Remote monitoring of single-lead electrocardiography enables detection of heart failure status

Eriko Hasumi
The University of Tokyo
Katsuhito Fujiu (fujiu-tky@umin.ac.jp)
The University of Tokyo https://orcid.org/0000-0002-1889-9291
Ying Chen
The University of Tokyo
Mitsunori Oida
The University of Tokyo
Yu Shimizu
The University of Tokyo
Kunihiro Kani
The University of Tokyo
Kohsaku Goto
The University of Tokyo
Ryoko Uchida
The University of Tokyo
Yuxiang Liu
The University of Tokyo
Tsukasa Oshima
The University of Tokyo
Jun Matsuda
The University of Tokyo
Takumi Matsubara
The University of Tokyo
Junichi Sugita
The University of Tokyo
Yukiteru Nakayama
The University of Tokyo
Hiroshi Matsunaga
The University of Tokyo Graduate School of Medicine
Akiko Saga

The University of Tokyo Graduate School of Medicine
Gaku Oguri
The University of Tokyo
Toshiya Kojima
The University of Tokyo
Yujin Maru
Nippon Medical School
Satoshi Kodera
The University of Tokyo
Hiroshi Akazawa
The University of Tokyo https://orcid.org/0000-0002-3574-9607
Morio Shoda
Tokyo Women's Medical University
Issei Komuro
The University of Tokyo Graduate School of Medicine https://orcid.org/0000-0002-0714-7182

Article

Keywords:

Posted Date: February 2nd, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1308790/v1

Abstract

Repeated hospitalization for heart failure (HF) is a strong predictor of mortality among HF patients. While recent cardiac electrical implantable devices (CIEDs) can detect worsening HF through remote monitoring[1,2], there is no early detection system for HF progression in patients at home without a CIED. We therefore developed an artificial intelligence-based HF detection system that uses single-lead electrocardiograms (ECGs) recorded at home. Our convolutional neural network (CNN) model calculated a novel HF-index from the estimated *NYHA* grades as a quantitative indicator of HF severity. Retrospective data revealed a strong correlation between HF-indexes and plasma BNP levels (R=0.91).

A prospective clinical study confirmed the accuracy of the HF severity judged from the estimated HF-index using a portable single-lead ECG monitor at home.

We have thus successfully constructed a novel, at-home HF monitoring system for a portable single-lead ECG device, which enables early detection and early medical intervention in HF.

Full Text

An estimated 64.3 million people worldwide are living with HF[3]. Moreover, that number is rising as a result of aging populations[4]. Treatments for chronic HF have advanced remarkably over the past two decades, and these interventions have improved survival and reduced the rate of HF progression. Nonetheless, most patients still require rehospitalization for decompensated HF[5].

Recent studies suggest that multiple rehospitalizations for HF are strong independent predictors of mortality[6,7], but preventing disease progression and recurrent hospital admissions in patients with established HF continues to be a challenge[8]. In particular, the incidence of HF with preserved ejection fraction (HFpEF) is rapidly becoming the most common form of HF[9]. And one in five inpatients with HFpEF are readmitted within 30 days of hospital discharge, and more than half are readmitted within one year[10].

However, diagnosing initial HF symptoms in HFpEF patients is challenging for several reasons, including the lack of specific symptoms, lower plasma BNP levels, etc.[11]. Although echocardiography is the primary tool for diagnosing HFpEF, based on left ventricular hypertrophy or signs of diastolic dysfunction[12], it is difficult for a non-expert to rule out noncardiac disorders, such as pulmonary disease, through echocardiographic estimation[13]. Therefore, novel strategies have been developed for early and easy detection of HF progression[14,15,16]. One current approach is remote monitoring of cardiac implantable electrical device (CIED)[1,17]. CIEDs provide helpful information for managing HF patients[17,18,19]. Remotely captured parameters can help predict future adverse clinical events[20,21,17] as well as all-cause mortality[22]. It was also reported that remote monitoring-based HF management reduces the risk of emergency department visits for worsening HF[23]. However, there is no method for remotely monitoring

the cardiac condition of patients without CIEDs. Therefore, new strategies for detecting HF status are needed.

Because HF is a complex syndrome, its correct diagnosis can be challenging for non-specialists[24]. Recently, convolutional neural network (CNN) models have been used to assist physicians in diagnosing HF[25,26]. One recent report showed that an AI-based clinical decision system provided high diagnostic accuracy for HF patients presenting with dyspnea at an outpatient clinic. The system used multiple clinical findings and other data, including symptoms, 12-lead ECG, N-terminal pro-B-type natriuretic peptide (NT-proBNP) level, and echocardiographic data[27]. And because the ECG features of HF patients are varied and complex for non-cardiologists, CNNs have been developed to surmount the inadequacies of manual analysis of ECG signals through the use of highly accurate systems for automatic classification of 2-s ECG signals[28,29]. These CNNs are able to discriminate the features of ECG waveforms in HF from normal ECGs, but they are unable to determine the severity of the HF. To provide adequate medical treatment and avoid HF recurrence, detailed classification of HF symptoms collected from daily life is required; simply detecting HF is not sufficient. Our aim, therefore, was to develop a CNN model for automatic classification of HF status using lead-I ECGs, which can be easily and frequently recorded in a patient's home.

Results

To construct a CNN algorithm for diagnosing HF status, we retrospectively collected 10-s, 12-lead ECGs from adult (20 years old or older) HF patients and healthy individuals. The collected data were from outpatients with or without CVD and inpatients with HF. Lead I ECG data were extracted from the 12-lead ECGs, and NYHA classifications determined by multiple cardiologists at the time of ECG recording were collected. The flow of data creation for the CNN is shown in Figure 1a. Participants implanted with a CIED or for whom the ECG recording was poor or NYHA information lacking were excluded. After the screening, 23,029 ECGs from 9,518 participants were deemed eligible for inclusion in the subsequent analysis. After preprocessing as described in the Methods, the lead-I ECG data were segmented into single heartbeats, and the averaged heartbeat waveforms from 10-s ECGs were obtained for each NYHA grade.

The CNN was designed to use single beats segmented from the lead-I ECG waveform to distinguish HF patients from healthy controls or to classify the severity of HF according to the NYHA classification. We employed hold-out validation to develop and evaluate the CNN. Among all patient-by-patient datasets, 70% were used to develop the CNN model, and the remainder were divided into two equal parts and used for validation and testing (Figure 1a). The patients whose data were used for training/validation were excluded from the testing process. The patients' characteristics are shown in Table 1.

The CNN included three convolution layers and three max-pooling layers for both training and identification. We configured the CNN to classify HF severity into three classes: healthy control, NYHA I-II (asymptomatic to mild HF) or III-IV HF (moderate to severe HF). We roughly classified HF severity into just three classes because a main aim of this model was to extract patients in an obvious decompensated

state using a single-lead ECG. The learned feature maps were then fed into two fully connected neural networks with an output layer of three nodes, which refer to participants without CVD (healthy controls) or patients with NYHA I-II or III-IV HF (Figure 1b). The CNN model accurately classified the patients into three grades with an accuracy of 91.6% (Figure 2a). In the analysis of the retrospective cohort data, the ROC curve showed that the AUCs for the three grades were 0.993 for control, 0.961 for NYHA I/II and 0.973 for NYHA III/IV (Figure 2b and Table 2).

