Clinical Utility of Machine Learning-Derived Vocal Biomarkers in the Management of Heart Failure

Kozo Okada, MD; Daisuke Mizuguchi; Yasuhiro Omiya; Koji Endo;
Yusuke Kobayashi, MD; Noriaki Iwahashi, MD; Masami Kosuge, MD;
Toshiaki Ebina, MD; Kouichi Tamura, MD; Teruyasu Sugano, MD;
Tomoaki Ishigami, MD; Kazuo Kimura, MD; Kiyoshi Hibi, MD

Background: This study aimed to systematically evaluate voice symptoms during heart failure (HF) treatments and to exploratorily extract HF-related vocal biomarkers.

Methods and Results: This single-center, prospective study longitudinally acquired 839 audio files from 59 patients with acute decompensated HF. Patients' voices were analyzed along with conventional HF indicators (New York Heart Association [NYHA] class, presence of pulmonary congestion and pleural effusion on chest X-ray, and B-type natriuretic peptide [BNP]) and GOKAN scores based on the assessment of a cardiologist. Machine-learning (ML) models to estimate HF conditions were created using a Light Gradient Boosting Machine. Voice analysis identified 27 acoustic features that correlated with conventional HF indicators and GOKAN scores. When creating ML models based on the acoustic features, there was a significant correlation between actual and ML-derived BNP levels (r=0.49; P<0.001). ML models also identified good diagnostic accuracies in determining HF conditions characterized by NYHA class ≥2, BNP ≥300 pg/mL, presence of pulmonary congestion or pleural effusion on chest X-ray, and decompensated HF (defined as NYHA class ≥2 and BNP levels ≥300 pg/mL; accuracy: 75.1%, 69.1%, 68.7%, 66.4%, and 80.4%, respectively).

Conclusions: The present study successfully extracted HF-related acoustic features that correlated with conventional HF indicators. Although the data are preliminary, ML models based on acoustic features (vocal biomarkers) have the potential to infer various HF conditions, which warrant future studies.

Key Words: Acoustic features; Artificial intelligence; Telemedicine; Vocal biomarker

Heart failure (HF) is a leading cause of mortality and morbidity worldwide, and is a chronic, progressive disorder with repeating remission and exacerbation.[1,2] Additionally, HF has immeasurable physical and economic burdens on patients.[3,4] In the year following HF hospitalization, the rate of rehospitalization is approximately 50%, and the 1-year mortality rate is 15–20%.[4,5] In contrast, even when HF exacerbations occur, it has been reported that appropriate intervention in the early phase of HF decompensation can reduce the rehospitalization rate and disease progression of HF.[6–10] Therefore, early detection of signs and symptoms of cardiac decompensation is important in HF management after hospital discharge.

Previous attempts at early detection of exacerbations have primarily proposed telemonitoring, aimed at capturing signs of exacerbations early while at home. However, telemonitoring of changes in symptoms, vital signs, and body weight alone has not offered consistent results in reducing HF rehospitalization.[9–15] In contrast, it has been reported that increasing face-to-face and conversational opportunities through frequent home visits and telephone interviews, in addition to monitoring symptoms and vital signs, could improve rehospitalization rates and prognosis of HF.[7–10] These results suggest there are possible benefits of assessing biometric information acquired from a patient's face and voice together for early detection of exacerbations. In daily practice, it is not uncommon for skilled clinicians to recognize the early signs of HF from changes in a patient's face and voice.[16] However, the diagnostic skills of

medical professionals in capturing changes in face and voice symptoms vary among individuals and it is difficult to verbalize and communicate these alterations as well. To apply this biometric information as a 'versatile, evidence-based new biomarker', it is necessary to confirm their association with established HF indicators and make them indicators that can be objectively evaluated. As mentioned earlier, HF is a recurrent disease with repeated remissions and exacerbations. Considering this, we hypothesized that the exacerbation process of HF could be efficiently inferred by retrospectively reviewing the process of remission after inpatient treatment. Therefore, the present study aimed to systematically evaluate voice symptoms in conjunction with conventional HF indices during HF treatments and to extract HF-related acoustic features as vocal biomarkers.

Methods

Study Population
This study was the voice part of the GOKAN-HF study. The GOKAN-HF study is a prospective, observational study that aimed to extract biometric symptoms obtained from face and voice alterations, the most familiar objects of recognition, as digital biomarkers to complement the 'five senses' (GOKAN in Japanese) of humans. Through this, we could create a digital biomarker that comprehensively incorporates and reflects HF-related symptoms, and physical and laboratory findings, and aim to propose a solution to the clinically challenging problem of 'early detection of HF deterioration'.

Between June 2021 and February 2023, patients who were hospitalized for acute decompensated HF at Yokohama City University Medical Center were eligible for enrollment if they were aged 20–89 years and provided written informed consent to participate in this study. To qualify the quality of the analysis, patients were excluded if they had any of the following factors that could affect voice and facial assessments: diseases of the vocal tracts including vocal cords; severe respiratory disease; sepsis; novel coronavirus infection; dermatologic disorders; facial trauma or tumors; psychiatric disorders; neurodegenerative diseases; disturbance of consciousness; delirium; stroke or a history of stroke; dialysis treatment; cancer treatment; diseases with a prognosis of ≤1 year; or hemodynamic instability that required inotropic therapy or mechanical circulatory support. Patients were also excluded if they were unable to maintain a sitting position for recording, or if they could not be recorded because of treatments including endotracheal intubation or noninvasive positive pressure ventilation (detailed patient flow is shown in **Supplementary Figure 1**). All patients received standard medical treatment for HF according to clinical guidelines.[17] Cardiologists made all medical decisions regarding indications for hospitalization, timing of discharge, and treatment during hospitalization, and were blinded to voice analysis results. Patients were followed for up to 541 (mean 239±139) days and evaluated for HF exacerbation or rehospitalization for HF. HF exacerbation was defined as ≥2 of the following: worsening symptoms; significant fluid accumulation such as leg edema; pleural effusion; and pulmonary congestion; and/or significant increase in B-type natriuretic peptide (BNP) levels. The study protocol was approved by the institutional review board (IRB) at Yokohama City University, and followed the Declaration of Helsinki and ethical standards of the responsible committee on human experimentation.

