# Demystifying Heart Failure with Mid-Range Ejection Fraction using Machine Learning

Achal Dixit, Soumi Chattopadhyay

Indian Institute of Information Technology Guwahati, India

## Abstract

*Treating Heart Failure (HF) patients with mid-range Ejection Fraction (HFmrEF) is a challenging task due to prognostic uncertainty and transitional behaviour of HFmrEF, often referred to as "grey-area". In this study, we address the uncertainty of HFmrEF through Machine Learning (ML) by classifying it into two well studied phenotypes: HF with preserved Ejection Fraction and HF with reduced Ejection Fraction, using the data from clinical attributes. We propose a semi-supervised Active Learning based model that uses significantly lesser data to tackle the need of supervised label validation and performs on-par with supervised ML models developed for comparison. We believe the use of proposed ML models can enable experts in making informed data-driven decisions leading to the accurate prognosis of* HF *patients.*

## 1. Introduction

Heart failure (HF) is heart's inability to pump an adequate supply of blood to the body. HF is a long-term condition that is potentially life threatening if left untreated. According to U.S. Centers for Disease Control and Prevention, about 6.2 million adults in the United States had HF in 2018 [1]. The HF was reported on 13.4% of the total death certificates issues in 2018 [1]. According to the European Society of Cardiology (ESC), 26 million adults globally are diagnosed with HF. Within the first year itself, 17–45% of the patients suffering from HF die and the remaining die within 5 years [2].

Clinically, HF has two primary subtypes: Heart Failure with Reduced Ejection Fraction(HFrEF) and Heart Failure with Preserved Ejection Fraction (HFpEF), distinguished on the basis of Left Ventricular Ejection Fraction (LVEF) [3,4]. LVEF is a percentage of the volume of blood ejected out of left ventricle in each contraction [4]. HFrEF is classified by LVEF $\leq$ 40% and HFpEF by LVEF $\geq$ 50% [3,4]. However, the HF patients having 40%< LVEF <50% are not accounted well enough through this classification [3].

Classification of HF in itself is a complicated task for clinicians. Therefore numerous researches have either changed the boundary conditions around HFmrEF in their studies or referred to it as a "grey-area" [5,6]. This classification is essential for determining patients' prognosis and treatment but becomes even more challenging due to 5%-10% inter-operator variability in measuring LVEF [7].

Furthermore, patients with HFmrEF had clinical characteristics although intermediate between the HFrEF and HFpEF groups, yet more similar to those of HFpEF [8,9]. However, in the presence of ischemic disease, different studies have found that HFmrEF resembles HFrEF [10,11]. These complex and ambiguous behavior of HFmrEF characteristics motivated our study into developing a data-driven ML-based approach for tackling the uncertainty in prognosis of HFmrEF. The models developed in our study use a combination of clinical attributes such as patient characteristics, clinical test and echocardiogram results of a patient to accurately classify them into the primary phenotypes: HFpEF and HFrEF. We want to explore how well a data-driven approach can perform for resolving the uncertainty of HFmrEF by assigning it to one of the established phenotypes based on the data from clinical attributes.

## 1.1 Literature Review

In the past, only a few studies have been conducted by researchers addressing this problem from ML perspective. In 2013, Austin et al. [12] compared the performance of different classification methods with conventional classification trees and Logistic regression to classify patients with heart failure (HF). However, their main goal was to compare the algorithms and score their predictive ability. Their models achieved area-under-curve between 0.683 - 0.780. In 2015, Alonzo et al. [13] focused on the distinction of HFpEF subtypes using ML techniques. They used 397 HFpEF patients and performed detailed clinical, laboratory and electrocardiographic phenotyping of the participating patients. The study mainly focused on finding a better volumetric discriminator as compared to LVEF. In 2016, Isler [14] performed a heart rate variability(HRV) analysis to distinguish patients with systolic Congestive Heart Failure (CHF) from patients with diastolic CHF. Short-term HRV measures were given as input to train the classifiers. 18 patients with systolic and 12 patients with diastolic CHF were enrolled.

