# A machine learning-based lung ultrasound algorithm for the diagnosis of acute heart failure

Stefano Coiro, Claire Lacomblez, Kevin Duarte, Luna Gargani, Tripti Rastogi, Tahar Chouihed, Nicolas Girerd

# A Machine Learning-based Lung Ultrasound Algorithm For the Diagnosis of Acute Heart Failure

## Short title: ML lung ultrasound diagnostic algorithm for acute heart failure

**Stefano Coiro**[1,2]
**Claire Lacomblez**[2]
**Kevin Duarte**[2]
**Luna Gargani**[3]
**Tripti Rastogi**[2]
**Tahar Chouihed**[4]
**Nicolas Girerd**[2]

[1] **Cardiology Department, Santa Maria della Misericordia Hospital, Perugia**
[2] **Université de Lorraine, Centre D'Investigation Clinique-Plurithématique Inserm CIC-P 1433, Inserm U1116, CHRU Nancy hopitaux de Brabois, F-CRIN INI-CRCT (Cardiovascular and Renal Clinical Trialists), Institut Lorrain du Coeur et des Vaisseaux Louis Mathieu, 4 Rue du Morvan, 54500, Vandoeuvre lès Nancy, France.**
[3]**Department of Surgical, Medical and Molecular Pathology and Critical Care Medicine, University of Pisa., Pisa, Italy**
[4]**Emergency Department, University Hospital of Nancy, Nancy, France**
**INSERM, UMRS 1116, University Hospital of Nancy, Nancy, France**

**E-mail addresses: stefano.coiro@hotmail.it; c.lacomblez@chru-nancy.fr; k.duarte@chru-nancy.fr; luna.gargani@unipi.it; t.gupta@chru-nancy.fr; t.chouihedmahjoub@chru-nancy.fr; nicolas_girerd@yahoo.com.**

**Correspondence:**
**Professor Nicolas Girerd,**
**Service de Cardiologie, Institut Lorrain du Coeur et des Vaisseaux Louis Mathieu, Centre Hospitalier Universitaire de Nancy, 4 Rue duMorvan, 54500 Nancy, France.**
**Tel: +33 3 83 15 74 96; Fax: +33 3 83 15 73 24.**
**Email: nicolas_girerd@yahoo.com**

**Word count (abstract + manuscript text): 22099 characters.**

**This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.**
**All the authors have no relationship with industry to disclose.**

**Abstract (word count: 256)**

**Background:** Lung ultrasound (LUS) is an effective tool for diagnosing acute heart failure (AHF). However, several imaging protocols currently exist and how to best use LUS remains undefined. We aimed at developing a lung ultrasound-based model for AHF diagnosis using machine learning.

**Methods:** Random forest and decision trees were generated using the LUS data (via an 8-zone scanning protocol) in patients with acute dyspnea admitted to the Emergency Department (PLUME study, N=117), and subsequently validated in an external dataset (80 controls from the REMI study, 50 cases from the Nancy AHF cohort).

**Results:** Using the random forest model, total B-line sum (i.e., in both hemithoraces) was the most significant variable for identifying AHF, followed by the difference in B-line sum between the superior and inferior lung areas. The decision tree algorithm had a good diagnostic accuracy [area under the curve (AUC)=0.865] and identified three risk groups (i.e. low 24%, high 70%, and very high-risk 96%) for AHF. The very high-risk group was defined by the presence of 14 or more B-lines in both hemithoraces while the high-risk group was described as having either B-lines mostly localized in superior points or in the right hemithorax. Accuracy in the validation cohort was excellent (AUC=0.91). Importantly, adding the algorithm on top of a validated clinical score and classical definition of positive LUS scanning for AHF resulted in a significant improvement in diagnostic accuracy (continuous net reclassification improvement=1.21, P<0.001).

**Conclusions:** Our simple lung ultrasound-based machine learning algorithm features an excellent performance, and may constitute a validated strategy to diagnose AHF.

**Keywords** Lung ultrasound; machine learning, acute heart failure; diagnostic value.

## Introduction

Signs and symptoms of congestion are a common cause of heart failure (HF) hospitalization [1]. Gradual accumulation or rapid redistribution of intravascular and interstitial fluids (i.e., extravascular lung water) are the main causes of the two most common clinical presentations of acute HF (AHF), namely decompensated HF and pulmonary edema [1, 2].

Lung ultrasound (LUS) imaging has emerged as a simple semi-quantitative method to assess pulmonary congestion in HF patients through B-line quantification [3-5], the diagnostic value of which has been demonstrated in various settings, including pre-hospital, emergency department (ED), intensive care unit, and cardiology/inpatient units [6-12].

Several imaging protocols for LUS have been described, varying from 4- to 28-zones [8]; an Expert Consensus Document on LUS suggested that at least six zones should be examined in HF patients [13]. In an ED setting (N=117), the 8-zone method, followed by the 6-zone method, demonstrated the highest diagnostic accuracy for AHF, also showing a significant incremental diagnostic value over a clinical score [14].