We next investigated which portions of the ECG wave are crucial features of each NYHA classification. We employed Grad-CAM to highlight the regions of the input heartbeats that were essential for the classification[30]. The averaged ECG waveforms for all heartbeats within each of the three groups in the training dataset differed from one another (Figure 2c). The histograms (gray bars) overlaying on the average heartbeat waveforms indicate the essential information transmitted through the network to determine the class-discriminative information in the input time-series. A widened QRS and prolonged PR interval were previously reported to be independently associated with in-hospital death, post-discharge death, and rehospitalization[31]. However, there is little consensus on the predictive value of ECG characteristics in the HF population[32]. In the present study, we used all of the patients' ECGs to create an average lead-I ECG for each HF status (Figure 2c). Comparison among the average ECG waveforms revealed that severer HF associated with smaller amplitude P, R and T waves. In addition, a wider QRS complex was observed in the NYHA III/IV group In the Grad-CAM histogram overlayed on the average ECG forms, the latter half of the P wave, the QRS complex, and the beginning and last part of the T wave were outstanding features of NYHA I/II, but not the others. The specific features of the NYHA III/IV group were on broader parts of average wave in addition to an outstanding feature at the onset of the P wave. These HF class-specific ECG waveforms may provide new insight into how heart burden affects the electrical activity of the heart.

NYHA classification is a symptom-based HF classification system that is useful for monitoring the effects of therapy[33], and NYHA classes after therapy are closely related to prognosis[34]. Qualitative classification of HF, like the NYHA classification, is useful and easy to understand for non-cardiologists. However, to individualize medication and disease management for HF patients and to predict prognoses, quantitative assessment to detect daily changes in the heart's condition is necessary. We therefore defined a unique quantitative index of HF severity (HF-index), which the CNN algorithm calculates from the lead-I ECG-based NYHA grade using the following equations.

$$E[X] = \sum_{i=1}^{3} x_i P\left(X = x_i\right),\ x \in \{0,1,2\}$$

The HF-Index is then scaled in the range of [0,100] by normalizing $E[X]$ (*)

$$\text{HF-Index} = \frac{E[X] - x_{min}}{(x_{max} - x_{min})} \times 100,\ 0 \leq \text{HF-Index} \leq 100\ (*)$$

From the input data, the CNN selects the output class (X) as the class to which the input data has the highest probability of belonging. Values of 0, 1 or 2, which correspond to healthy control, NYHA I/II and NYHA III/IV, respectively, are assigned to X, and $P(X)$ is the probability of the corresponding class (X). $E[X]$ is the expected value, calculated as the probability-weighted average of the three output classes.

We then tested whether the HF-index could indicate HF severity as quantitatively as the plasma BNP level. Plasma BNP levels measured on the same day's ECGs were recorded were compared to HF-indexes calculated from all data sets used for CNN construction (Figure 3a). The scatter plot relating plasma BNP levels and HF-indexes for all patients for whom we had BNP levels (n=40,728) is shown in Figure 3a. When we extracted the patients without renal dysfunction (creatinine (Cre) ≤ 1 mg/dL), simple regression showed a strong positive correlation between the HF-index and BNP level, which was averaged among the group corresponding to the same HF-index (R = 0.89) (Figure 3b). Because plasma BNP levels are reportedly affected by renal dysfunction[35], the impact of renal dysfunction on the correlation between the HF-index and BNP level was analyzed. Notably, simple regression also showed a strong correlation between HF-indexes and BNP levels in the patients with renal dysfunction (Figure 3c). This result suggests the CNN can identify the severity of HF based on the HF-index, irrespective of renal function.

To evaluate the feasibility of using the HF-index to detect HF progression, we retrospectively collected long-time-series data, including clinical symptoms and multiple ECGs and plasma BNP levels from another 30 patients who had been hospitalized for HF at least once (male: 13 (43%), mean age: 66 ± 2.4 years). We then compared the time-courses of changes in the HF-index and plasma BNP level. Representative cases are shown in Figure 3d and e. In most cases, changes in BNP levels and HF-indexes were similar over the observed period, which suggests the HF-index may be a promising tool with which to detect worsening HF during continuous observation of HF patients.

Lastly, a prospective observational study was conducted in which we verified the performance with HF patients at home using self-recorded single-lead ECGs (Figure 4). In this study, we used a portable ECG device that recorded lead-I ECGs and transferred the data to a hospital in real time via 4G/LTE (Figure 4a). The wave form from the mobile ECG device sometimes differed from the lead-I ECG extracted from a standard 12-lead ECG. The major difference appeared to reflect the patient's position during the ECG recording (i.e., supine, standing or seated). We therefore compared the differences between the lead-I ECG wave forms from the portable device with the patient in various positions to lead-I ECGs from a standard 12-lead ECG device with the patient in a supine position. The correlations between the two were 0.98 (both in supine position), 0.97 (portable ECG in sitting position vs standard ECG in supine) and 0.84 (portable ECG in standing position vs standard ECG in supine). Based on the results, a sitting or supine position was recommended for recording lead-I ECGs at home in the prospective study.

Fifteen patients (male: 13 (87%), mean age: 54±3.7 years, EF: 45.8±4.4%) were enrolled in the prospective pilot study, beginning in November 2020; their characteristics are shown in Table 3. All patients used portable devices to record 30-s ECGs by themselves at home at least once a day. The mean observation period was 307.5±45.9 days. One patient who had dilated cardiomyopathy (DCM) was hospitalized during the observation period due to worsening HF with reduced ejection fraction (HFrEF) for which they underwent implantation of a CRTD. None of the other patients were hospitalized for HF.

The CNN model first divided the 30-s ECGs into three 10-s ECG segments and gave a NYHA class prediction for each segment. The final decision for the 30-s ECG was based on a majority of the three predictions. We then averaged the probabilities of the three classes from predictions that were the same as the final decision and calculated the HF-index based on the averaged probabilities as previously mentioned (*). Two representative cases are shown in Figure 4b and c. The time-course of the HF-index calculated from daily ECGs self-collected by patients at home correlated well with the BNP levels in both cases (Figure 4b and c). The patient in case 1 was admitted to the hospital due to worsening HF and discharged as indicated (Figure 4b); however, HF worsened again 4.5 months after the discharge (Figure 4b). The clinical course was clearly reflected in the serum BNP levels. Likewise, the HF-index nicely traced the time course of the BNP levels.

The patient in case 2 was also admitted due to worsening HF and discharged soon after (Figure 4c). This patient did not require rehospitalization, and the decreases in the HF-index corresponded to low plasma BNP levels. Like BNP, the HF-index well represented the worsening and improvement of this patient's HF status. These results suggest it may be feasible to use the HF-index to monitor HF status at home.

Discussion

Biomarkers such as BNP and NT-pro-BNP are often used for diagnosing or monitoring HF[36], and reductions in these biomarkers may indicate a better prognosis[37]. Our CNN achieved greater accuracy for detecting HF than BNP, though the accuracy of BNP is high (AUC 0.60)[38]. In fact, the abnormal ECG patterns seen in chronic HF are consequences of such pathologies as left ventricular hypertrophy and left ventricular dysfunction, and these abnormal ECG features have proven to effectively capture changes in cardiac status during the management with HF[39].