Past research which correlates to the problem addressed in our study does not directly focus on classification of HFmrEF patients. The studies have used different, often conventional methodologies with fewer patients to pursue less similar objectives. The challenge is to classify

## 2. Dataset

The dataset used in our study consists of 495 patients diagnosed with Heart Failure. The published dataset consists of patients retrospectively selected from electronic healthcare records admitted with HF between December 2016 to June 2019 in Zigong Fourth People's Hospital Sichuan, China [15]. The dataset had large number of missing values. As a result, dataset boiled down to 495 samples with 26 features from 2008 samples after cleaning was performed in a meticulous way to retain maximum values across essential features, without imputation, in order to prevent unwanted bias. Finally through feature importance obtained from ML algorithms for selection, 10 features selected for training the models are Pulse, Systolic Blood Pressure, Body Mass Index, Killip Grade, Left Ventricular End-Diastolic Diameter (LVEDD), Brain Natriuretic Peptide (BNP), Creatine Kinase, Cholesterol, Creatinine Enzymatic Method and Potassium. LVEF is used as a target variable encoded into HF classes for our study.

Among 495 patients, 212 are Male and 283 are Female. The number HFpEF patients having LVEF ≥ 50% is 267, HFrEF having LVEF ≤ 40% is 117, and HFmrEF having 41% ≤ LVEF ≤ 49% is 111. Killip Grade is the only categorical feature with 163 in Class I, 248 Class II, 74 in Class III and 10 in Class IV. Killip grade ranging from Class I to IV is used mainly for stratification of patients suffering from acute myocardial infarction [16]. Class I means no sign of Congestive HF, class II depicts the presence S3 gallop or bibasilar rales or both, class III reflects the presence of pulmonary edema and class IV patients suffer from cardiogenic shock [16].

## 3. Methods

Given the similar characteristics of HFmrEF to the two primary phenotypes HFrEF and HFpEF [8-11], we classify HFmrEF into the primary phenotypes. This task can be mapped to a binary classification problem. Binary classification can be performed through supervised ML techniques but it requires labelled data for validation of HFmrEF labels obtained from a supervised learning classifier trained on HFpEF and HFrEF samples. Since, we are exploring to resolve the uncertainty in HFmrEF samples, hence, obtaining the classified HFmrEF labels is not possible because it exactly what we want to do. Additionally, we use unsupervised learning when we do not have any labelled data and we wish to obtain the classes labels from hidden or unknown patterns in data [18]. Therefore, we proceed with a semi-supervised classification approach using Active Learning (AL).

Active Learning is an ML paradigm which is useful when obtaining the labels for data samples is expensive but we still need to find labels for unlabeled data samples [19]. This method allows us to leverage the use of HFrEF and HFpEF samples by learning the label of only a handful data samples which have high classification uncertainty, and thus, obtain HFmrEF labels in a semi-supervised way via primary phenotype class clusters.

Another obstacle is class-imbalance of HFrEF and HFpEF samples in dataset. We resolved the issue of class imbalance by using Adaptive Synthetic Sampling (ADASYN) [17] to oversample the minority class HFrEF to 267, among primary phenotypes, initially having 117 samples. ADASYN sampled dataset not only provides a balanced representation of the data distribution, but it also forces the learning algorithm to focus on difficult to learn samples. ADASYN algorithm adaptively updates the distribution to oversample the minority class based on those classes' data distribution characteristics. After oversampling, training dataset consisted of an equal number of data points in both HFrEF and HFpEF classes with a total of 534 samples, 11 features including the target variable.