Artificial intelligence (AI), a rapidly evolving field of computer science that mimics human cognition via technology, can be used to improve disease diagnosis [15, 16]. One sub-field of AI is machine learning (ML), which provides computers with the ability to identify patterns and relationships of variables based on given data. Supervised ML approaches, which use labeled datasets with human input to predict outcomes or classify observations, can provide simple or multiple decision trees (i.e., random forest) to classify patients into clinical categories which estimate the risk for a medical condition [17]. By adopting this approach, we can establish an evidence-based framework for utilizing LUS to improve the diagnosis of AHF in an ED setting. This innovative approach serves as a "peacemaker" by leveraging the unbiased nature of AI. Using ML technology, the present study aimed to develop and validate an algorithm capable of identifying AHF based on LUS findings in patients admitted to the ED with acute dyspnea.

## Methods

*Study protocol and design*

A detailed description of the methods is available in the *Online Resource.* This study is part of the prospective Pathway and Urgent caRe of Dyspneic Patient at the Emergency Department in LorrainE District (PURPLE) study (NCT 03194243, CNIL DR- 2017-098)[18]. The derivation cohort was comprised of a subset of patients admitted to the ED in four different hospitals, including a University Hospital, over a 3-month period (Pulmonary Lung Ultrasound for acute dyspnea in eMergency dEpartment, PLUME study).

The validation cohort was constructed by merging two datasets, one including cases (i.e. patients with an established diagnosis of AHF, the Nancy AHF cohort [19]), and the other including controls (the relation between aldosterone and cardiac REmodeling after Myocardial Infarction (REMI) study (controls, treated by primary percutaneous transluminal coronary angioplasty for a first ST-segment elevation myocardial infarction) [20]. All studies were approved by the respective local Ethics Committees.

*Lung ultrasound methods*

All LUS examinations were performed by experienced ultrasound-certified physicians with the following ultrasound equipment and probes: 1) PLUME dataset, Vivid S60 ultrasound (GE HealthCare), 3Sc-RS probe (frequency range 1.3–4.5 MHz); 2) REMI dataset, Vivid E9 ultrasound (GE HealthCare), M5S-D probe (frequency range 1.5–4.5 MHz); 3) Nancy AHF dataset, Vivid 7 ultrasound (GE HealthCare), 3S probe (frequency range 1.5–3.6 MHz). Preset mechanical index (MI) was reduced to ≤0.7 for all examinations, and the scanning time was minimized as short as possible [21]. The focus depth was set at the pleural line level.

All LUS examinations were performed by adopting the 28-zone scanning protocol [19, 22]. For each zone, a B-line semiquantification grading from zero to 10 was adopted, assessed in real time. When a scanning zone displayed only a few separate B-lines (usually less than 5), they were easily discernible and were counted individually. If B-lines were confluent (i.e. not sufficiently distinguishable from one another to be counted individually), the percentage of the area of the screen below the pleural line occupied by B-lines

was considered, and then divided by 10, as previously described and suggested[10, 23]. For the purpose of the present analysis, an 8-zone scanning protocol was derived by reviewing the patients' datasets as described elsewhere [14]. Briefly, among the original 28 scanning zones, four scanning sites (i.e. zones) on each hemithorax were identified (i.e. two anterior zones between the sternum and the anterior axillary line and two lateral zones between the anterior and the posterior axillary line): B-line counts from the superomedial zones of each of the 8 derived scanning sites were used for the analysis (i.e. second and fourth intercostal space at the parasternal and anterior axillary line on each hemithorax) (see Online Resource, *Supplementary Figure 1*).

In order to explore the different approaches to LUS, seven variables were then obtained by either counting the number of B-lines by the 8-zone scanning protocol or by comparing the number of B-lines between scanning zones in terms of absolute or relative differences: 1) total B-line sum (on both hemithoraces), 2-3) B-line sum on each hemithorax considered separately, 4) absolute difference between B-line sum of superior and inferior zones, 5) absolute difference between B-line sum of anterior and lateral zones, 6) absolute difference between B-line sum of right and left hemithorax, and 7) relative difference between B-line sum of right and left hemithorax.