Conventional HF Indicators and GOKAN Scores
Conventional HF indicators included New York Heart Association (NYHA) class, vital signs, percutaneous oxygen saturation, body weight, rales, leg edema, BNP, chest X-ray findings (pulmonary congestion and pleural effusion) and left ventricular ejection fraction (LVEF) on echocardiology. In our daily practice, blood tests (including BNP) and chest X-rays were performed daily during the acute phase of HF inpatient treatment, every few days at the attending physician's discretion once the patient's condition stabilized, and every 1–3 months as needed after discharge. LVEF was assessed at admission, around discharge (17 [10–41] days from admission), and at the chronic phase (167 [107–242] days from admission). According to NYHA class and BNP levels, the HF status of patients was clinically classified into 3 subgroups: (1) NYHA class ≥II and BNP levels ≥300 pg/mL[18] for decompensated HF; (2) NYHA class ≥II or BNP levels ≥300 pg/mL for compensated-to-decompensated HF; and (3) NYHA class <II and BNP levels <300 pg/mL for compensated HF. The GOKAN score was the original score composed of 11-item scoring to visualize potential information on HF obtained from the faces and voices of patients and scored by an experienced cardiologist who was blinded to clinical and voice analysis data in the same session voice recording (**Supplementary Table**). Each parameter was scored at −2, −1, 0, 1 and 2 for a total of −22 to 22, with lower values indicating worse HF status and higher values indicating better HF status. The intra- and interobserver intra-class correlation coefficients for the GOKAN score were 0.999 and 0.998 (P<0.0001 for both comparisons).

Voice Recording and Analysis
Patient voice acquisition was performed on the same schedule as measurements of conventional HF indicators. Recordings were made using a directional pin microphone (MX150B, SHURE, USA) connected to a portable, linear pulse-code modulation recorder (DR-100mkIII, TASCAM, Japan), at a sampling rate of 192 kHz with 24-bit resolution. The microphone was attached to the patient's clothes at chest level, approximately 15 cm from the mouth, and the voices were recorded in a coronary care unit (recording with monitor sound off) or in a specific quiet room in the cardiovascular ward. In each recording session, the patient performed 2 simple language-independent tasks: (1) after taking a deep breath, utter the sustained vowel sound (/a/) as long as possible; and (2) repeat the trisyllable (/pataka/) 5 times or more as quickly as possible. The recording of the two tasks took less than 1 min.

For voice analyses, all measurement data were used and their associations with concomitant assessments of conventional HF indicators and GOKAN scores were assessed. The audio signals were downsampled to 16 kHz with 16-bit resolution for acoustic feature extraction. Voice analysis identified many acoustic features associated with changes in HF conditions. Among them, 27 acoustic features (21 and 6 features for /a/ and /pataka/, respectively), which correlated with conventional HF indicators and GOKAN scores, were extracted. The features derived from the phrases /a/ and /pataka/ included the statistics of pitch-related or voice quality-related features (e.g., shimmer, jitter, and harmonics-to-noise ratio) and peak intensity-related features, respectively. For the calculation of pitch-related or voice quality-related features, the audio signal was processed for each 10 ms window length. In addition,

Table. Clinical Characteristics	
Variable	All (n=59)
Age (years)	74 [63~81]
Male sex (%)	57.6
BMI (kg/mm²)	24.3 [22.2~28.3]
Past medical history (%)	
Current or former smoker	57.7
Hypertension	80.0
Dyslipidemia	44.1
Diabetes	39.0
Type of HF (%)	
HFrEF/HFpEF	83.1/16.9
First admission for HF (%)	83.1
Primary underlying cardiac disease (%)	
Coronary artery disease	13.6
Non-ischemic cardiomyopathy	44.1
Valvular heart disease	23.7
Arrhythmia	18.6
Length of hospital stay (days)	14 [11~18]
Days from symptom onset to hospitalization	25 [10~51]
No. patients who visited a clinic prior to admission (%)	72.9
Days from symptom onset to a clinic	8 [0~28]
Days from clinic visit to hospitalization	13 [4~29]
NYHA functional class I/II/III/IV (%)	
Admission	0/8.5/25.4/66.1
Discharge	90.0/10.0/0/0
GOKAN score	
Admission	−20 [−22~−12]
Discharge	22 [15~22]
BW (kg)	
Admission	62 [54~81]
Discharge	57 [48~71]
BW changes during hospitalization	−6.0 [−9.8~−3.6]
Heart rate (beats/min)	
Admission	96±21
Discharge	72±14
Sinus rhythm (%)	
Admission	54.2
Discharge	67.8
Systolic blood pressure (mmHg)	
Admission	150 [126~162]
Discharge	111±15
Diastolic blood pressure (mmHg)	
Admission	90±25
Discharge	66±14
Respiratory rate (breaths/min)	
Admission	21 [17~24]
Discharge	16 [15~16]
Rales (%)	
Admission	96.6
Discharge	0
Leg edema (%)	
Admission	76.3
Discharge	0
Blood biomarkers	
Hemoglobin (g/L)	
Admission	13.9 [11.6~14.9]
Discharge	13.9±2.3

(**Table** continued the next column.)

Variable	All (n=59)
Blood biomarkers	
Albumin (g/dL)	
Admission	3.7 [3.5~4.0]
Discharge	3.6±0.4
eGFR (mL/min/1.73 m²)	
Admission	49.5 [35.5~62.4]
Discharge	44.9 [34.5~54.7]
BNP (pg/mL)	
Admission	968 [610~1,915]
Discharge	312 [154~543]
Medications on admission (%)	
ACEI/ARB/ARNI	
Admission	45.8
Discharge	93.2
β-blockers	
Admission	39.0
Discharge	84.5
MRA	
Admission	18.6
Discharge	78.0
SGLT-2i	
Admission	13.6
Discharge	71.2
Loop diuretic agents	
Admission	40.7
Discharge	74.6
Tolvaptan	
Admission	3.4
Discharge	14.0
Echocardiography	
LVEF (%)	
Admission	30 [23~38]
Discharge	34 [29~50]
Follow-up*	47±16
LVDd (mm)	
Admission	55±8
Discharge	53±8
Follow-up*	50±8
LVDs (mm)	
Admission	45±0
Discharge	42±10
Follow-up*	38±10
LVEF changes from admission	
Discharge	6±12
Follow-up*	16±16