## 3.1. Model Development

We developed two Active Learning models using: 1) Random Forest (RF) 2) Multi-Layer Perceptron (MLP), as base estimators simultaneously. Random Forest is an ensemble tree-based discriminative modelling algorithm that fits a number of decision trees on various sub-samples of the dataset [20]. MLP is a feed-forward Artificial Neural Network(ANN) consisting of at least an input layer, a hidden layer and an output layer and can be used for classification and regression problems [21].

A Stream-based query sampling strategy was used to train AL models [22], in which, each unlabeled data sample is examined one at a time for informativeness of that sample. In simpler terms, the model decides whether it can learn enough from knowing the true label of that sample or not. We use classification uncertainty to examine this informativeness or as a "query strategy", which is defined by:

$$U(x) = 1 - P(\hat{x}|x)$$

Where $x$ is the data sample to be predicted, $\hat{x}$ is the most likely prediction and $P(\hat{x}|x)$ is class probability for that sample. In a dataset with $n$ features and $m$ samples we first train the base estimators on $k$ samples ($k < m$) to prevent “cold start problem” [23] based on the size of $m$, where the estimator has a very high bias due to very less initial data samples. The rest of the $(m-k)$ samples act as our unlabeled-training pool. In every learning cycle, we randomly draw samples for $i$ iterations from unlabeled pool and based on the query strategy, classification uncertainty $U(x)$ in our case, the estimator decides to query the label for that sample or not as per the uncertainty threshold $u_c$. The estimator is trained in $k$ samples including the newly queried label. These learning cycles can be repeated while decreasing the uncertainty threshold by $\alpha$ every time.

The base estimators of our AL models were initially trained on $k = 150$ randomly drawn samples (Initial-set) from the training set consisting of HFpEF and HFrEF samples. At least 150 samples initially are necessary to prevent “cold-start problem” [23]. Then, for $i$ = 500 iterations, randomly samples were drawn from the unlabeled-training pool and if the classification uncertainty was below $u_c = 0.4$, the label for that sample was queried from the training set and the estimator was trained. This process of querying was repeated 4 times and the classification uncertainty was reduced by $\alpha = 0.05$ in each AL cycle.

For comparing the classification performance of models trained on HFpEF and HFrFE, we also trained a Random Forest, and Logistic Regression (LR) models on the complete training set.

## 3.1. Model Evaluation

Scoring metrics selected to evaluate the model performance are Receiver Operator Characteristic- Area Under the Curve (ROC-AUC), Accuracy, Precision and Recall. The score were obtained by evaluating the performance through 5-fold cross validation (CV). The dataset is divided into 5 groups or folds of approximately equal size. One fold is held-out while the model is trained on the rest of the 4 folds. The process is repeated 5 times with different hold-out set in each run used for testing. An average of all the 5 scores, for all the metrics, gives us the overall performance of the model. The hold-out sets were preserved through the pipeline and same folds were used to train-test all the models to ensure fairness.

## 4. Results

Table 1. Average Scores of Models on 5-fold CV.

| Model | Accuracy | Roc-Auc | Recall | Precision |
|---|---|---|---|---|
| LR | 85.81% | 86.50% | 84.59% | 90.92% |
| RF | 88.75% | **89.81%** | 85.82% | **93.82%** |
| AL-RF | **88.87**% | 88.89% | **86.88**% | 93.26% |
| AL-MLP | 86.66% | 85.32% | 85.90% | 88.15% |

The dataset consisted of 534 samples of balanced HFpEF and HFrEF classes. For each fold, the dataset was split into 85% training and 15% testing sets. The models were trained on 454 samples and tested on remaining 80 in each run of 5-fold CV. Although the scores of AL model with RF estimator are similar to RF trained directly on all the training data, it must be noted that AL models were trained on average 43% lesser data as compared to RF. Additionally, a supervised classifier cannot be used to predict the labels of HFmrEF due to lack of validation in the absence of labelled data. Still, for comparing the classification performance we have considered including them in our study. The proposed AL based semi-supervised method allows us to have validation from the classification algorithm at the time of querying the unlabeled data. Hence, the AL model can learn using both labelled and unlabeled data, and eventually assign labels to HFmrEF class based on the inherent clusters in the data distribution as explored by Gao. M et.al [24]. Finally, the model classified 111 HFmrEF samples into 66 HFrEF and 45 HFpEF.