*Statistical methods*

To assess the relevance of the use of LUS in diagnosing AHF, a classification tree was generated based on the derivation cohort (PLUME dataset, N=117) using the Rpart package, which implements the Classification and Regression Tree (CART) algorithm. AHF diagnosis was used as class, and the seven aforementioned variables as predictors. A first complex tree was generated and then streamlined to reduce cross-validation errors (using the Rpart built-in cross-validation). Additionally, to assess the predictive power of the considered variables, a random forest (generating 100 trees) was performed using the randomForest package in the same conditions as the classification tree. One hundred subsets were randomly generated from the PLUME dataset using the bagging method. These subsets were used to build classification trees, after which the results of the models were combined. The discrimination power of each

variable was evaluated using the average gain of purity by splits of a given variable. The risk of AHF was subsequently categorized into three subgroups: a) low risk (i.e. between 22% and 27%, with an average of 24%), b) moderate risk (i.e. between 68% and 75%, with an average of 70%), and c) very high risk (i.e. 96%). The average percentages were determined based on the number of patients in each profile. In addition, in the derivation cohort, the incremental value of our model on top of i) a validated clinical score in an ED setting (the BREST score) [24] and ii) a "classical definition" of positive LUS for AHF diagnosis (i.e. bilateral presence of three or more B-lines in two or more zones) [3, 25] was assessed by C- index, continuous net reclassification improvement (cNRI), and integrated discrimination improvement (IDI). The Brest score, a diagnostic tool able to improve the diagnosis of AHF in patients admitted with undifferentiated dyspnea, is based on eleven variables: age >65, comorbidities [i.e., prior history of HF, myocardial infarction, and chronic obstructive pulmonary disease (COPD)], pattern of dyspnea, ECG findings (atrial fibrillation/flutter, ST segment abnormalities), and signs/symptoms of congestion (i.e., rales and leg edema)[24]. The Brest score ranges from zero to 15, identifying three categories: low (score 0-3), intermediate (score 4-8), and high (score 9-15) probability of AHF. All analyses were performed using R software (the R foundation for Statistical Computing). A P-value $< 0.05$ was considered statistically significant.

**Results**

Of the 117 patients included in the derivation cohort, 73 (62.4%) received a diagnosis of AHF, 23 (19.7%) were diagnosed with pneumopathy, 9 (7.7%) were diagnosed with chronic obstructive pulmonary disease (COPD), and 12 (10.2%) had other etiologies.

*Predictive power of the considered variables*

The random forest model analysis (*Figure 1*) showed that the total B-line sum (i.e. in both hemithoraces) had the most relevant impact on AHF classification (mean impurity decrease =12.3), followed by the B-line sum in the right or left hemithorax (mean impurity decrease 7.7 and 7.6, respectively), while

the difference between B-line sum in the superior and inferior points were also relevant for classification, but with less impurity-based importance comparatively to total B-line sum. Lastly, the differences between right and left hemithorax and anterior and lateral aspect of the lungs were less determinant for patient classification.

*Decision tree for the prediction of acute heart failure diagnosis*

The classification tree identified total B-line-sum, B-line sum in the right hemithorax, and the difference between superior and inferior points as the most relevant variables for correctly classifying patients, confirming the results obtained with the random forest model. The resulting area under the curve (AUC) of 0.865 indicated a good classification of the patients (*Figure 2*). The random forest model had an AUC of 0.754, with an out-of-bag estimated error of 35% (see Online Resource, *Supplementary Figure 2*).

The resulting classification model (*Figure 2*) described three groups at low, high, and very high risk for AHF. According to the model, the very high-risk group was defined by the presence of 14 or more B-lines in both hemithoraces. The low-risk group was characterized by either having a) no B-lines in the right hemithorax or b) having B-lines mostly localized in the inferior points and the left hemithorax. The high-risk group was described as having either B-lines mostly localized in superior points (i.e. difference between B-line sum in superior and inferior points ≥2) or when B-lines were not mostly localized in superior areas, having a relatively high number of B-lines in the right hemithorax (*Figure 2*).

*Description of the risk groups*

Patient characteristics according to the decision tree model-based phenotypes in both the derivation and validation cohorts are presented in *Tables 1* and *2*. In the derivation cohort, the low-risk group (N=45) displayed the most favorable clinical, biological, and radiological profiles. Patients in the very high-risk group (N=45) were more likely to display signs of pulmonary congestion (crackles, pulmonary congestion at chest X-ray), and cardiomegaly (all P-values <0.01) comparatively to those at low risk; lower hemoglobin concentration and higher Brest score were also observed in this group (all P-values <0.05) (*Table 1*). The high-risk group (N=27) mostly displayed intermediate values, which were closer to the high-risk profile than

the low-risk profile. In the validation cohort, patients in the very high-risk group were older and more likely to be male, also displaying significantly higher natriuretic peptide concentrations and lower systolic BP as compared with the other two groups (all P-values <0.05) (*Table 2).*

*Model validation in the external dataset*

To confirm the validity of the generated classification models, the latter were subsequently applied in the validation cohort. The two models showed good performances once applied to the new dataset with an AUC of 0.906 for the classification tree model and 0.918 for the random forest model (see Online Resource, *Supplementary Figure 3, and Figure 2*).