Values are presented as percentage (%), mean±standard deviation (SD), or median [interquartile range]. LVEF at discharge was performed at 17 (10~41) days from admission and LVEF at follow up was performed at 167 (107~242) days from admission. *Results of echocardiography at follow-up for 8 out of 59 patients were unavailable due to referral to another clinic/hospital after discharge. ACEI, angiotensin-converting enzyme inhibitor; ARB, angiotensin II receptor blocker; ARNI, angiotensin receptor neprilysin inhibitor; BMI, body mass index; BNP, B-type natriuretic peptide; BW, body weight; eGFR, estimated glomerular filtration rate; HF, heart failure; HFpEF, heart failure with preserved ejection fraction; HFrEF, heart failure with reduced ejection fraction; LVDd, left ventricular end-diastolic diameter; LVDs, left ventricular end-systolic diameter; LVEF, left ventricular ejection fraction; MRA, mineralocorticoid receptor antagonists; NYHA, New York Heart Association; SGLT-2i, sodium-glucose cotransporter 2 inhibitors.
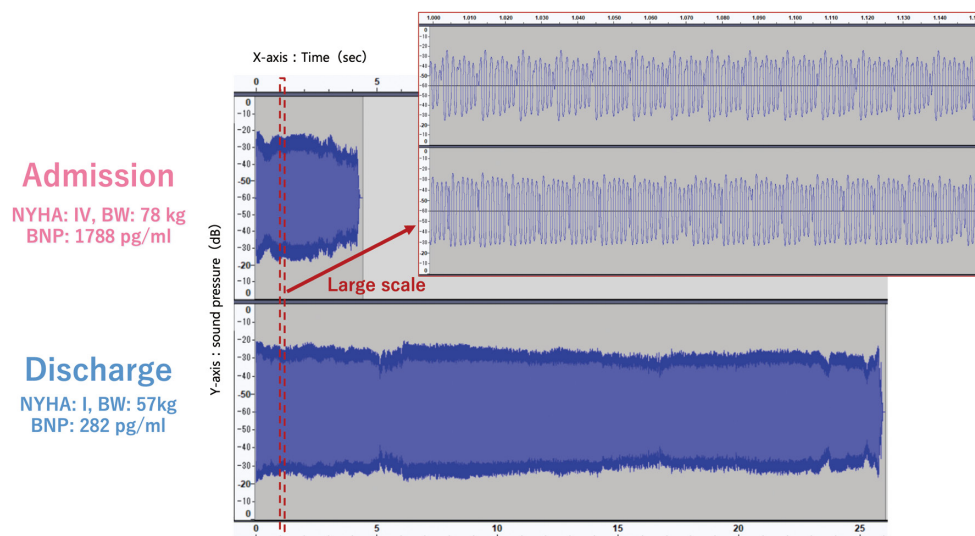
Figure 1. Example of changes in voice symptoms. In response to heart failure (HF) treatments, New York Heart Association (NYHA) class improved from IV at admission to I at discharge, and body weight (BW) decreased by 21 kg. The duration of sustained vowel sound at admission was 4 s but extended approximately 6-fold to 26 s at discharge. Conversely, B-type natriuretic peptide (BNP) levels decreased to approximately 1/6, from 1,788 pg/mL at admission to 282 pg/mL at discharge.

hand-crafted features that reflect the 'roughness' of the voice were measured based on the dynamics in the power spectrum of the waveform envelope as a novel index. For the intensity-related features, peaks in the waveforms were extracted by calculating the relative maxima in the time series data of intensity values.

Machine Learning (ML) Models
Based on 27 acoustic features extracted, ML models were created using a Light Gradient Boosting Machine (LightGBM), a gradient boosting tree algorithm for classification, to estimate various HF conditions, which included NYHA class ≥2, BNP levels ≥300 pg/mL,[18] presence of pulmonary congestion or pleural effusion on chest X-ray, decompensated HF (NYHA class ≥2 and BNP levels ≥300 pg/mL[18]). To confirm that AI methods would work with the present study's sample size, ML models were also created and tested by a support vector machine (SVM), which is considered the optimal ML method for small-sample studies.[19] Shapley Additive exPlanations (SHAP) values were calculated to interpret the influence of individual acoustic features on HF conditions and improve explainability of the models.[20]

Hyperparameters for the LightGBM classifiers were optimized using the Optuna hyperparameter optimization framework.[21] To avoid overfitting in the ML models, we performed feature selection from 27 acoustic features based on the null importance, which compares the null importance distributions with the actual importance of the features gathered by fitting models on the original target. A 5-fold cross-validation was applied to evaluate model performance. Group 5-fold validation was performed as internal validation, where all data from a given patient were categorized in the test set or training set, but not in both. Receiver operating characteristics (ROC) curve analysis was performed to evaluate diagnostic accuracy of

ML models. Associated sensitivity, specificity, positive predictive value, negative predictive value, area under the curve (AUC), and overall accuracy were calculated.

Statistical Analysis
Statistical analyses were performed with JMP Pro® 16 (SAS Institute Inc., Cary, NC, USA). Data were expressed as mean±standard deviation (SD) or median with interquartile range for continuous variables, and as percentages for categorical variables. Continuous values were compared using the Wilcoxon rank-sum test or t-test. Categorical comparisons were performed using a chi-square test or Fisher's exact test. Associations between continuous variables were investigated using linear regression analysis. Repeated measures correlation was also performed to determine the common within-individual association for paired measures assessed on two or more occasions for multiple individuals. A P value <0.05 was considered statistically significant.