Several earlier studies reported a CNN able to distinguish HF patients from healthy controls using single beats or 2 s or 5 s of single-lead ECG from a public data base[40,28,29]. However, that algorithm was unable to classify the severity of HF, though it could potentially help non-specialists avoid overlooking apparent HF. To improve current HF management and avoid HF recurrence, detailed classification of symptoms collected from daily life, including mild symptoms, is required. Our proposed CNN model enables us to distinguish HF patients from healthy controls with high accuracy (AUC 0.961-0.973). Moreover, it can assess the severity of HF based on NYHA classification and distinguish between NYHA I/II and NYHA III/IV, which suggests it could be used for early detection of HF worsening at home. Such early detection of worsening HF would enable prescription of additional medications before decompensation, an

approach that has been linked to decreasing the incidence major adverse cardiovascular events and unplanned hospital admissions[41]. Ours is the first report of a CNN able to classify the severity of HF. Moreover, we succeeded in developing a more quantitative index that provides the severity of HF expressed as a HF-index, which is very simple. Indeed, retrospective evaluation showed that the HF-index could be used to quantitatively represent the clinical course of HF patients (Figure 3d and e) and that it is linearly related to the plasma BNP level (Figure 3). We anticipate that applying our proposed HF-index, which, like BNP, represents HF severity quantitatively, to portable single-lead ECG monitors will enable early detection of changes of their HF condition by patients through self-monitoring. A prospective study from Choi et al. showed that their AI-Clinical Decision Support System could distinguish between patients presenting with dyspnea due to HF or a non-HF ailment, such as lung disease. That AI support system diagnosed these patients based on information that included data from their physical examination, medical history, laboratory findings, ECGs and echocardiography collected when they visited an outpatient clinic[27]. However, HF usually develops outside of the hospital. To improve patient prognosis, it is more important to detect HF progression before the recurrence of decompensation. From that viewpoint, our prospective analysis makes a novel contribution to the implementation of early detection of HF recurrence at home and early medical intervention.

Study Limitations

First, the lead I ECGs used for modeling in this study were recorded with the patient in a supine position using 12-lead ECG equipment. Further evaluation will be needed to verify the efficacy of the proposed model when applied to other ECG devices, including other portable lead-I ECG recording systems. Second, the size of the sample used for verification of the clinical diagnostic efficacy of the AI algorithm was small, and all of the patients were Japanese. A long-term, multicenter prospective investigation will be required to make precise estimates of the algorism's contribution to clinical outcome.

Conclusions

This study showed the feasibility of a CNN model for HF detection using single- lead ECG data transferred from patients' homes. Because daily ECG recording using a portable or wearable ECG device is convenient and non-invasive, the HF-index calculated from the transferred ECGs could be used for remote HF monitoring.

Declarations

Author contributions

E.H. and K.F. contributed to the study design, study coordination, data analysis and interpretation, and writing of the manuscript. E.H., K.F. and Y.C. contributed to data collection and data management. M.O., Y.S., K.K., K.G., R.U., T.O., J.M., T.M., J.S., Y.N., A.S., H.M., G.O., T.K., M.Y. and S.K. contributed to analyzing the ECG interpretation. H.A., M.S., and I.K were senior supervisors on the project. All authors take

responsibility for all aspects of the reliability and freedom from bias of the data presented and their discussed interpretation.

Acknowledgments

We gratefully acknowledge the excellent technical assistance of M. Takahashi. This study was supported by the Grant-in-Aid for Scientific Research from JSPS, AMED, and JST (JP17H05594, JP17H04171, 17H04171, 19K22615, 20ek0109423h0001 to K.F., JPMJMS2023 to K.F.

Competing Interests

The authors declare no competing financial interests.

Corresponding author

Katsuhito Fujiu and Issei Komuro

Funding

This study was partial financially supported by SIMPLEX QUANTUM INC., Tokyo, Japan.

Conflicts of interest

One researcher had positions at both The University of Tokyo and SIMPLEX QUANTUM INC. No other employees of SIMPLEX QUANTUM INC. were involved in any procedures of this study, including its design or the collection and analysis of the data.

References

1. Catanzariti, D. *et al.* Monitoring intrathoracic impedance with an implantable defibrillator reduces hospitalizations in patients with heart failure. Pacing Clin Electrophysiol **32**, 363–370, doi:10.1111/j.1540-8159.2008.02245.x (2009).
2. Landolina, M. *et al.* Remote monitoring reduces healthcare use and improves quality of care in heart failure patients with implantable defibrillators: the evolution of management strategies of heart failure patients with implantable defibrillators (EVOLVO) study. Circulation **125**, 2985–2992, doi:10.1161/circulationaha.111.088971 (2012).
3. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet **392**, 1789–1858, doi:10.1016/s0140-6736(18)32279-7 (2018).
4. Groenewegen, A., Rutten, F. H., Mosterd, A. & Hoes, A. W. Epidemiology of heart failure. Eur J Heart Fail, doi:10.1002/ejhf.1858 (2020).
5. O'Connor, C. M. *et al.* Predictors of mortality after discharge in patients hospitalized with heart failure: an analysis from the Organized Program to Initiate Lifesaving Treatment in Hospitalized

Patients with Heart Failure (OPTIMIZE-HF). Am Heart J **156**, 662–673, doi:10.1016/j.ahj.2008.04.030 (2008).

6. Setoguchi, S., Stevenson, L. W. & Schneeweiss, S. Repeated hospitalizations predict mortality in the community population with heart failure. Am Heart J **154**, 260–266, doi:10.1016/j.ahj.2007.01.041 (2007).
7. Lindmark, K. *et al.* Recurrent heart failure hospitalizations increase the risk of cardiovascular and all-cause mortality in patients with heart failure in Sweden: a real-world study. ESC Heart Fail **8**, 2144–2153, doi:10.1002/ehf2.13296 (2021).
8. Metra, M. & Teerlink, J. R. Heart failure. The Lancet **390**, 1981–1995, doi:https://doi.org/10.1016/S0140-6736(17)31071-1 (2017).
9. Steinberg, B. A. *et al.* Trends in patients hospitalized with heart failure and preserved left ventricular ejection fraction: prevalence, therapies, and outcomes. Circulation **126**, 65–75, doi:10.1161/circulationaha.111.080770 (2012).
10. Dunlay, S. M., Roger, V. L. & Redfield, M. M. Epidemiology of heart failure with preserved ejection fraction. Nature Reviews Cardiology **14**, 591–602, doi:10.1038/nrcardio.2017.65 (2017).
11. Anjan, V. Y. *et al.* Prevalence, clinical phenotype, and outcomes associated with normal B-type natriuretic peptide levels in heart failure with preserved ejection fraction. Am J Cardiol **110**, 870–876, doi:10.1016/j.amjcard.2012.05.014 (2012).
12. Flachskampf, F. A. *et al.* Cardiac Imaging to Evaluate Left Ventricular Diastolic Function. JACC Cardiovasc Imaging **8**, 1071–1093, doi:10.1016/j.jcmg.2015.07.004 (2015).
13. Shah, S. J. *et al.* Research Priorities for Heart Failure With Preserved Ejection Fraction. Circulation **141**, 1001–1026, doi:doi:10.1161/CIRCULATIONAHA.119.041886 (2020).
14. Abraham, W. T. *et al.* Wireless pulmonary artery haemodynamic monitoring in chronic heart failure: a randomised controlled trial. Lancet **377**, 658–666, doi:10.1016/s0140-6736(11)60101-3 (2011).
15. Inan, O. T. *et al.* Novel Wearable Seismocardiography and Machine Learning Algorithms Can Assess Clinical Status of Heart Failure Patients. Circ Heart Fail **11**, e004313, doi:10.1161/circheartfailure.117.004313 (2018).
16. Siontis, K. C., Noseworthy, P. A., Attia, Z. I. & Friedman, P. A. Artificial intelligence-enhanced electrocardiography in cardiovascular disease management. Nature Reviews Cardiology **18**, 465–478, doi:10.1038/s41569-020-00503-2 (2021).
17. Ahmed, F. Z. *et al.* Triage-HF Plus: a novel device-based remote monitoring pathway to identify worsening heart failure. ESC Heart Fail **7**, 107–116, doi:10.1002/ehf2.12529 (2020).
18. Morgan, J. M. *et al.* Remote management of heart failure using implantable electronic devices. Eur Heart J **38**, 2352–2360, doi:10.1093/eurheartj/ehx227 (2017).
19. Abraham, W. T. & Perl, L. Implantable Hemodynamic Monitoring for Heart Failure Patients. J Am Coll Cardiol **70**, 389–398, doi:10.1016/j.jacc.2017.05.052 (2017).