*Incremental diagnostic value of the LUS-based machine learning algorithm*

In the derivation cohort, the addition of our LUS-based ML algorithm on top of the Brest score significantly improved the reclassification of AHF diagnosis in terms of C-index increase, cNRI and IDI (all P-values <0.001) (*Figure 3*). The classical definition of positive LUS also resulted in improved accuracy on top of the Brest score (all P-values <0.01) but of smaller magnitude than that observed for the ML algorithm (i.e., increase in C-index of 0.17 with ML algorithm vs. 0.11 with the classical definition). Furthermore, the ML algorithm significantly improved the reclassification of AHF diagnosis on top of the combined use of the classical definition of positive LUS and the Brest score (C-index increase [95% CI] = 0.07 [0.02 to 0.12], P = 0.006); cNRI [95% CI] = 1.21 [0.91 to 1.51], P < 0.001, IDI [95% CI] = 0.16 [0.09 to 0.23], P <0.001). In contrast, the added diagnostic value of the classical definition of positive LUS on top of the Brest score and the ML algorithm was overall not significant, aside from cNRI (*Figure 3*).

**Discussion**

To the best of our knowledge, the present analysis is the first to provide a LUS-based ML algorithm for the diagnosis of AHF. Our practical algorithm, including a few targeted internal nodes, showed good diagnostic accuracy in the derivation cohort (AUC=0.87), and further confirmed in the validation cohort

(AUC=0.91). Importantly, adding our algorithm on top of a validated clinical score and classical definition of positive LUS for AHF resulted in a significant improvement in diagnostic accuracy. A user-friendly online calculator (https://cic-p-nancy.fr/en/ml_lus_hf/) furthermore facilitates its use on an individual patient basis by clinicians in demanding acute clinical settings.

*Usefulness of machine learning models in heart failure diagnosis*

Machine learning uses a model based on training data to make decisions and program algorithms to solve the problem [26]. Machine learning has been increasingly used in the field of HF research, including classification of HF phenotypes, early diagnosis, prediction of incident HF, risk stratification, optimal titration of medical therapy, as well as patient selection for device implantation and clinical trial recruitment [15].

Generally, available data regarding the utilization of AI and its subfields such as ML in the diagnosis of HF, particularly in acute medical settings, is notably sparse. Baloescu et al. recently compared a 2-zone (anterior-superior) scanning protocol with an 8-zone approach in a cohort of 205 patients admitted to the ED for dyspnea or hypoxia with an automated B-line scoring system using a deep learning algorithm [27]; overall, differences in terms of average score were not significant between the two methods. In another study, when an AI deep learning algorithm was applied to images acquired by LUS novices in a cohort of 29 patients with suspected AHF, B-line quantification had a moderate-to-fair correlation when compared to semi-quantitative analysis by expert review [28]. A secondary analysis of the BLUSHED-AHF study (N=130, LUS-guided treatment strategy versus structured usual care) showed that an AI/ML-based lung congestion score had good agreement with two expert B-line readers (Spearman $r > 0.88$ for both) [29]; additionally, both experts' scores had significantly better agreement with the AI-ML-based score than their correlation with an ultrasound operator's score. Lung ultrasound relies on clinician experience in acquiring and interpreting images. Artificial intelligence software packages embedded in ultrasound equipment allow quantifying B-lines thus reducing subjectivity in the evaluation of LUS patterns [30], ultimately leading to more reproducible analyses. These results underscore the concept that, beyond the number of zones

assessed, a correct interpretation of findings is crucial, and an automated method of B-line quantification could enable non-experts to perform quality LUS examinations. On the other hand, the current limited availability of deep learning software does not allow a widespread application of this methodology at present.

In a setting of stable HF, Choi et al. assessed the diagnostic accuracy of an Artificial Intelligence-Clinical Decision Support System (AI-CDSS) using a hybrid approach (i.e. expert- and ML-driven) for HF diagnosis in a retrospective cohort of 1998 patients [31], compared with that of HF specialists as reference. The concordance rate was excellent both for HF diagnosis and for phenotypic classification (i.e. >91% for all). The AI-CDSS approach confirmed an excellent concordance rate (97%), as opposed to non-HF specialists (76%) in a prospective cohort of 97 dyspneic outpatients. However, such ML approach has never been used for LUS. The currently available algorithms used to diagnose HF with LUS are in effect human/eminence-driven.

*Pathophysiology of B-line appearance and distribution over chest zones in acute heart failure*

In AHF patients, B-line appearance is typically bilateral without showing a right or left side predominance in terms of positive zone (i.e. B-lines ≥3) [32]. However, our approach sheds new light on B-line distribution. According to our decision tree model, a higher B-line sum in the right hemithorax (in patients with a relatively low number of B-lines overall, i.e. <14) compared with the left hemithorax was an indicator of a high risk of AHF. This is in keeping with previous reports described in other settings (i.e. mixed population of inpatients [33] and hemodialysis outpatients [34]. Patients presenting to the ED with acute dyspnea can exhibit various degrees of pulmonary congestion. Milder forms can be identified by LUS only in certain chest areas and B-line distribution may not necessarily follow a typical symmetrical pattern [35]. From a pathophysiological standpoint, a right-sided B-line predominance may be taken as the equivalent of the right-sided predominance of pleural effusions in patients with decompensated HF [36]. While a number of theories have been proposed (i.e. compression of the azygous vein by right ventricular enlargement, predominant right pulmonary vein compression by enlarged atria, difference in lung

dimensions and compliance) [36], no theory has been universally accepted as a plausible explanation for pleural effusion, and this can be presumably extended to B-lines.