Results
Clinical Characteristics and Clinical Course
The present study enrolled 59 patients with acute decompensated HF. Clinical characteristics and clinical course are summarized in **Table**. The median age was 74 (63–81) years, and 57.6% of patients were male. Most patients had HF with reduced ejection fraction (HFrEF) and first-time HF admissions. The primary underlying cardiac diseases that caused HF were relatively well balanced in distribution. All patients were relieved by HF treatments and discharged in 14 (11–18) days after admission. Key HF medications were introduced during hospitalization as part of the standard HF treatment. Body weights decreased by a median of 6.0 (3.6–9.8) kg in response to HF treatment. NYHA class improved from admission to discharge, with
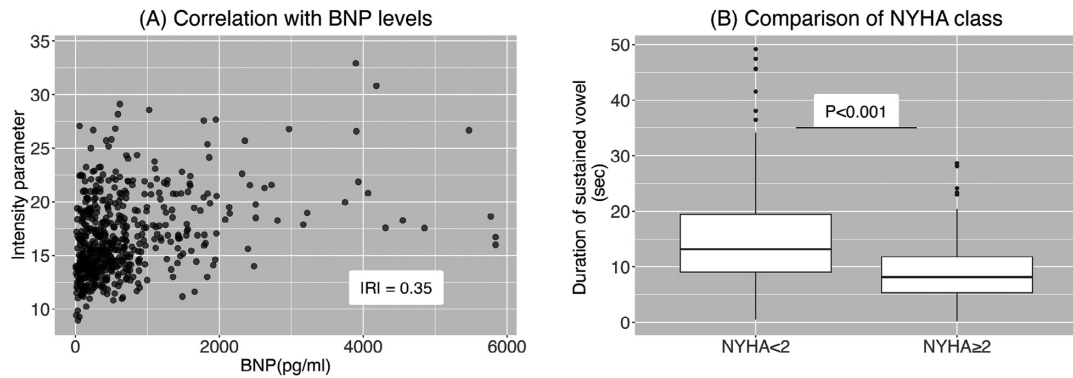
Figure 2. Association of a single acoustic feature with B-type natriuretic peptide (BNP) and New York Heart Association (NYHA). A sound intensity parameter was correlated with BNP levels (**A**), while patients with NYHA class ≥2 had shorter durations of sustained vowel sounds compared with those with NYHA class <2 (**B**).
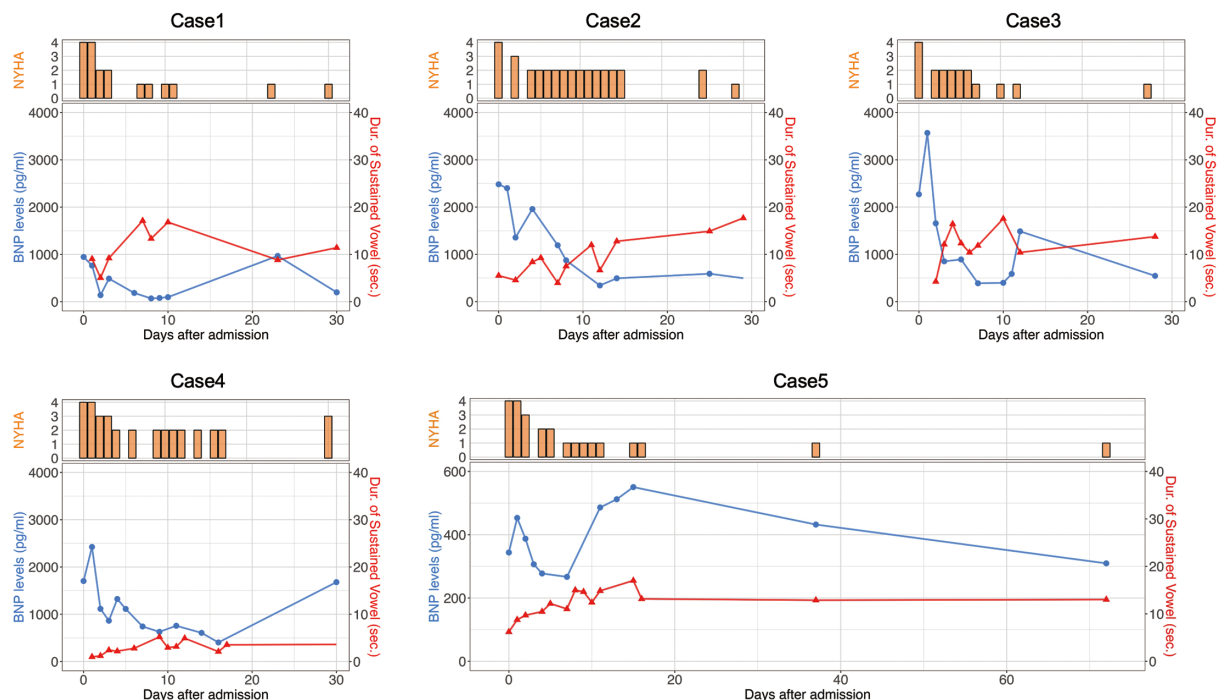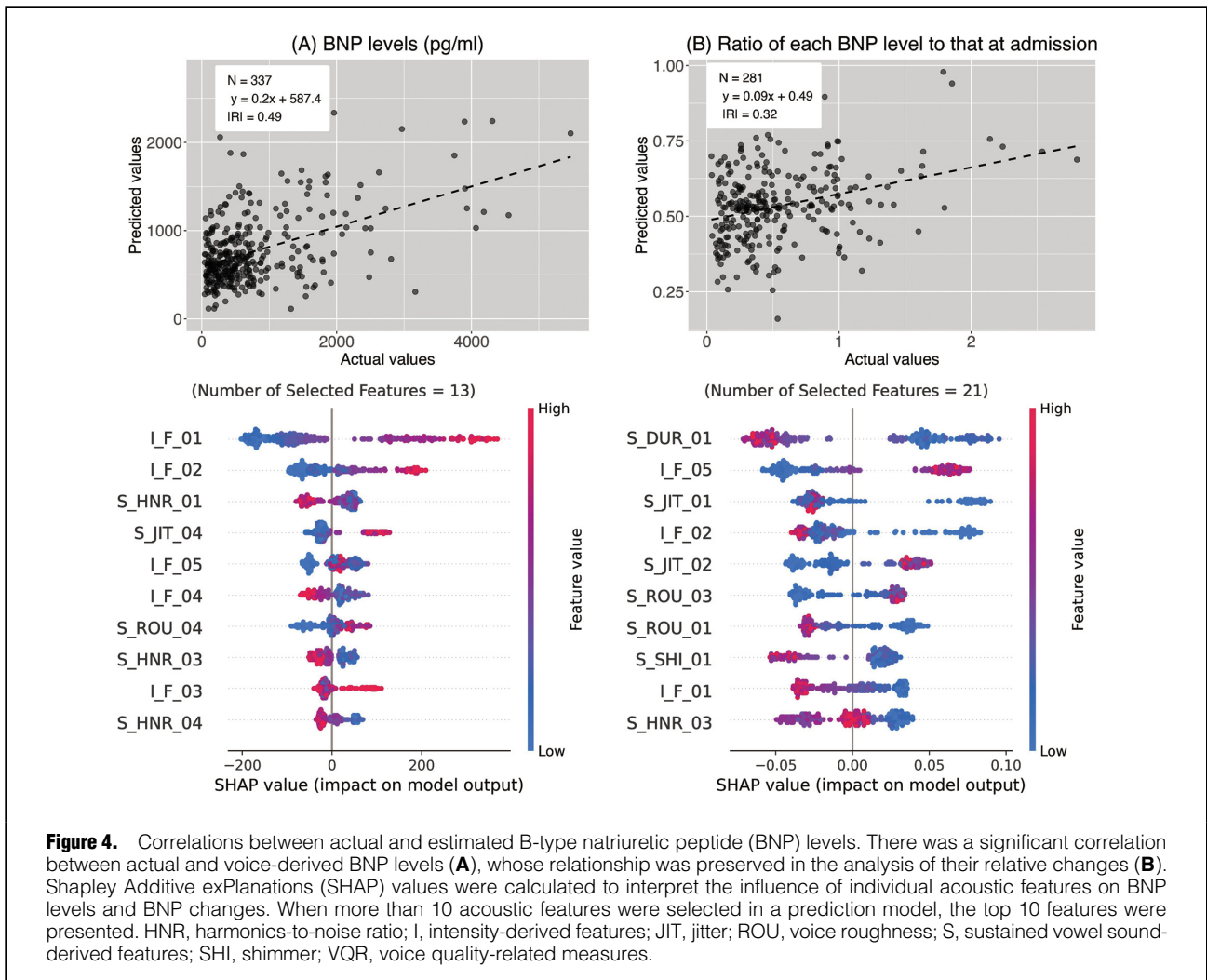