20. Burri, H. *et al.* Risk stratification of cardiovascular and heart failure hospitalizations using integrated device diagnostics in patients with a cardiac resynchronization therapy defibrillator. Europace **20**, e69-e77, doi:10.1093/europace/eux206 (2018).
21. Virani, S. A. *et al.* Prospective evaluation of integrated device diagnostics for heart failure management: results of the TRIAGE-HF study. ESC Heart Fail **5**, 809–817, doi:10.1002/ehf2.12309 (2018).
22. Ahmed, F. Z. *et al.* Remote monitoring data from cardiac implantable electronic devices predicts all-cause mortality. Europace, doi:10.1093/europace/euab160 (2021).
23. Landolina, M. *et al.* Remote Monitoring Reduces Healthcare Use and Improves Quality of Care in Heart Failure Patients With Implantable Defibrillators. Circulation **125**, 2985–2992, doi:doi:10.1161/CIRCULATIONAHA.111.088971 (2012).
24. Pascual-Figal, D. & Bayes-Genis, A. The misperception of 'stable' heart failure. Eur J Heart Fail **20**, 1375–1378, doi:10.1002/ejhf.1248 (2018).
25. Chiou, Y. A., Hung, C. L. & Lin, S. F. AI-Assisted Echocardiographic Prescreening of Heart Failure With Preserved Ejection Fraction on the Basis of Intrabeat Dynamics. JACC Cardiovasc Imaging **14**, 2091–2104, doi:10.1016/j.jcmg.2021.05.005 (2021).
26. Hirata, Y. *et al.* Deep Learning for Detection of Elevated Pulmonary Artery Wedge Pressure Using Standard Chest X-Ray. Can J Cardiol **37**, 1198–1206, doi:10.1016/j.cjca.2021.02.007 (2021).
27. Choi, D.-J., Park, J. J., Ali, T. & Lee, S. Artificial intelligence for the diagnosis of heart failure. npj Digital Medicine **3**, 54, doi:10.1038/s41746-020-0261-3 (2020).
28. Acharya, U. R. *et al.* Deep convolutional neural network for the automated diagnosis of congestive heart failure using ECG signals. Applied Intelligence **49**, 16–27, doi:10.1007/s10489-018-1179-1 (2019).
29. Sudarshan, V. K. *et al.* Automated diagnosis of congestive heart failure using dual tree complex wavelet transform and statistical features extracted from 2s of ECG signals. Computers in Biology and Medicine **83**, 48–58, doi:https://doi.org/10.1016/j.compbiomed.2017.01.019 (2017).
30. Selvaraju, R. R. *et al.* Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. International Journal of Computer Vision **128**, 336–359, doi:10.1007/s11263-019-01228-7 (2020).
31. Park, S.-J. *et al.* Short- and long-term outcomes depending on electrical dyssynchrony markers in patients presenting with acute heart failure: Clinical implication of the first-degree atrioventricular block and QRS prolongation from the Korean Heart Failure registry. American Heart Journal **165**, 57–64.e52, doi:https://doi.org/10.1016/j.ahj.2012.10.009 (2013).
32. Gouda, P., Brown, P., Rowe, B. H., McAlister, F. A. & Ezekowitz, J. A. Insights into the importance of the electrocardiogram in patients with acute heart failure. European Journal of Heart Failure **18**, 1032–1040, doi:10.1002/ejhf.561 (2016).
33. Ammar, K. A. *et al.* Prevalence and prognostic significance of heart failure stages: application of the American College of Cardiology/American Heart Association heart failure staging criteria in the

community. Circulation **115**, 1563–1570, doi:10.1161/circulationaha.106.666818 (2007).

34. Dickstein, K. *et al.* ESC guidelines for the diagnosis and treatment of acute and chronic heart failure 2008: the Task Force for the diagnosis and treatment of acute and chronic heart failure 2008 of the European Society of Cardiology. Developed in collaboration with the Heart Failure Association of the ESC (HFA) and endorsed by the European Society of Intensive Care Medicine (ESICM). Eur J Heart Fail **10**, 933–989, doi:10.1016/j.ejheart.2008.08.005 (2008).
35. Takase, H. & Dohi, Y. Kidney function crucially affects B-type natriuretic peptide (BNP), N-terminal proBNP and their relationship. European Journal of Clinical Investigation **44**, 303–308, doi:https://doi.org/10.1111/eci.12234 (2014).
36. Members, A. T. F. *et al.* ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure 2012: The Task Force for the Diagnosis and Treatment of Acute and Chronic Heart Failure 2012 of the European Society of Cardiology. Developed in collaboration with the Heart Failure Association (HFA) of the ESC. European Heart Journal **33**, 1787–1847, doi:10.1093/eurheartj/ehs104 (2012).
37. Masson, S. *et al.* Prognostic value of changes in N-terminal pro-brain natriuretic peptide in Val-HeFT (Valsartan Heart Failure Trial). J Am Coll Cardiol **52**, 997–1003, doi:10.1016/j.jacc.2008.04.069 (2008).
38. Bhalla, V. *et al.* Diagnostic ability of B-type natriuretic peptide and impedance cardiography: testing to identify left ventricular dysfunction in hypertensive patients. Am J Hypertens **18**, 73s-81s, doi:10.1016/j.amjhyper.2004.11.044 (2005).
39. Xie, L., Li, Z., Zhou, Y., He, Y. & Zhu, J. Computational Diagnostic Techniques for Electrocardiogram Signal Analysis. Sensors **20**, 6318 (2020).
40. Porumb, M., Iadanza, E., Massaro, S. & Pecchia, L. A convolutional neural network approach to detect congestive heart failure. Biomedical Signal Processing and Control **55**, 101597, doi:https://doi.org/10.1016/j.bspc.2019.101597 (2020).
41. Halliday, B. P. *et al.* Withdrawal of pharmacological treatment for heart failure in patients with recovered dilated cardiomyopathy (TRED-HF): an open-label, pilot, randomised trial. The Lancet **393**, 61–73, doi:10.1016/S0140-6736(18)32484-X (2019).
42. Dolgin, M. & Committee, N. Y. H. A. C. *Nomenclature and criteria for diagnosis of diseases of the heart and great vessels*. 9th ed. / editor, Martin Dolgin edn, (Little, Brown, 1994).
43. Shin, H.-C. *et al.* Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. IEEE Trans Med Imaging **35**, 1285–1298, doi:10.1109/TMI.2016.2528162 (2016).
44. Pan, J. & Tompkins, W. J. A Real-Time QRS Detection Algorithm. *IEEE Transactions on Biomedical Engineering* **BME-32**, 230-236, doi:10.1109/TBME.1985.325532 (1985).

Methods

Study population and data collection for construction of a CNN

All data used for construction of a CNN for automatic HF classification were retrospectively acquired at the University of Tokyo hospital from participants who were at least 20 years old between 2013 and 2020. The flow of data creation for the CNN is shown in Figure 1a. The healthy control group was composed of individuals who underwent a comprehensive medical examination and were diagnosed as having no cardiovascular disease (CVD). All patients with HF were diagnosed as New York Heart Association (NYHA) I, NYHA II, NYHA III, or NYHA IV according to NYHA functional classification criteria[42]. The definitions of the classifications are as follows:

1. Healthy control: Participants have no CVD.
2. NYHA I: Patient has been hospitalized for HF but has no limitation of physical activity; physical activity does not cause fatigue, palpitation, or dyspnea.
3. NYHA II: Patient has been hospitalized for HF, and physical activity is slightly limited by fatigue, palpitation, or dyspnea, but they have no symptoms at rest.
4. NYHA III: Patient has been hospitalized for HF, and physical activity is limited by fatigue, palpitation, or dyspnea, even at rest.
5. NYHA IV: Patient has been hospitalized for HF and cannot carry on any physical activity without discomfort, and they experience symptoms of HF at rest.