A higher B-line sum in superior zones comparatively to inferior zones was another relevant variable according to our decision tree. In the course of AHF, development of pulmonary congestion (and thus B-line appearance) follows a gravitational gradient, by initially involving basal (i.e. inferior) zones [35]. Previous studies that have assessed regional distributions of B-lines in AHF patients have shown that the proportions of positive zones range from 80-90% in superior-middle areas to nearly 100% in inferior areas using an 11-zone protocol [32, 37]. Additionally, if the degree of lung congestion is mild, B-lines can be absent on the more anterior apical zones and be visible only in the most gravitationally dependent region of the thorax [38]. After receiving treatment, B-line clearance was found to be rapid, showing a significant reduction within a few hours [37], with persistence of positive zones observed in $\simeq$ 30% and $\simeq$ 5% of superior-middle areas at 24 h and at pre-discharge, respectively [32, 37]. On the other hand, clearance of the inferior area showed slower kinetics [32, 37] with only 70% of inferior areas cleared at pre-discharge [32].

A possible explanation for our findings in the present study is that a higher B-line sum in superior vs. inferior areas can increase specificity in diagnosing AHF. Indeed, Lichteinstein et al. showed that 50% of patients with exacerbation of COPD and 25% of patients without acute respiratory failure exhibited positive zones laterally, along the last intercostal space above the diaphragm [39]. Thus, a higher B-line sum in superior zones may reflect upper lobe redistribution of lung vasculature resulting in fluid accumulation in the interstitium and interlobular septa. Additionally, in a cohort of 135 stable outpatients with known/suspected HF or coronary artery disease undergoing stress echocardiography, higher B-line counts were found along the 3rd intercostal space (usually included in superior areas in the various LUS protocols) compared with other intercostal spaces (from 2nd to $5^{th}$), both at rest and during stress [40]. It should be emphasized that co-morbidities such as COPD may cause morphological changes in the pulmonary parenchyma leading to a possible asymmetric distribution of hydrostatic edema [35] and thus hamper the ability of LUS to diagnose AHF during acute exacerbations [41].

*Clinical implications*

In the latest ESC Guidelines on HF, LUS was incorporated in the proposed diagnostic work-up as an additional investigation "to confirm AHF diagnosis, especially when natriuretic peptide testing is not available"[1]; however, this indication was not supported by a specific class of recommendations. No mention of the latter has been made by the latest American HF Guidelines [42].

Data from randomized clinical trials and meta-analyses have demonstrated that integration of LUS with clinical assessment for the diagnosis of AHF in an ED setting was more accurate than a "traditional" diagnostic work-up and allowed reducing diagnostic misclassification [8, 43]. We overall confirm previous findings by showing that LUS has good capability in diagnosing AHF. Additionally, by taking advantage of an ML analysis, we first report practical information on the optimal use and interpretation of LUS findings in an acute setting by providing specific details on B-line distribution; this approach showed an incremental diagnostic value over classical definitions of positive LUS based on positive zones.

Indeed, in the setting of acute dyspnea, it may be challenging to differentiate AHF from other diseases such as acute distress respiratory syndrome or pneumonia. Certain LUS findings (i.e. inhomogeneous interstitial pattern, fragmented pleural line, reduced/abolished lung sliding, lung consolidation) are not encountered in AHF and may help point to alternative diagnoses [8]. We provide herein a practical LUS-based algorithm, which exclusively relies on B-line findings (i.e., B-line count and distribution); this "B-line-centered" approach can moreover be widely applied in clinical practice by non-expert practitioners. Indeed, assessment of pleural and parenchymal abnormalities may be difficult to perform by LUS beginners.

*Limitations*

The limited number of patients could potentially limit the generalizability of our results. It should be noted that the detection of B-lines does not necessarily imply their cardiogenic origin, as they are a very sensitive but non-specific sign; they are encountered in all conditions primarily or secondarily affecting the pulmonary interstitium, from acute respiratory syndrome to pulmonary fibrosis [8]. Indeed, like any other diagnostic tool, B-line findings need an integrated assessment with the overall clinical presentation. We

acknowledge that these results need to be validated in further larger studies; in the absence of these validating studies, our work should be considered as hypothesis-generating. In the derivation cohort, the control subgroup exclusively included myocardial infarction patients, and case and control subjects were not matched for age and sex which may have partially influenced LUS diagnostic accuracy.