Figure 3. Examples of changes in New York Heart Association (NYHA) class, B-type natriuretic peptide (BNP), and voice symptoms. Data are shown for 30 (**Cases 1–4**) and 75 (**Case 5**) days after hospitalization. The left Y-axis represents BNP levels. The right Y-axis represents the duration of the sustained vowel sound. With heart failure (HF) treatments, NYHA, BNP levels, and sustained vowel sound duration improved overall. Changes in sustained vowel sound duration and BNP and NYHA were relatively well matched, but some cases showed different trends. For example, even with transient BNP increases and residual symptoms, if the duration of sustained vowel sounds was maintained or tended to prolong, the subsequent HF status spontaneously improved (**Cases 1–3** and **5**). Conversely, patients who could not prolong the duration of sustained vowel sounds showed worsening HF characterized by increased BNP levels and NYHA class thereafter (**Case 4**). We were also convinced of the success of HF treatments through improved patients' voices, even if there were no significant changes in conventional HF indicators.

BNP levels decreasing from 968 (610–1,915) pg/mL at admission to 312 (154–543) pg/mL at discharge. LVEF improved over time from admission to discharge and chronic phase.

Based on medical history, patients' HF symptoms began to appear at a median of 25 (10–51) days before admission, and 72.9% of the patients visited other clinics and hospitals 8 (0–28) days after symptom onset and were then admitted 13 (4–29) days later without receiving an appropriate diagnosis, treatments, and follow-up for worsening HF (**Table**).

Figure 4. Correlations between actual and estimated B-type natriuretic peptide (BNP) levels. There was a significant correlation between actual and voice-derived BNP levels (**A**), whose relationship was preserved in the analysis of their relative changes (**B**). Shapley Additive exPlanations (SHAP) values were calculated to interpret the influence of individual acoustic features on BNP levels and BNP changes. When more than 10 acoustic features were selected in a prediction model, the top 10 features were presented. HNR, harmonics-to-noise ratio; I, intensity-derived features; JIT, jitter; ROU, voice roughness; S, sustained vowel sound-derived features; SHI, shimmer; VQR, voice quality-related measures.

Acoustic Features Related to HF Conditions

During hospitalization, a total of 839 audio files from 59 patients (median of 10 [8–17] recording sessions for each patient) was longitudinally obtained and analyzed. Overall, voice symptoms changed over time in response to HF treatments (**Figure 1**). Although commercial and intellectual property concerns preclude detailed descriptions, 27 acoustic features extracted from voice analyses had statistically significant correlations with HF indicators and GOKAN scores (**Supplementary Figure 2**). For example, I_F01, one of the indices related to sound intensity, correlated with BNP levels, while the duration of sustained vowel sound (S_DUR_01) differed significantly between NYHA class ≥2 and <2 (**Figure 2**). In contrast, acoustic features at stable HF symptoms did not differ significantly between the 2 clinical settings (hospitalization and outpatient). These acoustic features were also suggested to complement the interpretation of changes in conventional HF indicators. We observed a certain number of patients who were able to speak with energetic voices even though they had residual symptoms (NYHA ≥2) or observed temporarily elevated BNP levels. In such patients, we could see that HF did not worsen in subsequent clinical courses, and both symptoms and BNP levels spontaneously improved there-

after (**Figure 3**). Conversely, if voice symptoms worsened as BNP levels increased, we often found that HF was likely to be in an exacerbating trend (**Figure 3**).

