Interpretable AI Model-Based Predictions of ECG changes in COVID-recovered patients

Anubha Gupta
SBILab, Department of ECE
IIIT-Delhi, India
anubha@iiitd.ac.in

Jayant Jain
SBILab, Department of ECE
IIIT-Delhi, India
jayant17155@iiitd.ac.in

Shubhankar Poundrik
SBILab, Department of ECE
IIIT-Delhi, India
shubhankar17194@iiitd.ac.in

Manu Kumar Shetty
Department of Pharmacology
Maulana Azad Medical College, Delhi India
shetty080@gmail.com

Girish M. P.
Department of Cardiology
G.B.Pant Institute of Post Graduate Medical Education, Delhi, India
mpgirish_1999@yahoo.com

Mohit D. Gupta
Department of Cardiology
G.B.Pant Institute of Post Graduate Medical Education, Delhi, India
drmohitgupta@yahoo.com

***Abstract*—COVID-19 has caused immense social and economic losses throughout the world. Subjects recovered from COVID are learned to have complications. Some studies have shown a change in the heart rate variability (HRV) in COVID-recovered subjects compared to the healthy ones. This change indicates an increased risk of heart problems among the survivors of moderate-to-severe COVID. Hence, this study is aimed at finding HRV features that get altered in COVID-recovered subjects compared to healthy subjects. Data of COVID-recovered and healthy subjects were collected from two hospitals in Delhi, India. Seven ML models have been built to classify healthy versus COVID-recovered subjects. The best-performing model was further analyzed to explore the ranking of altered heart features in COVID-recovered subjects via AI interpretability. Ranking of these features can indicate cardiovascular health status to doctors, who can provide support to the COVID-recovered subjects for timely safeguard from heart disorders. To the best of our knowledge, this is the first study with an in-depth analysis of the heart status of COVID-recovered subjects via ECG analysis.**

***Index Terms*—ECG signal analysis, COVID-19, post-COVID subjects, interpretable AI, XAI**

I. Introduction

An electrocardiogram (ECG) is a method to measure the electrical activity of the heart. It is a low-cost, non-invasive, and painless data-capture modality that provides timely measurements of heart functionality. Several ECG features such as heart rate and heart rate variability are critical for early disease diagnosis and prognosis. Doctors study these graphs to identify cardiovascular disorders (CVD). Of late, machine learning (ML) models exploiting ECG data are being developed and utilized actively. Because of the increased prevalence of heart diseases in the present times, there is an immense scope of ML models in CVD diagnosis and prediction. Small sensors are being developed to capture the ECG data while a subject is moving or sleeping. Mobile applications are being developed for the automated assessment of such ECG to help people live healthy lifestyles. Energy-efficient data capture techniques such as compressive sensing are being developed for the ECG data to support telemedicine [1]. While it is easy to utilize such smart devices for monitoring heart rate in day-to-day life, clinical diagnosis requires an interpretable ML model. Today, accuracy and interpretability are two crucial aspects for the acceptance of an ML model by clinicians.

Attempts are being made to develop time-series models to predict the COVID-19 infection rate in order to support the agencies with appropriate policy decisions [2], [3]. Similarly, medical efforts are also underway. Recently, some studies have indicated cardiac problems in patients recovered from COVID [4], [5], while one recent study indicates reduced heart rate variability (HRV) in COVID-recovered subjects [6]. Thus, it is important to track the heart status of subjects recovered from COVID and provide timely assistance for better survival.

HRV is a biomarker for balanced autonomic nervous system [7]. The autonomic nervous system (ANS) controls various body functions such as blood circulation, digestion, and breathing. ANS is divided into two main systems: The sympathetic autonomic nervous system that is activated or stimulated when body's fight or flight response takes effect, and the parasympathetic autonomic nervous system that aims to bring the body to a state of calm [8]. Some studies have confirmed that COVID-19 infection affects the autonomic nervous system that results in post-COVID-19 fatigue and anxiety [9]. The study showed a decrease in heart rate variability in post-COVID vs. control at 3 and 6 months post-discharge. It is observed that viral infections impair the autonomic nervous system. This change indicates increased cardiovascular risk among survivors of COVID [6].

Cardiac abnormalities in Post-COVID can be detected using ECG data, but it requires expertise in the field and is a time-consuming process. Recent advances in AI can help to analyze and interpret ECG data accurately. In general, the analysis of ECG data involves preprocessing of data including

Authors thank the Centre of Excellence in Healthcare, IIIT-Delhi, India.