ECG data were obtained from the control group at an annual health check center at the University of Tokyo hospital, where individuals receive regular yearly health checks. ECG data were obtained from patients clinically diagnosed with HF during their hospitalization in the Department of Cardiology. Standard 10-s, 12-lead ECGs were recorded from all patients with HF on admission, and a NYHA class was assigned based on the patient's symptoms and evaluation of their medical examination by expert cardiologists. Data lacking a NYHA classification were excluded, as were data from patients with a pacemaker or poor ECG recordings due to motion artifacts, inaccurate electrode application or excessive noise. The data used for CNN construction were 25,368 10-s ECGs from 6,901 participants, including both HF patients and healthy individuals.

Data preprocessing

Standard 10-s, 12-lead ECGs were recorded at 500 Hz with patients in a supine position in a resting state. Only lead I ECG data were selected for this study. To train the CNN model, we segmented the 10-s ECG recordings into heartbeat waveforms as independent input data[43]. Before the heartbeat segmentation, however, the recordings were preprocessed to eliminate baseline drift and noise. First, the baseline drift was removed using the wavelet decomposition method. ECG data were decomposed into sublevels, and the final approximation coefficient was taken as the baseline drift and subtracted from the original signal. Next, a Butterworth bandpass filter was applied to remove power-line noise and high-frequency distortion. Thereafter, heartbeats were detected by using the Pan-Tompkins algorithm to recognize the peaks of the R waves on the ECG recordings[44]. A window was then used to segment the heartbeats between 0.34 s

before and 0.72 s after the R-wave peaks to capture the PQRST complexes. This process enabled us to adjust the alignment of heartbeats using the R-wave peaks.

Heartbeats were annotated based on the NYHA classification of corresponding 10-s ECG recordings. To evaluate the generalizability and stability of the proposed algorithm, we combined two NHYA classes into one group. The resultant three classes include healthy controls, NYHA I-II and NYHA □-IV. We then trained and tested the model using the three-group dataset. To increase computational efficiency, we removed outlier heartbeats from the NYHA classes and retained the remaining heartbeats for training and validation. Euclidean distance was applied to identify the outliers based on their distances to the center heartbeat.

CNN modeling and validation

The learned features reflected by the CNN parameters were used to identify the NYHA class from the test heartbeat waveform. The CNN takes heartbeat waveforms as one-dimensional time-series inputs and outputs label predictions as NYHA classes. Each convolutional layer was followed by rectified linear unit (ReLU) activation and a 10% dropout to regularize many parameters. To be specific, the kernels (filters) used to generate the convolutional layers were chosen as [1×128], [1×2] and [1×2], respectively. The learned feature maps were then fed into two flattened and fully connected layers followed by a softmax layer with an output layer of nodes corresponding to the NYHA classes.

The proposed model was trained, validated and tested using data that was randomly split into three datasets to avoid overlap of the same ECG data between sets. Resampling imbalanced the three group datasets, and the heart beats in the three classes were adjusted to the same size. The performance was evaluated based solely on the heartbeats in the independent test dataset, which enabled an efficient global evaluation.

Visual explanation of the CNN model used to identify heart failure and NYHA classification

To understand which parts of the heartbeat waveform are most important for NYHA classification, gradient-weighted class activation mapping (Grad-CAM) was used to show the gradient of the classification score for the convolutional features determined by the network. This can help a clinician understand why the CNN model makes a given classification. The idea of Grad-CAM is to calculate the gradient of the final classification score for the final convolutional feature map. The places where this gradient is large are the places in the data upon which the final score most depends. In other words, the data points on the heartbeat waveform that have the highest Grad-CAM scores contribute most to the classification. To create an “average” Grad-CAM for each NYHA class, we calculated the respective pointwise Grad-CAM scores (normalized to between 0 and 1) for each heartbeat. We then calculated the average heartbeat waveform for each class and accumulated the occurrence of data points that had Grad-CAM scores equal to 1. From the obtained “average” heartbeat waveform and the frequency map, the places where high Grad-CAM scores most frequently occur indicate significant features that have the most impact on the classification.

Retrospective cohort study

To evaluate the CNN's ability to detect a temporal change in HF patients, we obtained longitudinal ECG data from the electronic medical records of patients who had been hospitalized for HF at least once. The retrospective time series data was adopted to analyze the performance of the CNN for diagnosis of HF severity. Thirty patients with a history of HF hospitalization (n=30, male: 13(43%), mean age: 66±2.4 years) were randomly selected from the database. All patients were over 20-years-old and had no history of CIED implantation. Lead-I ECG data extracted from 12-lead ECGs and NYHA classifications were collected. Then using these datasets we evaluated the performance of the CNN.

Prospective observational pilot study

To evaluate the feasibility of using a CNN model to detect HF based on remote ECG monitoring, we performed a prospective pilot study (**Trial registration:** UMIN Clinical Trials Registry, UMIN000042073 (http://www.umin.ac.jp/ctr/index.htm)). During their hospitalization, we recruited patients who had been hospitalized because of worsening HF. The inclusion criteria were: (1) age 20 or over; (2) history of hospitalization for HF; and (3) histologically confirmed diagnosis of cardiovascular disease. The exclusion criteria were: (1) under age 20; (2) histologically confirmed diagnosis of pulmonary hypertension; and (3) prior CIED implantation. Fifteen consecutive patients who met the inclusion criteria (n=15, male: 13 (87%), mean age: 54±3.7 years, EF: 45.8±4.4%) participated in the prospective study. Participants were instructed to record their ECGs using a portable ECG monitor every day at home, after which they transmitted the ECG data using a remote monitoring system within the portable ECG monitor. All patients visited our outpatient department every one or two months. Plasma BNP levels and NYHA classification diagnosed based on the patient's *symptoms by cardiologists were collected at the time of each outpatient visit.* We conducted the first patient enrollment and started a pilot study in November 2021. To evaluate the performance of the CNN, the time-course data for a HF-index defined by the CNN using the lead-I ECG waveform were compared with the plasma BNP levels.

All clinical studies were approved by the institutional ethical committee of the University of Tokyo (No. 2020024NI-(3)). For the retrospective cohort, the requirement for written informed consent was waived by the institutional ethical committee. Each patient in the prospective cohort provided informed consent before study enrollment. The study protocol complied with the Declaration of Helsinki.

Single-lead ECG device and remote home monitoring system

In the prospective study, participants recorded ECGs using a portable ECG monitor. We used a wireless single-lead device with two electrodes placed at either end of the body (SHINDENKUN®, SIMPLEX QUANTUM Inc. Tokyo, Japan). Upon gripping the stick-shaped portable ECG device, the bipolar lead-I ECG was recorded. The ECG data was then sent to a data server via the internet, and the CNN assigned the data a NYHA classification and calculated the HF-index in real time.

Statistical considerations.

We compared baseline demographics among patients. Statistical significance was tested between the groups using the chi-square test for categorical variables. The other categorical variables were expressed as numbers (percent) and compared using unpaired Student's t-tests. All analyses were two-sided, and values of *P* less than 0.05 were considered statistically significant. To quantify the validation performance, we estimated the area under the curve (AUC) from receiver-operating characteristic (ROC) curves and the sensitivity, specificity, and accuracy with 95% confidential intervals. Accuracy was evaluated for the optimal operating point on the ROC curve that maximized the sensitivity and specificity. The 95% CI was estimated by bootstrapping 1,000 random and variable sampled instances. All training and validation were implemented using MATLAB R2020a and performed on an NVIDIA GeForce RTX 2080 Ti platform.