ML Models for Estimating HF Conditions

Various ML models to estimate HF conditions were created and evaluated. For example, there was a statistically significant correlation between actual and voice-derived BNP levels (r=0.49; P<0.001; **Figure 4**). While standardization of the acoustic features did not further improve the correlation with a similar correlation coefficient (r=0.47; P<0.001) in the present study, this correlation was preserved in multiple linear regression analysis with explained variance scores (variance of BNP explained=19%, adjusted for the other HF conditions and acoustic features) and in the analysis of relative changes between actual and voice-derived BNP levels. The other ML models to estimate NYHA class ≥2 (n=477), BNP ≥300 pg/mL (n=337), presence of pulmonary congestion (n=259) or pleural effusion (n=253) on chest X-ray, and HF decompensation (n=219), also showed good diagnostic accuracies (**Figure 5**), although some variation in the diagnostic accuracy of each indicator was observed due to differences in sample size and characteristics (e.g., NYHA is susceptible to deconditioning
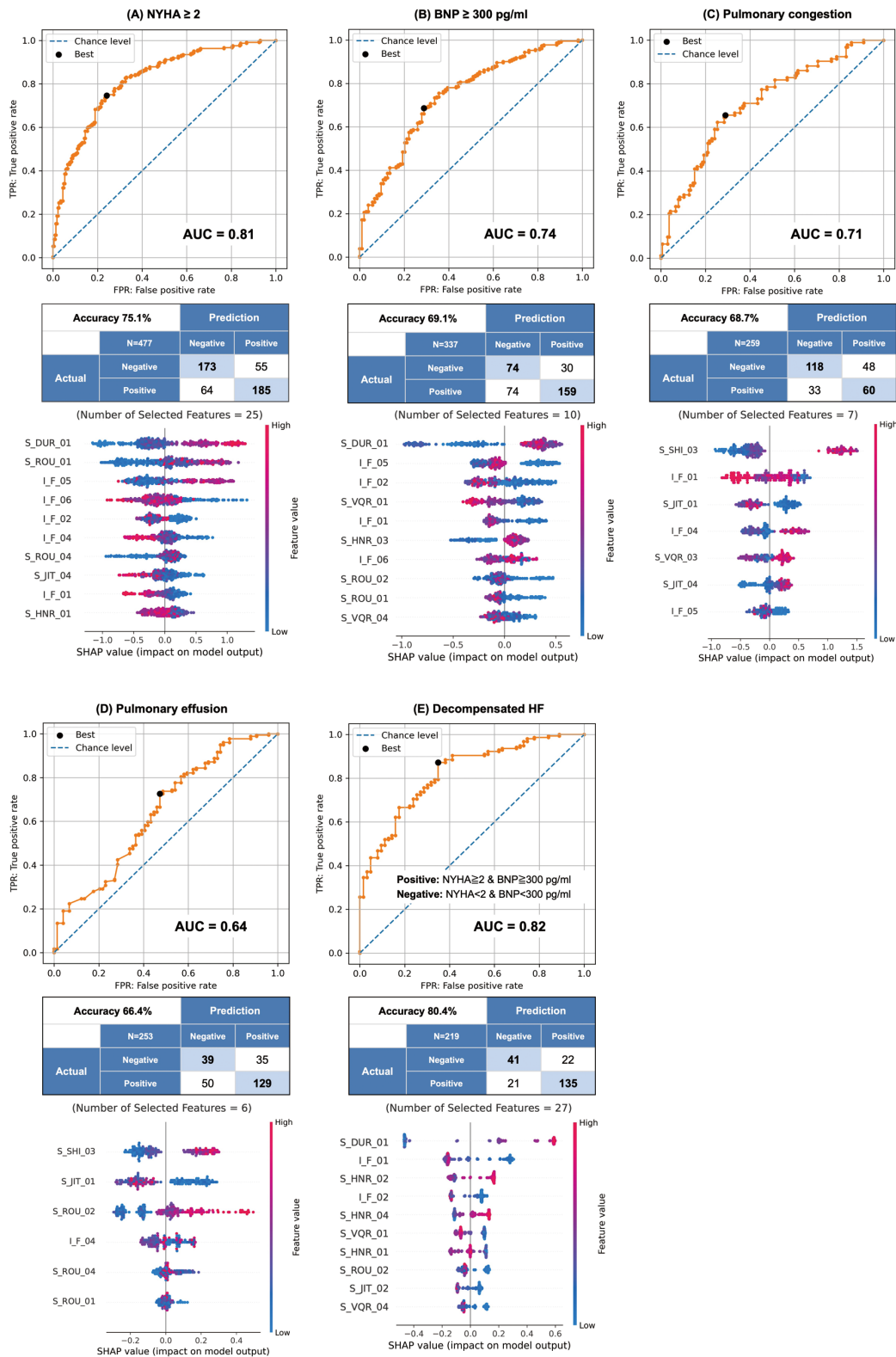
Figure 5. Diagnostic accuracy of ML models to estimate HF conditions. Machine-learning models estimated heart failure (HF) conditions characterized by New York Heart Association (NYHA) class ≥2 (**A**), B-type natriuretic peptide (BNP) ≥300 pg/mL (**B**), presence of pulmonary congestion or pleural effusion on chest X-ray (**C,D**), and worsening HF statuses (decompensated vs. compensated HF) (**E**). Influence of individual acoustic features on HF conditions was expressed using Shapley Additive exPlanations (SHAP). HNR, harmonics-to-noise ratio; I, intensity-derived features; JIT, jitter; ROU, voice roughness; S, sustained vowel sound-derived features; SHI, shimmer; VQR, voice quality-related measures.

during hospitalization, BNP varies widely among individuals). When including voice data in outpatients in the analysis, the ML model estimated NYHA class ≥2 with a similar diagnostic accuracy (sensitivity 70.6%; specificity 75.9%; accuracy 72.2%) to that based on the data of inpatients alone. Similar results were preserved in group 5-fold analyses using separate independent cohorts and analyses by SVM (diagnostic accuracy for inferring HF decompensation: 73.5% and 76.3%, respectively). SHAP value analysis demonstrated that different acoustic features contributed to the model prediction with various degrees of contribution in each model (**Figures 4,5**). For example, intensity-derived features (I_F01 and I_F02) predominantly contributed to the model for BNP-level estimation, while sustained vowel sound-derived acoustic features (S_DUR_01 or S_SHI_03) predominantly contributed to the models for the other HF indicators.

The cutoff point (i.e., classification threshold) of the model could be changed according to clinical purposes. For example, in the ROC analysis of the NYHA class ≥2, sensitivity, specificity, and accuracy calculated at the cutoff point based on the Youden index (where sensitivity and specificity were balanced) were 74.3%, 75.9%, and 75.1%, respectively. In contrast, in a case where a higher-sensitivity model was preferred (e.g., 90.0% for sensitivity), specificity and accuracy were 50.4% and 71.2%, respectively.

Voice Symptoms and Worsening HF
During the follow-up period, seven (11.9%) patients had apparent worsening HF, 6 of whom were rehospitalized for HF. Although the present study was not able to follow patients' drug compliance and lifestyle care after discharge in detail for all patients, 1 patient forgot to take the medication and showed HF exacerbation (not rehospitalization). HF indicators, such as NYHA ≥2 (28.6% vs. 8.0%; P=0.15), BNP levels (404 [171–622] vs. 301 [141–514]; P=0.46), and LVEF (32.9 [26.4–48.2] vs. 34.3 [28.6–50.6]; P=0.71) at discharge, did not differ significantly between patients with or without worsening HF. In contrast, significant differences in voice symptoms were observed between the two groups. For example, patients with worsening HF had a significantly shorter duration of sustained vowel sound at discharge compared with those without)7.47 [3.62–16.5] vs. 14.9 [5.6–45.6]; P=0.03). Conversely, patients who could sustain vowel sound production for a certain time (≥10.62 s) at discharge more frequently observed improvements in LVEF at the chronic phase than those who could not.

Discussion

To the best of our knowledge, this is the first study to systematically evaluate changes in patients' voice symptoms during treatments of acute decompensated HF and to identify HF-related acoustic features that correlated with HF conditions.