978-1-6654-0810-3/21/$31.00 ©2021 IEEE

noise removal [10]–[12], feature extraction, normalization, and classification. Traditionally, ECG signals are analyzed using the time domain and the frequency domain features, say HRV features, extracted from the one-dimensional waveforms of different leads. Motivated with the above discussion, this study makes the following significant contributions:

1) ECG data of COVID-recovered and healthy subjects were collected at two hospitals in Delhi, India.
2) Seven ML models have been built to classify healthy versus COVID-recovered subjects.
3) The best performing model was further analyzed to explore the ranking of the altered ECG features in COVID-recovered subjects via AI interpretability.
4) To the best of our knowledge, this is the first study that carries out an in-depth analysis of the heart status of COVID-recovered subjects via AI on ECG data.

II. Methodology

A. Dataset

Data were collected by the Department of Cardiology, G.B. Pant Hospital, Delhi, and Lok Nayak Hospital, Delhi. COVID-19 patients who had recovered (30-60 days after the date of infection) were initially screened for the eligibility criteria. Patients with preexisting cardiac conditions and pathological conditions before COVID-19 infections were excluded from the study. After screening, 117 subjects were eligible for the study. 12 lead, 500 Hz, 60 second ECG data was collected and stored as raw ECG data. These data were recorded during supine paced breathing using VESTA 301i (500 Hz). Similarly, ECG data of 430 healthy subjects recorded in the study [13] at the same hospitals using the same machines were used as the control group data. We removed 25 post-COVID-19 and 4 healthy samples because their ECG data were very noisy. Finally, the ECG data of 94 post-COVID-19 and 426 healthy subjects were included for analysis in the study. Lead II of the 12-lead raw ECG was used for the study. Baseline wander, and R-wave noise were corrected in some samples [11]. After this, feature extraction algorithms and libraries were used to extract various ECG features in the time and frequency domains. Min-Max normalization was executed on the dataset to bring all the values to a common scale of [0,1].

B. Heart Rate Variability Features

We used the lead II of ECG to extract the time-domain and frequency-domain HRV features using the python module (hrv-analysis version 1.0.4). In each ECG signal, R-wave (also called N point) is determined to measure various time-domain features. Features derived from direct measurements of the RR intervals include MEAN HR (mean heart rate), MEAN NNI (mean of NN intervals), MEDIAN NNI (median of the successive difference between NN intervals), and SDNN (standard deviation of the NN intervals). The SDNN is the gold standard to determine cardiac risk and is used to predict morbidity and mortality. Sympathetic and parasympathetic activities contribute to SDNN [14]. RMSSD (root mean square of differences between successive NN intervals) is calculated by measuring the successive difference between two heart rhythms in milliseconds, and the average of the squared sum values is taken. RMSSD is determined mainly by the parasympathetic estimate of short-term components of HRV. RMSSD is a better measure than pNN50 (percentage of successive NN interval greater than 50ms) and pNNi20 (Percentage of successive NN intervals greater than 20ms). For short-term ECG recording, SDNN and RMSDD are good markers of HRV. The normal value of RMSDD is 27± 12ms [14]. RANGE NNI (difference between the maximum and minimum NN intervals) and STD HR (Standard deviation of heart rate) in pathological conditions are typically elevated, and HRV is decreased. CVNNI (Co-efficient of variation equal to the ratio of SDNN divided by mean NN interval) and CVSD (Coefficient of variation of successive difference equal to the RMSSD divided by mean NN interval) are other important features. Frequency domain HRV features HF (High frequency), LF (Low Frequency), and VLF (Very Low Frequency) are distinguished in a short-term recording of 2 to 5 min of ECG [15]. HF (high-frequency power) is also called the respiratory band because HF increases during inspiration and decreases during expiration. It is determined by parasympathetic activity. HF is similar to time domain features pNNI50, and RMSSD [16]. HFNU is the normalized high-frequency power. LF (low-frequency power) is produced by parasympathetic activity, sympathetic activity, and baroreceptors. LFNU is the normalized low-frequency power value. VLF (very low frequency) strongly correlates with the time-domain feature SDNNI. Stress response and physical activity modulate the VLF amplitude, and frequency [17]. LF/HF RATIO (ratio of low frequency and high-frequency power) determines the ratio of sympathetic and parasympathetic activity. A high LF/HF ratio indicates the sympathetic over-activity typically seen in fight-or-flight behaviors, and a low value reflects the dominant parasympathetic system.