Tables

Table 1. Characteristics of the healthy controls and HF patients used for the convolutional neural network (CNN)

	Total (N=6901)	Training (N=5365)	Validation (N=1155)	Test (N=381)	*P*
Age, years	62.5±17.2	62.6±17.2	62.1±17.2	62.9±17.1	0.646
Female, sex, %	45.2	44.6	46.9	45.9	0.945
Mean EF, %	63.1±13.8	63.2±13.8	62.8±13.6	62.5±13.4	0.567
History of HF	57.2	42.1	39.8	63.0	
HFpEF, %	88.9	88.9	86.6	89.0	
HFmrEF, %	4.1	3.9	5.0	4.7	
HFrEF, %	7.1	7.2	6.8	6.3	

HFrEF, heart failure with reduced ejection fraction (EF<40%); HFmrEF, heart failure with mild reduced ejection fraction (EF 40-49%); HFpEF, heart failure with preserved ejection fraction (EF≦50%)

Table 2. CNN Model performance

	AUC	Sensitivity	Specificity	Accuracy
Healthy controls	0.993 (0.985-0.997)	0.976(0.942-0.994)	0.966(0.943-0.982)	0.969(0.950-0.982)
NYHA I /□	0.961(0.942-0.974)	0.896(0.844-0.941)	0.929(0.899-0.953)	0.918(0.896-0.943)
NYHA □/IV	0.973(0.956-0.984)	0.890(0.832-0.933)	0.966(0.940-0.982)	0.943(0.918-0.961)

Table 3. Clinical characteristics of the patients in the prospective observationl study

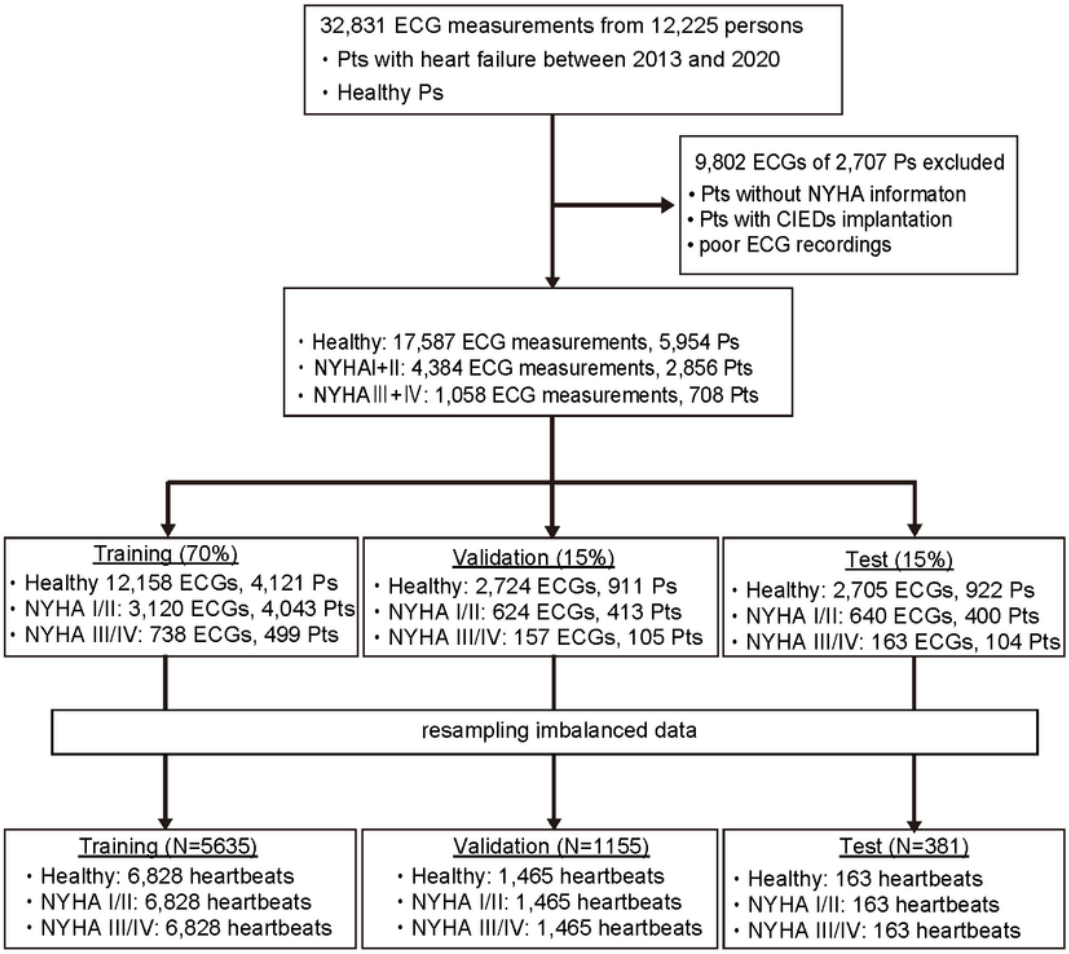
No	Age	Gender	EF (%)	Diagnosis
1	37	female	70	ASD
2	74	male	32	HFrEF
3	48	male	22	HFrEF, EGPA
4	65	male	38	HFrEF
5	59	male	65	OMI
6	64	male	43	ICM
7	63	male	63	HFpEF, PAF
8	82	female	52	ICM
9	43	male	62	DCM
10	48	male	51	HFpEF after MVR
11	37	male	30	Cardiac sarcoidosis
12	47	male	49	OMI
13	41	male	18	DCM
14	56	male	28	DCM
15	56	male	24	DCM

ASD, atrial septal defect; EGPA, Eosinophilic Granulomatosis with Polyangiitis; OMI, *old* myocardial infarction; ICM, ischemic cardiomyopathy; PAF, paroxysmal atrial fibrillation; DCM, dilated cardiomyopathy

Figures

Figure 1

a



b

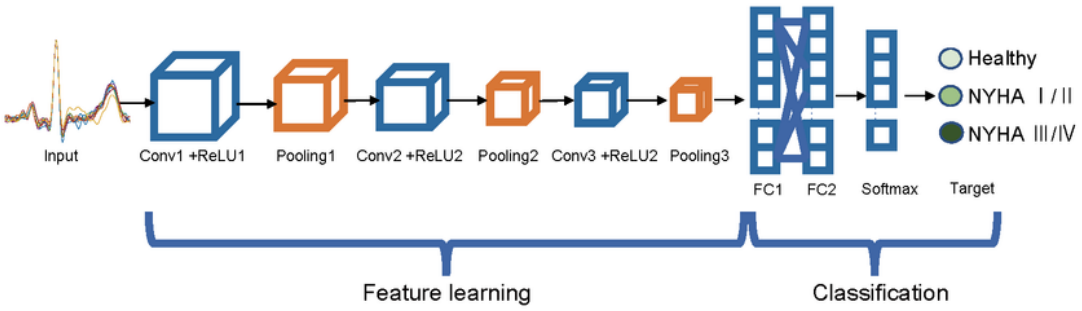


Figure 1

Construction of a heart failure detection CNN model using single-lead ECGs. a, Flow chart for creation of an ECG dataset from heart failure (HF) patients and healthy controls to construct a convolutional neural network (CNN). The CNN model was trained using 70% of the data and optimized using 15% as an internal validation set. To evaluate the CNN model, remaining 15% of the data was assigned as a test dataset. The heartbeats of patients in the training group were used only for development of the CNN. The

performance of the CNN was evaluated using heartbeats of an independent group of test patients. **b**, Construction of the neural network architecture for evaluation of ECGs and HF.