The clinical attributes which acted as best predictors are left ventricular end-diastolic diameter(LVEDD) and brain natriuretic peptide. These two features constituted 0.605 importance whereas as other eight features combined had importance of 0.395. The feature importance in RF is calculated as the decrease in node impurity weighted by the probability of reaching that node and higher value depicts higher importance [20]. These importance values are scaled in such a way that their sum is 1.

## 5. Conclusion

As per this study, it is quite evident that the Machine Learning models can perform heart failure classification effectively using clinical attributes. ML models can not only provide a data-driven answer to complex decision problems related to HF classification such as determining the prognosis of HFmrEF patients, which is being actively explored by the research at present. Therefore, the models can inherently help clinicians in making quicker classification and prognosis of HF patients, often inscrutable in HFmrEF and in borderline cases of HFrEF and HFpEF. This study is aimed to contribute to the development and research of ML models which can assist clinicians in making data-driven decisions. We believe that this study will motivate further development of ML models and methods to demystify HFmrEF and HF classification in general.

## References

[1] Salim S. Virani, et al., On behalf of the American Heart Association Council on Epidemiology and Prevention

Statistics Committee and Stroke Statistics Subcommittee "A Report From the American Heart Association", *Heart Disease and Stroke Statistics—2020 Update*

[2] Adam Timmis, Nick Townsend, et al., "European Society of Cardiology, European Society of Cardiology: Cardiovascular Disease Statistics 2019", *European Heart Journal*, Volume 41, Issue 1, 1 January 2020, Pages 12–85.

[3] Ponikowski P, Voors AA, Anker SD, Bueno H, Cleland JG, Coats AJ, et al. "2016 ESC guidelines for the diagnosis and treatment of acute and chronic heart failure: the Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC) developed with the special contribution of the Heart Failure Association (HFA) of the ESC", *Eur Heart J.* 2016;37(27):2129–2200

[4] Lupón J, Bayés-Genís A, "Left ventricular ejection fraction in heart failure: a clinician's perspective about a dynamic and imperfect parameter, though still convenient and a corner stone for patient classification and management" *Eur J Heart Fail*. 2018;20:433–5.

[5] LindenfeldJ., AlbertN.M., BoehmerJ.P. "HFSA 2010 comprehensive heart failure practice guideline". *J Card Fail*. 2010; 16: e1–94.

[6] McMurrayJ.J., AdamopoulosS., AnkerS.D.; "ESC Committee for Practice Guidelines. 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 & Heart Failure Association (HFA) of the ESC" *Eur Heart J.* 2012; 33(14): 1787–847.

[7] Wood PW, Choy JB, Nanda NC, Becher H. Left ventricular ejection fraction and volumes: it depends on the imaging method,*Echocardiography*,2014;31(1):87-100.

[8] Tsuji K, Sakata Y, Nochioka K, Miura M, Yamauchi T, Onose T, et al. "Characterization of heart failure patients with midrange left ventricular ejection fraction-a report from the CHART-2 study", *Eur J Heart Fail*. 2017;19(10):1258–1269.

[9] Cheng RK, Cox M, Neely ML, Heidenreich PA, Bhatt DL, Eapen ZJ, et al. "Outcomes in patients with heart failure with preserved, borderline, and reduced ejection fraction in the Medicare population", *Am Heart J.* 2014;168(5):721–730.

[10] Bhambhani V, Kizer JR, Lima JA, van der Harst P, Bahrami H, Nayor M, et al., "Predictors and outcomes of heart failure with mid-range ejection fraction", *Eur J Heart* Fail. 2018;20(4):651–659.