HF-Related, Explainable Vocal Biomarkers
Previous studies using language-dependent phrases (e.g., Rainbow passage) or spontaneous speech-based analysis have reported possible differences in acoustic features between admission and discharge in patients hospitalized for acute decompensated HF.[22,23] Another investigation of dialysis patients has reported that pre- vs. post-hemodialysis speech recordings showed significant differences in acoustic

parameters, including shimmer, fundamental frequency, maximum phonation time, and the noise-to-harmonics ratio.[24–26] Although previous studies suggest a potential association between voice alterations and fluid retention, their associations with established HF indicators are yet to be adequately proven;[22–27] further study may help to understand the role and rationale of a vocal biomarker in the management of HF. The answer to the clinically important question of 'whether voice symptoms could change as continuous variables in response to HF conditions' also remains unknown, since all acoustic features of previous studies were extracted from the difference between two extremely distinct conditions (admission and discharge, pre- and post-dialysis).[22–26] Additionally, the spontaneous speech-based, manually processed analyses might have other concerns regarding influence of different languages and bias due to the manual processes. The present study expanded on previous studies by providing particular solutions to the issues highlighted above. For example, the present study used simple, language-independent input tasks to extract HF-related acoustic features, including novel hand-crafted features, and confirmed that changes in these acoustic features were longitudinally correlated with changes in HF conditions and the established HF indicators. Importantly, the present study also showed that vocal biomarkers could be applied even in a population with a different language (e.g., Japanese) from the previous studies. These findings contribute to the accumulation of evidence and complement the results and mechanisms of previous studies. Additionally, it was potentially noteworthy that the present study found the differences in vocal biomarkers at discharge between patients with and without subsequent HF events (i.e., worsening or rehospitalization), as well as improvements in LVEF during follow up, although this was not the primary objective of the study. Furthermore, as our ML models were created based on acoustic features that were extracted from language-independent input tasks and correlated with conventional HF indicators, our results might contribute to the development of explainable AI models for HF monitoring, and could be used on a global scale to monitor HF, but this requires further study.

Potential Utility of Vocal Biomarkers in HF Management
The present study includes clinically important messages. First, vocal biomarkers based on HF-related acoustic features could allow noninvasive, repeatable, continuous monitoring of HF status even at home or in clinics, where blood draws, X-rays, and other tests are limited. Second, the classification threshold of the model could be changed according to clinical purposes. For example, higher-sensitivity models, even at the cost of lower specificity, might be useful in the early detection, or screening, of HF for further testing at hospitals. Conversely, higher-specificity models might enable the reduction of unnecessary tests or hospital visits by ruling out the possibility of HF exacerbations. Third, vocal biomarkers could translate a physician's subjective impression into an objective measure, as well as support diagnosis of HF by complementing the interpretation of conventional HF indicators. In the present study, often we could have been convinced of treatment success or HF exacerbation by changes in a patient's vocal symptoms. Fourth, vocal biomarkers could help enhance the identification of high-risk patients with future HF exacerbation by demonstrating significant differences in vocal biomarkers at discharge between patients with or

without subsequent HF events or improved LVEF. Last, importantly, the present study revealed that HF symptoms appeared a median of 25 (11–51) days before hospitalization and many patients had already visited other clinics and hospitals before hospitalization. The observation reaffirmed that early detection of HF exacerbations followed by appropriate therapeutic intervention may prevent HF rehospitalizations and suggests the potential benefits of vocal biomarkers in HF management.

Study Limitations

Several limitations should be noted. First, this was a hypothesis-generating, single-center analysis with a small sample size and no external validation, in which over 80% of the screened HF population were excluded. Therefore, clinical characteristics in the study population might be different from those with acute HF in general clinical settings. Also, further improvement in diagnostic accuracy may require ML models stratified by each age, sex, and LVEF. Although changes in voice symptoms were found in conjunction with HF indicators in the outpatient setting as well as in the inpatient setting in the present study, this study was within a proof of concept, and our findings cannot be generalized and need to be confirmed in a prospective, multicenter, external validation study enrolling larger populations with appropriate power calculations to ensure the generalizability, robustness and clinical implications. Second, the present study excluded patients with severe respiratory disease such as chronic obstructive pulmonary disease and pneumonia and all voice recordings were performed without oxygen administration. However, in addition to the effect of oxygen therapy on voice analysis, it would be important to confirm the differences in acoustic features between patients with HF and those with other etiology (especially respiratory diseases) to define HF-specific acoustic features. There may also be concerns about effects of mild to moderate respiratory diseases and supplementary oxygen therapy on voice evaluation. Third, we used language-independent simple tasks for the analysis, but our results need to be validated in other populations (races). It is highly likely that articulation (or smoothness and clearness in speaking) can vary significantly between individuals, even with simple input tasks. Fourth, NYHA classification is very subjective and could be affected by many factors other than HF (e.g., deconditioning). BNP levels also vary widely between individuals with HF. Thus, it would be helpful to include a control group with stable HF and pre-HF in the ML model. Last, it is not known how far in advance of HF hospitalization that the vocal biomarkers can infer HF exacerbations and whether there are any synergetic effects when integrating vocal biomarkers with conventional HF indicators, which warrants future studies. We are currently conducting a multicenter study incorporating at least 1,000 HF patients with various conditions, which we hope will address the above limitations.

Conclusions

The present study observed that patients' voice symptoms changed over time in response to HF conditions and was able to successfully extract 27 acoustic features that correlated with conventional HF indices. Although the data is preliminary, ML models based on 27 acoustic features (i.e., vocal biomarkers) have the potential to estimate HF

conditions. Further studies are warranted to confirm our results and address the possible clinical benefits of vocal biomarkers in the early detection of worsening HF.

IRB Information

This study was approved by the Ethics Committee of Yokohama City University (Reference no. B210100055).

Data Availability

The identified participant data will not be shared.