Figure 2

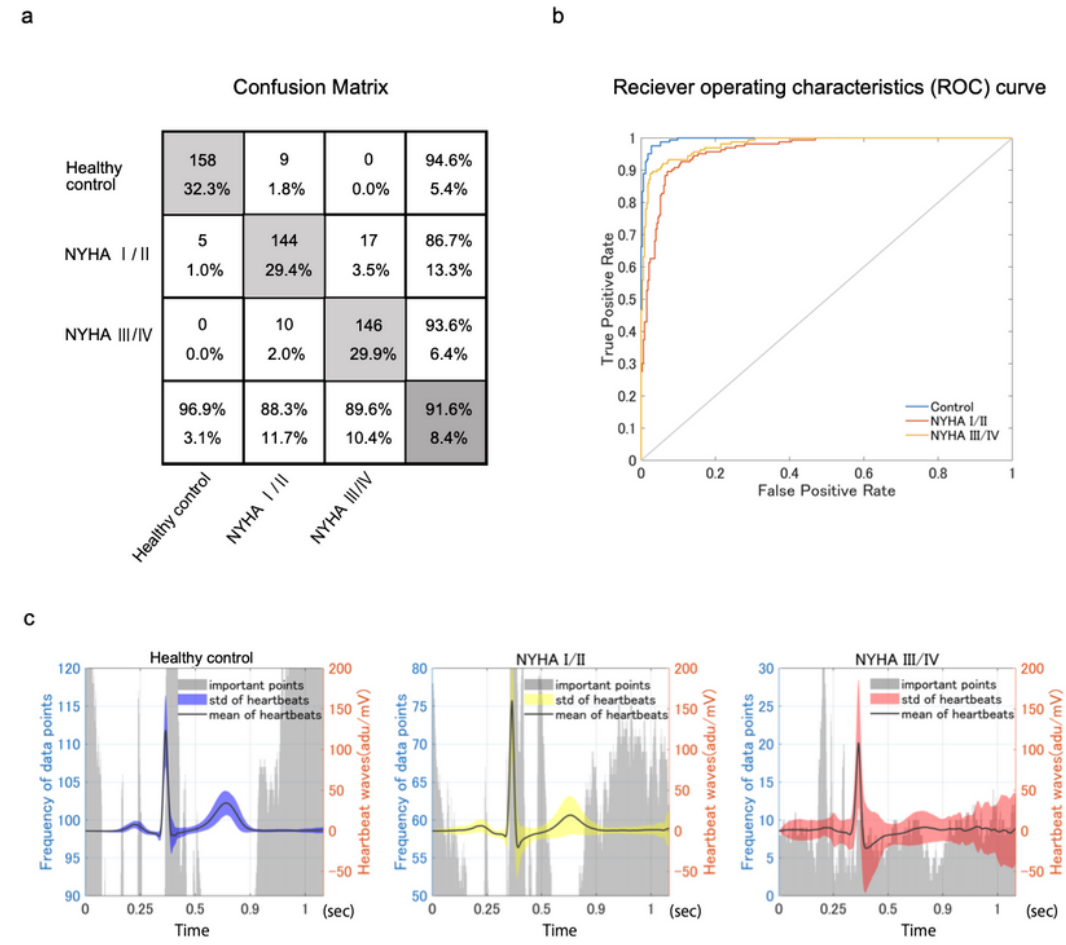


Figure 2

Evaluation of heart failure classification by the CNN model. a, Confusion matrix showing the accuracy of NYHA grade classification by the CNN model. The results show a classification accuracy of 91.6%. **b,**

Receiver-operating characteristic (ROC) curves for classification with three classes: healthy control, NYHA I/II and NYHA III/IV. The areas under the ROC curves (AUCs) were calculated from the results of performance analyses. The details of ROC-AUC curves are shown in Table 2. **c**, Visualization of the feature segments of ECGs used for HF classification by the CNN. The solid lines represent the mean with the respective error bands for all heartbeats corresponding to each classification group in the training datasets. The error bands represent the standard deviation of the heartbeats in each category. The gray colored vertical bars are histograms representing the sum of the sample points obtained by applying Grad-CAM methods to all training heartbeats that have Grad-CAM scores equal to 1 in the normalized heatmaps. Large amplitude histograms indicate the essential regions of a heartbeat used for classifying groups.