C. Models

We built seven machine learning models/methods for the binary classification of post-COVID versus healthy listed as: 1) random forest (with 100 trees), 2) support vector machine (SVC), 3) XGBoost (with 'logloss' evaluation metric), 4) naive Bayes, 5) *k*-nearest neighbors, 6) catBoost (100 estimators and 0.1 learning rate), and 7) logistic regression.

D. Training and Testing

The samples of both classes were divided into five folds. Seven classifiers (or models) were built to test each fold of the data for each of the ML methods listed in the previous subsection. For each model, the training and testing were carried out independently, where one fold was taken as the test data, while the remaining folds were used as the training data. To handle class imbalance in this training data, the random oversampling technique was employed that selects samples randomly from the minority class with replacement, and adds them to the training data. It synthetically generated data points corresponding to the post-COVID minority class to equal the

TABLE I
DESCRIPTIVE STATISTICS (MEAN±STD DEV) OF ECG FEATURES

	Total (n=520)	Post COVID (n=94)	Healthy (n=426)	p-value
LF/HF RATIO (%)	0.81±1.11	1.13±2.15	0.74±0.68	0.002
HF (Hz)	457.17 ±1600.38	477.13±3090.46	452.77±1018.6	0.89
HFNU (%)	64.24±20.63	62.20 ±22.16	64.69±20.27	0.28
LFNU (%)	35.75±20.63	37.79±22.16	35.30±20.27	0.28
STD HR (Heart Rate)	2.68±2.07	2.21±2.07	2.79±2.06	0.014
SDNN (milliseconds)	28.26±24.38	19.73±22.53	30.14±24.40	0.0001
LF (Hz)	220.76±632.87	221.5 ± 1110.2	220.6±468.38	0.99
TOTAL POWER (Hz)	718.05±2276.27	735.76 ±4361.65	714.15±1471.21	0.93
VLF (%)	40.11±122.57	37.13±176.05	40.7±107.53	0.79
MEAN NNI (milliseconds)	766.88±120.45	703.19±114.83	780.93±117.20	$\leq$ 0.0001
MEDIAN NNI (milliseconds)	766.61±122.15	702.13±115.5	780.84±119.06	$\leq$ 0.0001
RANGE NNI (milliseconds)	91.00±83.17	66.04±80.7	96.51±82.79	0.001
CVSD (Co-efficient of variation)	0.03±0.03	0.02±0.04	0.04±0.03	0.001
CVNNI (Co-efficient of variation)	0.03±0.02	0.02±0.03	0.03±0.02	0.001
MEAN HR (Heart rate/min)	80.35±12.84	87.73±14.53	78.72±11.85	$\leq$ 0.0001

number of minority class samples to the number of majority class samples. Otherwise, the model would have developed a bias in favor of the majority class. In other words, only the training folds were oversampled during the cross-validation. This was done to ensure that the test folds did not contain repeated samples and every sample was treated as the test sample only once for reporting the results of cross-validation.

E. Evaluation Metrics

Each of the seven ML models was evaluated on the basis of AUC score, accuracy, weighted accuracy, and MCC score. Accuracy and AUC scores are accurate estimators of performance in the balanced datasets. But they can be misleading in the evaluation of imbalanced datasets. Since our dataset is imbalanced, MCC score and weighted accuracy metrics were chosen because they are better indicators of performance for imbalanced datasets. The number of class-1 (post-COVID or COVID-recovered) samples identified correctly is called True Positive (TP), and the number of class-0 (healthy) samples identified correctly is called True Negative (TN). Similarly, the number of class-1 samples identified incorrectly is called False Negative (FN), and the number of class-0 samples identified incorrectly is called False Positive (FP).

$$\text{Weighted accuracy} = \frac{TP}{2(TP+FN)} + \frac{TN}{2(TN+FP)},$$

$$\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}},$$

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}.$$

AUC is the area under the ROC (receiver operating characteristic) curve that is a plot of the variation of True Positive Rate (TPR) with the False Positive Rate (FPR), where $TPR = \frac{TP}{TP+FN}$ and $FPR = \frac{FP}{FP+TN}$. The best performing ML model was chosen form the seven classifiers and feature ranking was carried out on this using SHAP analysis [18].