[11] Kapoor JR, Kapoor R, Ju C, Heidenreich PA, Eapen ZJ, Hernandez AF, et al., "Precipitating clinical factors, heart failure characterization, and outcomes in patients hospitalized with heart failure with reduced, borderline, and preserved ejection fraction", *JACC Heart Fail*. 2016;4(6):464–472..

[12] Austin, Peter C. et al., "Using methods from the data-mining and machine-learning literature for disease classification and prediction: a case study examining classification of heart failure subtypes", *Journal of Clinical Epidemiology*, Volume 66, Issue 4, 398 - 407

[13] Alonso-Betanzos A, Bolón-Canedo V, Heyndrickx GR, Kerkhof PLM. "Exploring Guide-lines for Classification of Major Heart Failure Subtypes by Using Machine Learning. ", *Clinical Medicine Insights: Cardiology*. January 2015.

[14] Yalcin Isler, "Discrimination of systolic and diastolic dysfunctions using multi-layer perceptron in heart rate variability analysis", *Computers in Biology and Medicine*, Volume 76, 2016, Pages 113-119.

[15] Zhang, Z., Zhao, Y., Cao, L., Xu, Z., Chen, R., Lv, L., & Xu, P. (2020). Hospitalized patients with heart failure: integrating electronic healthcare records and external outcome data (version 1.1). *PhysioNet*.

[16] Steven M. Hollenberg, Joseph E. Parrillo, Chapter 23 - Cardiogenic Shock, Editor(s): Jo-seph E. Parrillo, R. Phillip Dellinger, *Critical Care Medicine (Third Edition)*, Mosby, 2008, Pages 423-438, ISBN 9780323048415.

[17] He, Haibo & Bai, Yang & Garcia, Edwardo & Li, Shutao. "ADASYN: Adaptive Synthetic Sampling Approach for Imbalanced Learning", *Proceedings of the International Joint Conference on Neural Networks*, 1322 - 1328. 10.1109/IJCNN.2008.4633969

[18] Khadija El Bouchefry, Rafael S. de Souza ,"Knowledge Discovery in Big Data from Astronomy and Earth Observation", 2020, 12.2.2.

[19] Hasenjäger M., Ritter H. (2002) Active Learning in Neural Networks. In: Jain L.C., Kacprzyk J. (eds) New Learning Paradigms in Soft Computing. Studies in Fuzziness and Soft Computing, vol 84. Physica, Heidelberg.

[20] Pedregosa et al., "Scikit-learn: Machine Learning in Python", *JMLR* 12, pp. 2825-2830, 2011.

[21] Fionn Murtagh, "Multilayer perceptrons for classification and regression", *Neurocomputing*, Volume 2, Issues 5–6,1991, Pages 183-197, ISSN 0925-2312 [5] Houlsby, N., Hernandez-Lobato, J.M., Ghahramani, Z.: Cold-start active learning with robust ordinal matrix factorization. In: *ICML (2014).*

[22] E. Lughofer and M. Pratama, "Online Active Learning in Data Stream Regression Using Uncertainty Sampling Based on Evolving Generalized Fuzzy Models," in *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 1, pp. 292-309, Feb. 2018.

[23] Houlsby, N., Hernandez-Lobato, J.M., Ghahramani, Z.: Cold-start active learning with robust ordinal matrix factorization. In: *ICML (2014).*

[24] Gao M., Zhang Z., Yu G., Arık S.Ö., Davis L.S., Pfister T. (2020) Consistency-Based Semi-supervised Active Learning: Towards Minimizing Labeling Cost. In: Vedaldi A., Bischof H., Brox T., Frahm JM. (eds*) Computer Vision – ECCV 2020. Lecture Notes in Computer Science*, vol 12355. Springer.

Address for correspondence:

Achal Dixit.
Department of Computer Science and Engineering, Indian Institute of Information Technology Guwahati, India, 226101.
achal.dixit@iiitg.ac.in