**Conclusions**

Our practical algorithm, including a few internal nodes (including total B-line sum, but also B-line distribution), showed good diagnostic accuracy in the derivation and validation cohorts (with AUC close or equal to 0.90), and a significantly added accuracy on top of a validated clinical score and classical definition of positive LUS. These findings emerged from very different settings (ED, cardiology ward) thus ensuring generalizability. This ML evidence-based approach may constitute a validated/evidence-based strategy for diagnosing AHF [44].

**Acknowledgments:-.**

**Funding:-.**

**Data availability:** The data underlying this article will be shared upon reasonable request to the corresponding author.

-

# Reference list

1. McDonagh TA, Metra M, Adamo M, Gardner RS, Baumbach A, Böhm M, et al. 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: Developed by the Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC) With the special contribution of the Heart Failure Association (HFA) of the ESC. European Heart Journal. 2021;42(36):3599-726. https://doi.org/10.1093/eurheartj/ehab368.
2. Chioncel O, Mebazaa A, Harjola VP, Coats AJ, Piepoli MF, Crespo-Leiro MG, et al. Clinical phenotypes and outcome of patients hospitalized for acute heart failure: the ESC Heart Failure Long-Term Registry. Eur J Heart Fail. 2017;19(10):1242-54. https://doi.org/10.1002/ejhf.890.
3. Platz E, Jhund PS, Girerd N, Pivetta E, McMurray JJV, Peacock WF, et al. Expert consensus document: Reporting checklist for quantification of pulmonary congestion by lung ultrasound in heart failure. Eur J Heart Fail. 2019;21(7):844-51. https://doi.org/10.1002/ejhf.1499.
4. Kobayashi M, Gargani L, Palazzuoli A, Ambrosio G, Bayés-Genis A, Lupon J, et al. Association between right-sided cardiac function and ultrasound-based pulmonary congestion on acutely decompensated heart failure: findings from a pooled analysis of four cohort studies. Clinical Research in Cardiology. 2021;110(8):1181-92. https://doi.org/10.1007/s00392-020-01724-8.
5. Chouihed T, Coiro S, Zannad F, Girerd N. Lung ultrasound: a diagnostic and prognostic tool at every step in the pathway of care for acute heart failure. Am J Emerg Med. 2016;34(3):656-7. https://doi.org/10.1016/j.ajem.2015.12.030.
6. Lichtenstein DA, Mezière GA. Relevance of lung ultrasound in the diagnosis of acute respiratory failure: the BLUE protocol. Chest. 2008;134(1):117-25. https://doi.org/10.1378/chest.07-2800.
7. Pivetta E, Goffi A, Lupia E, Tizzani M, Porrino G, Ferreri E, et al. Lung Ultrasound-Implemented Diagnosis of Acute Decompensated Heart Failure in the ED: A SIMEU Multicenter Study. Chest. 2015;148(1):202-10. https://doi.org/10.1378/chest.14-2608.
8. Coiro S, Rastogi T, Girerd N. How and When to Use Lung Ultrasound in Patients with Heart Failure? RCM. 2022;23(6). https://doi.org/10.31083/j.rcm2306198.
9. Coiro S, Chouihed T, Girerd N. Lung ultrasound--the extension of clinical examination in patients with acute heart failure: Reply. European journal of heart failure. 2016;18(2):215. https://doi.org/10.1002/ejhf.460.
10. Pellicori P, Platz E, Dauw J, Ter Maaten JM, Martens P, Pivetta E, et al. Ultrasound imaging of congestion in heart failure: examinations beyond the heart. Eur J Heart Fail. 2021;23(5):703-12. https://doi.org/10.1002/ejhf.2032.
11. Gargani L, Frassi F, Soldati G, Tesorio P, Gheorghiade M, Picano E. Ultrasound lung comets for the differential diagnosis of acute cardiogenic dyspnoea: a comparison with natriuretic peptides. Eur J Heart Fail. 2008;10(1):70-7. https://doi.org/10.1016/j.ejheart.2007.10.009.
12. Pivetta E, Goffi A, Nazerian P, Castagno D, Tozzetti C, Tizzani P, et al. Lung ultrasound integrated with clinical assessment for the diagnosis of acute decompensated heart failure in the emergency department: a randomized controlled trial. European Journal of Heart Failure. 2019;21(6):754-66. https://doi.org/10.1002/ejhf.1379.
13. Platz E, Jhund PS, Girerd N, Pivetta E, McMurray JJV, Peacock WF, et al. Expert consensus document: Reporting checklist for quantification of pulmonary congestion by lung ultrasound in heart failure. European journal of heart failure. 2019;21(7):844-51. https://doi.org/10.1002/ejhf.1499.
14. Buessler A, Chouihed T, Duarte K, Bassand A, Huot-Marchand M, Gottwalles Y, et al. Accuracy of Several Lung Ultrasound Methods for the Diagnosis of Acute Heart Failure in the ED: A Multicenter Prospective Study. Chest. 2020;157(1):99-110. https://doi.org/10.1016/j.chest.2019.07.017.
15. Averbuch T, Sullivan K, Sauer A, Mamas MA, Voors AA, Gale CP, et al. Applications of artificial intelligence and machine learning in heart failure. European Heart Journal - Digital Health. 2022;3(2):311-22. https://doi.org/10.1093/ehjdh/ztac025.
16. Kobayashi M, Huttin O, Magnusson M, Ferreira JP, Bozec E, Huby A-C, et al. Machine Learning-Derived Echocardiographic Phenotypes Predict Heart Failure Incidence in Asymptomatic Individuals. JACC: Cardiovascular Imaging. 2022;15(2):193-208. https://doi.org/doi:10.1016/j.jcmg.2021.07.004.
17. Olsen CR, Mentz RJ, Anstrom KJ, Page D, Patel PA. Clinical applications of machine learning in the diagnosis, classification, and prediction of heart failure. American Heart Journal. 2020;229:1-17. https://doi.org/https://doi.org/10.1016/j.ahj.2020.07.009.