III. RESULTS AND ANALYSIS

All HRV features are expressed as mean ± SD (Table-1). The mean of every HRV feature was compared across two groups using the unpaired t-test. A feature with a p-value of ≤ 0.05 was considered to be statistically significant (Table 1). LF/HF RATIO, STD HR, SDNN, Mean NNI, Median NNI, Range NNI, CVSD, CVNNI, and Mean HR were observed to have $p \leq 0.05$, indicating these to be statistically significantly changed in post-COVID compared to healthy subjects. Confusion matrices obtained from each ML method are shown in Fig. 1. Various metrics of ML methods are computed to ascertain their performance. From Table- II, XGB classifier is observed to perform best in classifying healthy and post-COVID persons using the HRV features.

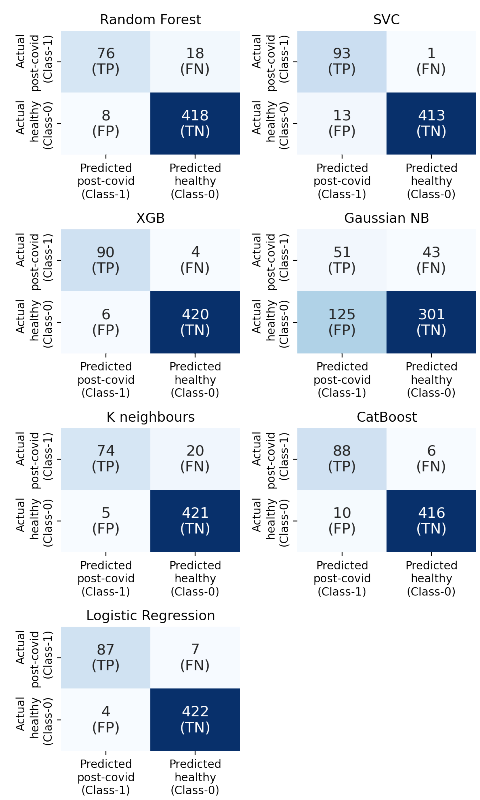

Fig. 1. Confusion matrix obtained from each ML method

Next, feature ranking was performed using SHAP analysis on the trained XGB classifier to determine the HRV features

TABLE II
CLASSIFICATION RESULTS USING DIFFERENT ML METHODS

	AUC	Accuracy	Wt. Accuracy	MCC
Random Forest	0.895	0.95	0.8949	0.8258
SVM	0.9795	0.9731	0.9794	0.9159
XGB	0.9719	0.9808	0.9717	0.9357
Naive Bayes	0.6262	0.6769	0.6246	0.2026
K neighbors	0.8871	0.9519	0.8877	0.8314
CatBoost	0.9564	0.9692	0.9563	0.8981
Logistic	0.9579	0.9788	0.9581	0.9279

influenced most by the COVID-19 infection. Each HRV feature is observed to play different importance in classifying the Post-COVID class in XGB classifier (Fig. 2. Overall, the LF/HF ratio, High-frequency power (HF), normalized high-frequency power (HFNU), standard deviation of heart rate, and low-frequency power were are found to be more important. These features were followed by total power, mean NN interval, mean heart rate, and standard deviation of successive NN intervals as seen from Fig. 2.

Low frequency indicates the mixture of sympathetic and parasympathetic activity. In long-duration ECG recordings, it indicates sympathetic activity. Since the ECG recordings used for classification in this paper are of relatively short duration, LF is of less importance [14]. Conversely, high frequency reflects the beat-to-beat variability of parasympathetic activity. A major contributor of HF is vagal activity [19]. Earlier studies have reported that vagal stimulation increases in the post-COVID recovery phase [20]. This is also evident in our interpretable ML model. An increase in HF indicates a better recovery phase. The LF/HF ratio was obtained by dividing the LF by the HF. This feature represents the balance of the autonomic nervous system (ANS). A higher value of this ratio indicates an imbalance in the ANS. This also leads to behavioral change, presenting itself as an increase in fight-or-flight behaviors. For the ratio value to return to its normal range, the ANS must return to normal. Therefore, a long-term follow-up is required [14]. This imbalance is reported in recent studies conducted in COVID patients and is recorded even 60 days after recovery [21]. This feature is visible in our interpretable model because the red dots that denote the higher ratio values are concentrated towards the right of the central line in Fig. 2 for the LF/HF ratio. The ranking of features according to importance from highest to lowest using SHAP analysis is: LF/HF ratio, HF, HFNU, STD HR, LFNU, total power, mean NNI, mean HR, SDNN, CVSD, VLF, CVNII, range NNI, median NNI, and LF.