References

1. Jencks SF, Williams MV, Coleman EA. Rehospitalizations among patients in the Medicare Fee-for-Service Program. *New Engl J Medicine* 2009; **360:** 1418–1428, doi:10.1056/nejmsa0803563.
2. Goodlin SJ. Palliative care in congestive heart failure. *J Am Coll Cardiol* 2009; **54:** 386–396, doi:10.1016/j.jacc.2009.02.078.
3. Koehler F, Winkler S, Schieber M, Sechtem U, Stangl K, Böhm M, et al. Impact of remote telemedical management on mortality and hospitalizations in ambulatory patients with chronic heart failure. *Circulation* 2011; **123:** 1873–1880, doi:10.1161/circulationaha. 111.018473.
4. Cowie MR, Anker SD, Cleland JGF, Felker GM, Filippatos G, Jaarsma T, et al. Improving care for patients with acute heart failure: Before, during and after hospitalization. *Esc Hear Fail* 2014; **1:** 110–145, doi:10.1002/ehf2.12021.
5. van Riet EE, Hoes AW, Wagenaar KP, Limburg A, Landman MAJ, Rutten FH. Epidemiology of heart failure: The prevalence of heart failure and ventricular dysfunction in older adults over time. A systematic review. *Eur J Heart Fail* 2016; **18:** 242–252, doi:10.1002/ejhf.483.
6. Miyoshi A, Nishii N, Okamoto Y, Fujita S, Kawamoto K, Okawa K, et al. Lifestyle modification or medication to improve condition of patients with asymptomatic heart failure: Monitoring and Management of OptiVol Alert to Reduce Heart Failure Hospitalization II (MOMOTARO II) Study. *Circ J* 2020; **84:** 456–462, doi:10.1253/circj.CJ-19-0986.
7. Tsuchihashi-Makaya M, Matsuo H, Kakinoki S, Takechi S, Kinugawa S, Tsutsui H, et al. Home-based disease management program to improve psychological status in patients with heart failure in Japan. *Circ J* 2013; **77:** 926–933, doi:10.1253/circj. CJ-13-0115.
8. Spall HGCV, Rahman T, Mytton O, Ramasundarahettige C, Ibrahim Q, Kabali C, et al. Comparative effectiveness of transitional care services in patients discharged from the hospital with heart failure: A systematic review and network meta-analysis. *Eur J Heart Fail* 2017; **19:** 1427–1443, doi:10.1002/ejhf.765.
9. Tsuchihashi M, Tsutsui H, Kodama K, Kasagi F, Setoguchi S, Mohr M, et al. Medical and socioenvironmental predictors of hospital readmission in patients with congestive heart failure. *Am Heart J* 2001; **142:** 20A–26A, doi:10.1067/mhj.2001.117964.
10. Koehler F, Koehler K, Deckwart O, Prescher S, Wegscheider K, Kirwan B-A, et al. Efficacy of telemedical interventional management in patients with heart failure (TIM-HF2): A randomised, controlled, parallel-group, unmasked trial. *Lancet* 2018; **392:** 1047–1057, doi:10.1016/s0140-6736(18)31880-4.
11. Völler H, Bindl D, Nagels K, Hofmann R, Vettorazzi E, Wegscheider K, et al. The first year of noninvasive remote telemonitoring in chronic heart failure is not cost saving but improves quality of life: The randomized controlled CardioBBEAT Trial. *Telemed J E Health* 2022; **28:** 1613–1622, doi:10.1089/tmj.2022.0021.
12. Koehler F, Koehler K, Deckwart O, Prescher S, Wegscheider K, Winkler S, et al. Telemedical Interventional Management in Heart Failure II (TIM-HF2), a randomised, controlled trial

investigating the impact of telemedicine on unplanned cardio-vascular hospitalisations and mortality in heart failure patients: Study design and description of the intervention. *Eur J Heart Fail* 2018; **20:** 1485–1493, doi:10.1002/ejhf.1300.

13. Ong MK, Romano PS, Edgington S, Aronow HU, Auerbach AD, Black JT, et al. Effectiveness of remote patient monitoring after discharge of hospitalized patients with heart failure: The Better Effectiveness After Transition–Heart Failure (BEAT-HF) randomized clinical trial. *JAMA Intern Med* 2016; **176:** 310, doi:10.1001/jamainternmed.2015.7712.

14. Chaudhry SI, Mattera JA, Curtis JP, Spertus JA, Herrin J, Lin Z, et al. Telemonitoring in patients with heart failure. *New Engl J Medicine* 2010; **363:** 2301–2309, doi:10.1056/nejmoa1010029.

15. Koehler F, Winkler S, Schieber M, Sechtem U, Stangl K, Böhm M, et al. Impact of remote telemedical management on mortality and hospitalizations in ambulatory patients with chronic heart failure. *Circulation* 2011; **123:** 1873–1880, doi:10.1161/circulationaha.111.018473.

16. Sagiroglu S, Doganer A. The effect of electrolyte balance on the voice in hemodialysis patients. *Eur Arch Otorhinolaryngol* 2018; **275:** 2755–2761, doi:10.1007/s00405-018-5098-x.

17. Tsutsui H, Ide T, Ito H, Kihara Y, Kinugawa K, Kinugawa S, et al. JCS/JHFS 2021 guideline focused update on diagnosis and treatment of acute and chronic heart failure. *Circ J* 2021; **85:** 2252–2291, doi:10.1253/circj.CJ-21-0431.

18. Kasahara S, Sakata Y, Nochioka K, Yamauchi T, Onose T, Tsuji K, et al; CHART-2 Investigators. Comparable prognostic impact of BNP levels among HFpEF, borderline HFpEF and HFrEF: A report from the CHART-2 study. *Hear Vessel* 2018; **33:** 997–1007, doi:10.1007/s00380-018-1150-4.

19. Shin KS, Lee TS, Kim H. An application of support vector machines in bankruptcy prediction model. *Expert Syst Appl* 2005; **28:** 127–135, doi:10.1016/j.eswa.2004.08.009.

20. Lundberg S, Lee SI. A unified approach to interpreting model predictions. *ArXiv* 2017, doi:10.48550/arxiv.1705.07874.

21. Akiba T, Sano S, Yanase T, Ohta T, Koyama M. Optuna: A next-generation hyperparameter optimization framework. *Arxiv* 2019, doi:10.48550/arxiv.1907.10902.

22. Amir O, Abraham WT, Azzam ZS, Berger G, Anker SD, Pinney SP, et al. Remote speech analysis in the evaluation of hospitalized patients with acute decompensated heart failure. *JACC Hear Fail* 2022; **10:** 41–49, doi:10.1016/j.jchf.2021.08.008.

23. Murton OM, Hillman RE, Mehta DD, Semigran M, Daher M, Cunningham T, et al. Acoustic speech analysis of patients with decompensated heart failure: A pilot study. *J Acoust Soc Am* 2017; **142:** EL401–EL407, doi:10.1121/1.5007092.

24. Sagiroglu S, Doganer A. The effect of electrolyte balance on the voice in hemodialysis patients. *Eur Arch Otorhinolaryngol* 2018; **275:** 2755–2761, doi:10.1007/s00405-018-5098-x.

25. Henry B, Royston TJ. A multiscale analytical model of bronchial airway acoustics. *J Acoust Soc Am* 2017; **142:** 1774–1783, doi:10.1121/1.5005497.

26. Unver S, Hardal U, Esertas K, Sezen A, Celikbilek F, Altundag A. Objective analysis of voice changes in a hemodialysis session and its correlation with ultrafiltration. *Ren Fail* 2015; **37:** 268–272, doi:10.3109/0886022x.2014.988108.

27. Maor E, Perry D, Mevorach D, Taiblum N, Luz Y, Mazin I, et al. Vocal biomarker is associated with hospitalization and mortality among heart failure patients. *J Am Heart Assoc* 2020; **9:** e013359, doi:10.1161/jaha.119.013359.

Supplementary Files