18. Central Hospital NF. Pathway and Urgent caRe of Dyspneic Patient at the Emergency Department in LorrainE District (PURPLE). 2023.
19. Coiro S, Porot G, Rossignol P, Ambrosio G, Carluccio E, Tritto I, et al. Prognostic value of pulmonary congestion assessed by lung ultrasound imaging during heart failure hospitalisation: A two-centre cohort study. Sci Rep. 2016;6:39426. https://doi.org/10.1038/srep39426.
20. Huttin O, Mandry D, Eschalier R, Zhang L, Micard E, Odille F, et al. Cardiac remodeling following reperfused acute myocardial infarction is linked to the concomitant evolution of vascular function as assessed by cardiovascular magnetic resonance. Journal of cardiovascular magnetic resonance : official journal of the Society for Cardiovascular Magnetic Resonance. 2017;19(1):2. https://doi.org/10.1186/s12968-016-0314-6.
21. Wolfram F, Miller D, Demi L, Verma P, Moran CM, Walther M, et al. Best Practice Recommendations for the Safe use of Lung Ultrasound. Ultraschall in der Medizin (Stuttgart, Germany : 1980). 2023;44(5):516-9. https://doi.org/10.1055/a-1978-5575.
22. Coiro S, Rossignol P, Ambrosio G, Carluccio E, Alunni G, Murrone A, et al. Prognostic value of residual pulmonary congestion at discharge assessed by lung ultrasound imaging in heart failure. European Journal of Heart Failure. 2015;17(11):1172-81. https://doi.org/10.1002/ejhf.344.
23. Gargani L, Girerd N, Platz E, Pellicori P, Stankovic I, Palazzuoli A, et al. Lung ultrasound in acute and chronic heart failure: a clinical consensus statement of the European Association of Cardiovascular Imaging (EACVI). European heart journal Cardiovascular Imaging. 2023;24(12):1569-82. https://doi.org/10.1093/ehjci/jead169.
24. Basset A, Nowak E, Castellant P, Gut-Gobert C, Le Gal G, L'Her E. Development of a clinical prediction score for congestive heart failure diagnosis in the emergency care setting: The Brest score. Am J Emerg Med. 2016;34(12):2277-83. https://doi.org/10.1016/j.ajem.2016.08.023.
25. Pivetta E, Goffi A, Nazerian P, Castagno D, Tozzetti C, Tizzani P, et al. Lung ultrasound integrated with clinical assessment for the diagnosis of acute decompensated heart failure in the emergency department: a randomized controlled trial. European Journal of Heart Failure. 2019;21(6):754-66. https://doi.org/https://doi.org/10.1002/ejhf.1379.
26. Yasmin F, Shah SMI, Naeem A, Shujauddin SM, Jabeen A, Kazmi S, et al. Artificial intelligence in the diagnosis and detection of heart failure: the past, present, and future. RCM. 2021;22(4):1095-113. https://doi.org/10.31083/j.rcm2204121.
27. Baloescu C, Chen A, Varasteh A, Toporek G, McNamara RL, Raju B, et al. Two- Versus 8-Zone Lung Ultrasound in Heart Failure: Analysis of a Large Data Set Using a Deep Learning Algorithm. Journal of ultrasound in medicine : official journal of the American Institute of Ultrasound in Medicine. 2023. https://doi.org/10.1002/jum.16262.
28. Russell FM, Ehrman RR, Barton A, Sarmiento E, Ottenhoff JE, Nti BK. B-line quantification: comparing learners novice to lung ultrasound assisted by machine artificial intelligence technology to expert review. The Ultrasound Journal. 2021;13(1):33. https://doi.org/10.1186/s13089-021-00234-6.
29. Goldsmith AJ, Jin M, Lucassen R, Duggan NM, Harrison NE, Wells W, et al. Comparison of pulmonary congestion severity using artificial intelligence-assisted scoring versus clinical experts: A secondary analysis of BLUSHED-AHF. European Journal of Heart Failure. 2023;25(7):1166-9. https://doi.org/https://doi.org/10.1002/ejhf.2881.
30. Mento F, Khan U, Faita F, Smargiassi A, Inchingolo R, Perrone T, et al. State of the Art in Lung Ultrasound, Shifting from Qualitative to Quantitative Analyses. Ultrasound Med Biol. 2022;48(12):2398-416. https://doi.org/10.1016/j.ultrasmedbio.2022.07.007.
31. Choi D-J, Park JJ, Ali T, Lee S. Artificial intelligence for the diagnosis of heart failure. npj Digital Medicine. 2020;3(1):54. https://doi.org/10.1038/s41746-020-0261-3.
32. Volpicelli G, Caramello V, Cardinale L, Mussa A, Bar F, Frascisco MF. Bedside ultrasound of the lung for the monitoring of acute decompensated heart failure. Am J Emerg Med. 2008;26(5):585-91. https://doi.org/10.1016/j.ajem.2007.09.014.
33. Jambrik Z, Monti S, Coppola V, Agricola E, Mottola G, Miniati M, et al. Usefulness of ultrasound lung comets as a nonradiologic sign of extravascular lung water. Am J Cardiol. 2004;93(10):1265-70. https://doi.org/10.1016/j.amjcard.2004.02.012.
34. Donadio C, Bozzoli L, Colombini E, Pisanu G, Ricchiuti G, Picano E, et al. Effective and timely evaluation of pulmonary congestion: qualitative comparison between lung ultrasound and thoracic bioelectrical impedance in maintenance hemodialysis patients. Medicine (Baltimore). 2015;94(6):e473. https://doi.org/10.1097/md.0000000000000473.