A recent study has shown that HRV is reduced in post-COVID patients after 3- and 6-month follow-up [6]. Our results, as seen in Fig. 2, are consistent with this. Higher STD HR indicates the healthy state or younger age, while lower STD HR is seen in diseases, stress conditions, or in ANS imbalance conditions [22]. This is clearly interpretable from our model. Lower STD HR (in blue color) is on the right side of the central line. Similarly, other HRV features like mean NNI and SDNN are higher in healthy subjects than post-COVID subjects. This is evident in the SHAP summary plot [18]. Some features of lower ranks do not contribute significantly to the classification. For these, the distribution of the two colors (blue and red) is not discernible as both colors are present on both sides of the line, with neither being dominant on either side.

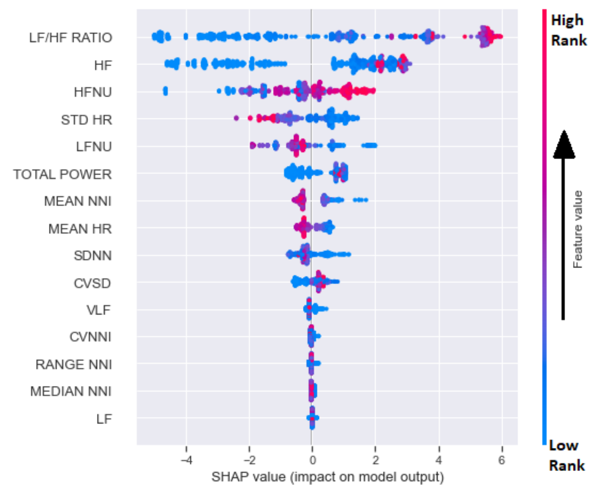

Fig. 2. SHAP summary plot using XGBoost ML model. Each point on a feature line is a SHAP value of a sample. The y-axis represents feature ranking in descending order. The x-axis represents SHAP value. The right-side (+ve SHAP value) of the central line indicates the post-COVID-19 class. A greater +ve SHAP value indicates higher impact on the prediction of the Post-COVID class. LF/HF ratio has the highest +ve SHAP value; therefore, it has the greatest impact on the post-COVID class. The color represents the value of the feature from low (blue) to high (red). For example, in the case of LF/HF ratio, red color is on the right side; therefore, a higher value of LF/HF ratio has a higher impact on the prediction of the post-COVID class.

IV. DISCUSSION AND CONCLUSION

Our interpretable XGBoost classifier with 97% weighted accuracy and a 93% MCC score clearly distinguishes the ECG features of Post-COVID-19 and healthy subjects. From the study, it is inferred that there is a significant difference between the HRV feature of healthy and moderate-to-severe Post-COVID subjects. Thus, a regular follow-up is required in post-COVID patients to monitor the cardiac status and early intervention to prevent complications. By interpreting the HRV features extracted from data collected by wearable technologies (such as smartwatches or fitbits) using our model, it is possible to monitor the cardiac condition of post-COVID patients at home.Moreover, our interpretable AI model can help develop cardiologists' trust to use the AI model in monitoring patients' cardiac status with less invasive or higher investigations like ECHO.

REFERENCES

[1] N. Ansari and A. Gupta, "WNC-ECGlet: Weighted non-convex minimization based reconstruction of compressively transmitted ECG using ECGlet," *Biomedical Signal Processing and Control*, vol. 49, pp. 1–13, 2019.

[2] P. Singh and A. Gupta, "Generalized SIR (GSIR) epidemic model: An improved framework for the predictive monitoring of COVID-19 pandemic," *ISA transactions*, 2021.

[3] P. Singh, A. Singhal, B. Fatimah, and A. Gupta, "An improved data driven dynamic sird model for predictive monitoring of covid-19," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 8158–8162, IEEE, 2021.

[4] A. Gasecka, M. Pruc, K. Kukula, N. Gilis-Malinowska, K. J. Filipiak, M. J. Jaguszewski, and L. Szarpak, "Post-covid-19 heart syndrome," *Cardiology journal*, vol. 28, no. 2, pp. 353–354, 2021.

[5] J. Hall, K. Myall, J. L. Lam, T. Mason, B. Mukherjee, A. West, and A. Dewar, "Identifying patients at risk of post-discharge complications related to covid-19 infection," *Thorax*, vol. 76, no. 4, pp. 408–411, 2021.