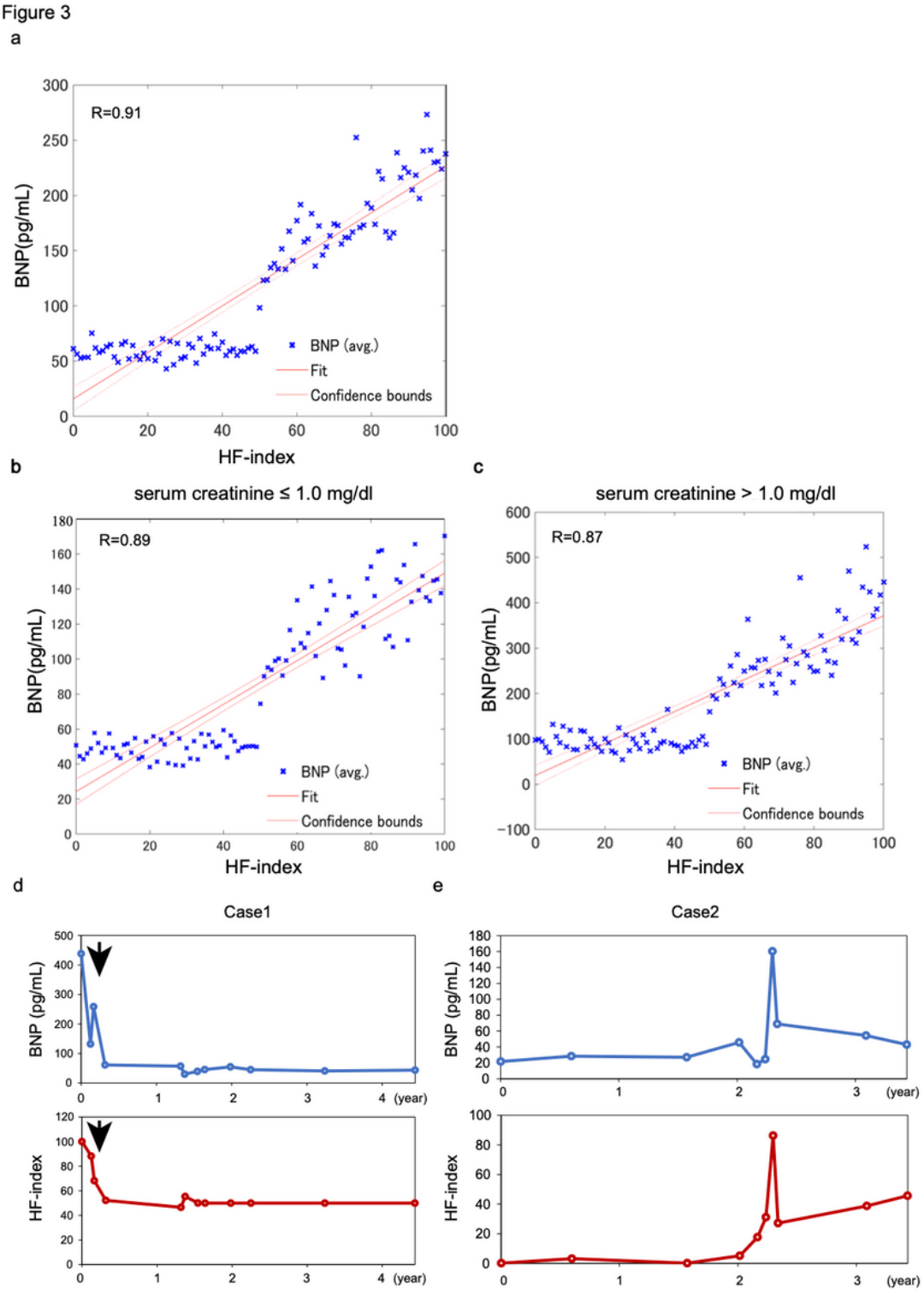


Figure 3

Strong positive correlation between plasma brain natriuretic peptide (BNP) levels and the HF-index calculated by the CNN from the estimated NYHA classification. ***a****, Scatter plot relating plasma BNP levels and HF-index in all patients. The HF-indexes were calculated from recorded ECGs (n=*56,600 *ECGs); BNP levels were simultaneously measured.* BNP levels were averaged among each group corresponding to same HF-index, (R=0.91). ***b and c****, Impact of renal dysfunction on the association between HF-index and*

serum BNP. The simple linear regression indicated a strong positive *relationship* in patients with Cre £ 1.0 mg/dL (Figure 3b, R = 0.89, n=40,728) (**b**) or > 1.0 mg/dL (Figure 3c, R = 0.87, n=15,872) (**c**). ***d and e****, Retrospective* time course analysis of the HF-index and plasma BNP levels in patients hospitalized for HF. **d**, Case 1: A 42-year-old male patient was admitted for acute congestive HF at 0 years. The HF symptoms were *relieved* with medical therapy. The patient was discharged at the indicated times. After discharge, the patient was never re-hospitalized. The HF-index and plasma BNP levels showed similar courses. **e**, Case 2: A 42-year-old female HF patient was stable for 1.5 years after her first hospital admission. However, this patient was re-hospitalized for acute *worsening* of her *HF at 2.2 years*. An acute BNP surge was accompanied by a corresponding surge in the HF-index. R, *coefficient of determination.*

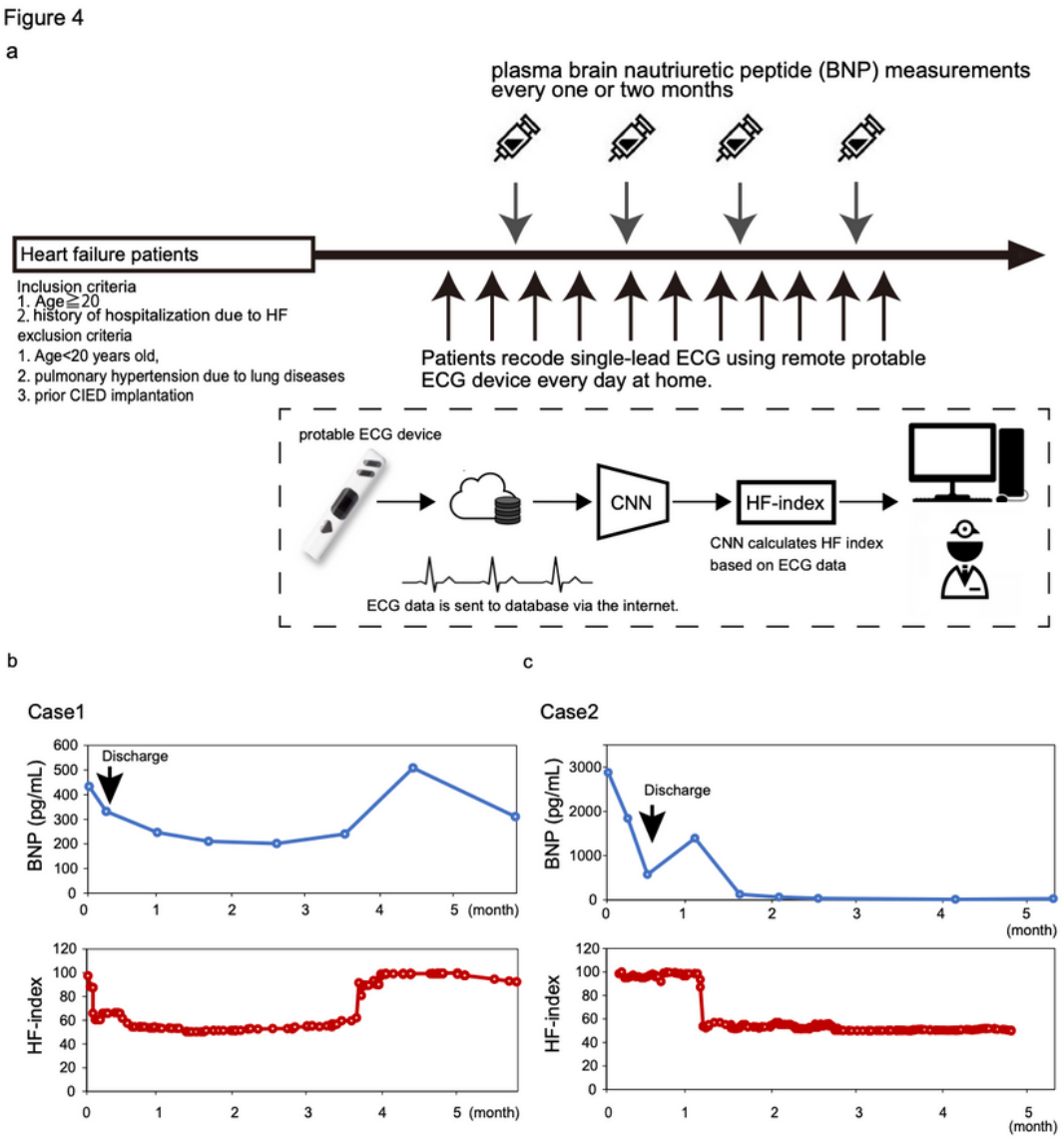


Figure 4

Prospective study: remote monitoring of single-lead ECGs predicts serum BNP levels. a, Study design of a prospective study to evaluate the feasibility of monitoring HF using HF-indexes based on single-lead ECGs recorded at home. Beginning in November 2020, fifteen consecutive patients admitted to the hospital for HF were enrolled. Once enrolled the patients began recording 30-s ECGs using portable devices we developed and shown here. The ECG recordings were first made by the patients at the hospital

before their discharge and then at least once a day at home after their discharge. Plasma BNP levels were also measured as indicated at the hospital**. b and c**, Comparison of HF-indexes and plasma BNP levels obtained during the prospective study. Lead-I ECG data were sent from patient's home to the hospital, and the HF-indexes were calculated by the CNN system in real time. **b**, Case 1: A 62-year-old female patient was admitted to the hospital with HFpEF (EF 63%). She started recording ECGs during her hospitalization at 0 months. She was stable after discharge. However, her HF worsened due to frequent premature ventricular contractions, which developed during the follow-up period, necessitating rehospitalization. Both her HF-indexes and BNP levels were increased at similar times. **c**, Case 2: A 43-year-old male patient was admitted to the hospital because with acute congestive HF. His left ventricular ejection function (LVEF) was reduced to 26%. He was diagnosed with *dilated cardiomyopathy*, and his HF symptoms were resolved with medical therapy. LVEF was 65% at discharge. He has continued to record ECGs using the portable ECG device after discharge. He has no symptoms, and his HF-index remains low, as do his BNP levels. In these two cases, the HF index well reflected the clinical changes in these HF patients.