a disagreement between diagnoses.

Statistical analysis

Descriptive statistics were calculated to determine the clinical characteristics and outcomes of the registry population. Data were presented as numbers and frequencies for categorical variables and as mean ± standard deviation or median with interquartile range for continuous variables. For the comparison between groups, the χ^2 test (or Fisher's exact test when any expected cell count was <5 for a 2 × 2 table) was used for categorical variables, whereas unpaired Student's t-test was used for continuous variables. Concordance was expressed as the percentage agreement. Pearson's correlation was used to calculate the association between expert opinion and AI-CDSS judgment.

A two-sided *P* value of <0.05 was considered statistically significant. Statistical tests were performed using IBM SPSS Statistics version 23 (SPSS Inc., Chicago, IL, USA).

Role of the funding source

The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Reporting summary

Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

DATA AVAILABILITY

The dataset generated during the current study is not publicly available due to restrictions in the ethical permit, but may be available from the corresponding author on reasonable request.

CODE AVAILABILITY

We have made all source code available at <https://github.com/ubiquitous-computing-lab/AI-CDSS-Cardiovascular-Silo>.

Received: 10 December 2019; Accepted: 12 March 2020;

Published online: 08 April 2020

REFERENCES

- Redfield, M. M. et al. Burden of systolic and diastolic ventricular dysfunction in the community: appreciating the scope of the heart failure epidemic. *JAMA* **289**, 194–202 (2003).
- Choi, D. J. et al. Characteristics, outcomes and predictors of long-term mortality for patients hospitalized for acute heart failure: a report from the Korean Heart Failure Registry. *Korean Circ. J.* **41**, 363–371 (2011).
- Bleumink, G. S. et al. Quantifying the heart failure epidemic: prevalence, incidence rate, lifetime risk and prognosis of heart failure The Rotterdam Study. *Eur. Heart J.* **25**, 1614–1619 (2004).
- Ponikowski, P. et al. ESC guidelines for the diagnosis and treatment of acute and chronic heart failure: The Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC). Developed with the special contribution of the Heart Failure Association (HFA) of the ESC. *Eur. J. Heart Fail.* **18**, 891–975 (2016).
- Kim, M. S. et al. Korean guidelines for diagnosis and management of chronic heart failure. *Korean Circ. J.* **47**, 555–643 (2017).
- Johnson, K. W. et al. Artificial intelligence in cardiology. *J. Am. Coll. Cardiol.* **71**, 2668–2679 (2018).
- Ali, T. et al. Multi-model-based interactive authoring environment for creating shareable medical knowledge. *Comput. Methods Prog. Biomed.* **150**, 41–72 (2017).
- Shortliffe, E. H. & Sepulveda, M. J. Clinical decision support in the era of artificial intelligence. *JAMA* **320**, 2199–2200 (2018).
- Ledley, R. S. & Lusted, L. B. Reasoning foundations of medical diagnosis; symbolic logic, probability, and value theory aid our understanding of how physicians reason. *Science* **130**, 9–21 (1959).
- Hussain, M. et al. Data-driven knowledge acquisition, validation, and transformation into HL7 Arden Syntax. *Artif. Intell. Med.* **92**, 51–70 (2018).
- Kline, J. A., Zeitouni, R. A., Hernandez-Nino, J. & Jones, A. E. Randomized trial of computerized quantitative pretest probability in low-risk chest pain patients: effect on safety and resource use. *Ann. Emerg. Med.* **53**, 727–735. e721 (2009).
- Kucher, N. et al. Electronic alerts to prevent venous thromboembolism among hospitalized patients. *N. Engl. J. Med.* **352**, 969–977 (2005).
- Roumie, C. L. et al. Improving blood pressure control through provider education, provider alerts, and patient education: a cluster randomized trial. *Ann. Intern. Med.* **145**, 165–175 (2006).
- Somashekhar, S. P. et al. Watson for oncology and breast cancer treatment recommendations: agreement with an expert multidisciplinary tumor board. *Ann. Oncol.: Off. J. Eur. Soc. Med. Oncol.* **29**, 418–423 (2018).
- Hinton, G. Deep learning—a technology with the potential to transform health care. *JAMA* **320**, 1101–1102 (2018).
- Romero, C., Olmo Ortiz, J. L. & Ventura, S. A meta-learning approach for recommending a subset of white-box classification algorithms for moodle datasets. *Proc. Educational Data Mining, Memphis, TN, USA* (2013).
- Berliner, D., Schneider, N., Welte, T. & Bauersachs, J. The differential diagnosis of dyspnea. *Dtsch. Arzteblatt Int.* **113**, 834–845 (2016).
- Ciocon, J. O., Fernandez, B. B. & Ciocon, D. G. Leg edema: clinical clues to the differential diagnosis. *Geriatrics* **48**, 34–40, 45 (1993).
- Stewart, S., MacIntyre, K., Hole, D. J., Capewell, S. & McMurray, J. J. More ‘malignant’ than cancer? Five-year survival following a first admission for heart failure. *Eur. J. Heart Fail.* **3**, 315–322 (2001).
- Lang, R. M. et al. Recommendations for cardiac chamber quantification by echocardiography in adults: an update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. *J. Am. Soc. Echocardiogr.* **28**, 1–39. e14 (2015).

ACKNOWLEDGEMENTS

This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2017-0-01629) supervised by the IITP (Institute for Information & Communications Technology Promotion). Figure 1 contains icons designed by Zlatko Najdenovski, Smashicons from www.flaticon.com.

AUTHOR CONTRIBUTIONS

D.-J.C. and S.L. designed the study; D.-J.C., J.J.P., T.A., and S.L. analyzed the data. J.J.P. and T.A. wrote the first draft of the manuscript which was reviewed, modified, and approved by all other authors. All the authors vouch for the accuracy and completeness of the data reported and for the fidelity of the study to the protocol.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information is available for this paper at <https://doi.org/10.1038/s41746-020-0261-3>.

Correspondence and requests for materials should be addressed to D.-J.C. or S.L.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020