[6] T. E. Adler, L. Norcliffe-Kaufmann, R. Condos, G. Fishman, D. Kwak, N. Talmor, and H. Reynolds, "Heart rate variability is reduced 3- and 6-months after hospitalization for covid-19 infection," *Journal of the American College of Cardiology*, vol. 77, no. 18_Supplement_1, pp. 3062–3062, 2021.

[7] J. F. Thayer, F. Åhs, M. Fredrikson, J. J. Sollers III, and T. D. Wager, "A meta-analysis of heart rate variability and neuroimaging studies: implications for heart rate variability as a marker of stress and health," *Neuroscience & Biobehavioral Reviews*, vol. 36, no. 2, pp. 747–756, 2012.

[8] L. K. McCorry, "Physiology of the autonomic nervous system," *American journal of pharmaceutical education*, vol. 71, no. 4, 2007.

[9] L. Townsend, D. Moloney, C. Finucane, K. McCarthy, C. Bergin, C. Bannan, and R.-A. Kenny, "Fatigue following covid-19 infection is not associated with autonomic dysfunction," *PloS one*, vol. 16, no. 2, p. e0247280, 2021.

[10] S. Agrawal and A. Gupta, "Fractal and EMD based removal of baseline wander and powerline interference from ECG signals," *Computers in biology and medicine*, vol. 43, no. 11, pp. 1889–1899, 2013.

[11] S. Agrawal and A. Gupta, "Removal of baseline wander in ECG using the statistical properties of fractional Brownian motion," in *2013 IEEE International Conference on Electronics, Computing and Communication Technologies*, pp. 1–6, IEEE, 2013.

[12] P. Singh, I. Srivastava, A. Singhal, and A. Gupta, "Baseline wander and power-line interference removal from ECG signals using Fourier decomposition method," in *Machine Intelligence and Signal Analysis*, pp. 25–36, Springer, 2019.

[13] M. D. Gupta, A. Bansal, P. G. Sarkar, M. Girish, M. Jha, J. Yusuf, S. Kumar, S. Kumar, A. Jain, S. Kathuria, *et al.*, "Design and rationale of an intelligent algorithm to detect Burnout in Healthcare workers in COVID era using ECG and artificiaL intelligence: The BRUCEE-LI study," *Indian heart journal*, vol. 73, no. 1, pp. 109–113, 2021.

[14] T. F. o. t. E. S. o. C. t. N. A. S. o. P. Electrophysiology, "Heart rate variability: standards of measurement, physiological interpretation, and clinical use," *Circulation*, vol. 93, no. 5, pp. 1043–1065, 1996.

[15] A. Malliani, M. Pagani, F. Lombardi, and S. Cerutti, "Cardiovascular neural regulation explored in the frequency domain.," *Circulation*, vol. 84, no. 2, pp. 482–492, 1991.

[16] P. Grossman and E. W. Taylor, "Toward understanding respiratory sinus arrhythmia: Relations to cardiac vagal tone, evolution and biobehavioral functions," *Biological psychology*, vol. 74, no. 2, pp. 263–285, 2007.

[17] L. Bernardi, F. Valle, M. Coco, A. Calciati, and P. Sleight, "Physical activity influences heart rate variability and very-low-frequency components in holter electrocardiograms," *ACC Current Journal Review*, vol. 3, no. 6, p. 82, 1997.

[18] P. Dehez and V. Ginsburgh, "Approval Voting and Shapley Ranking," *SSRN Electronic Journal*, 2018.

[19] F. Shaffer and J. P. Ginsberg, "An overview of heart rate variability metrics and norms," *Frontiers in public health*, vol. 5, p. 258, 2017.

[20] B. Bonaz, V. Sinniger, and S. Pellissier, "Targeting the cholinergic anti-inflammatory pathway with vagus nerve stimulation in patients with covid-19?," *Bioelectronic medicine*, vol. 6, no. 1, pp. 1–7, 2020.

[21] S. Xiao, D. Luo, and Y. Xiao, "Survivors of covid-19 are at high risk of posttraumatic stress disorder," *Global health research and policy*, vol. 5, pp. 1–3, 2020.

[22] G. Quer, P. Gouda, M. Galarnyk, E. J. Topol, and S. R. Steinhubl, "Inter-and intraindividual variability in daily resting heart rate and its associations with age, sex, sleep, bmi, and time of year: Retrospective, longitudinal cohort study of 92,457 adults," *Plos one*, vol. 15, no. 2, p. e0227709, 2020.