35. Volpicelli G, Noble VE, Liteplo A, Cardinale L. Decreased sensitivity of lung ultrasound limited to the anterior chest in emergency department diagnosis of cardiogenic pulmonary edema: a retrospective analysis. Critical Ultrasound Journal. 2010;2(2):47-52. https://doi.org/10.1007/s13089-010-0037-0.
36. Natanzon A, Kronzon I. Pericardial and pleural effusions in congestive heart failure-anatomical, pathophysiologic, and clinical considerations. Am J Med Sci. 2009;338(3):211-6. https://doi.org/10.1097/MAJ.0b013e3181a3936f.
37. Cortellaro F, Ceriani E, Spinelli M, Campanella C, Bossi I, Coen D, et al. Lung ultrasound for monitoring cardiogenic pulmonary edema. Intern Emerg Med. 2017;12(7):1011-7. https://doi.org/10.1007/s11739-016-1510-y.
38. Gargani L, Volpicelli G. How I do it: lung ultrasound. Cardiovascular ultrasound. 2014;12:25. https://doi.org/10.1186/1476-7120-12-25.
39. Lichtenstein D, Mezière G. A lung ultrasound sign allowing bedside distinction between pulmonary edema and COPD: the comet-tail artifact. Intensive Care Med. 1998;24(12):1331-4. https://doi.org/10.1007/s001340050771.
40. Scali MC, Zagatina A, Simova I, Zhuravskaya N, Ciampi Q, Paterni M, et al. B-lines with Lung Ultrasound: The Optimal Scan Technique at Rest and During Stress. Ultrasound Med Biol. 2017;43(11):2558-66. https://doi.org/10.1016/j.ultrasmedbio.2017.07.007.
41. Johannessen Ø, Reite FU, Bhatnagar R, Øvrebotten T, Einvik G, Myhre PL. Lung ultrasound to assess pulmonary congestion in patients with acute exacerbation of COPD - a feasibility study. medRxiv. 2022:2022.07.28.22277514. https://doi.org/10.1101/2022.07.28.22277514.
42. Heidenreich PA, Bozkurt B, Aguilar D, Allen LA, Byun JJ, Colvin MM, et al. 2022 AHA/ACC/HFSA Guideline for the Management of Heart Failure: A Report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. Circulation. 2022;145(18):e895-e1032. https://doi.org/doi:10.1161/CIR.0000000000001063.
43. Pivetta E, Goffi A, Nazerian P, Castagno D, Tozzetti C, Tizzani P, et al. Lung ultrasound integrated with clinical assessment for the diagnosis of acute decompensated heart failure in the emergency department: a randomized controlled trial. Eur J Heart Fail. 2019;21(6):754-66. https://doi.org/10.1002/ejhf.1379.
44. Girerd N, Kobayashi M. The new era of evidence-based echocardiographic algorithms using artificial intelligence. International Journal of Cardiology. 2023;380:35-6. https://doi.org/https://doi.org/10.1016/j.ijcard.2023.03.029.

## Figures

**Figure 1**: Variable importance as defined by the random forest algorithm based on the PLUME dataset.

The importance of the seven generated variables as defined by the random forest algorithm, depending on their average gain of purity by splits for the 100 generated trees.

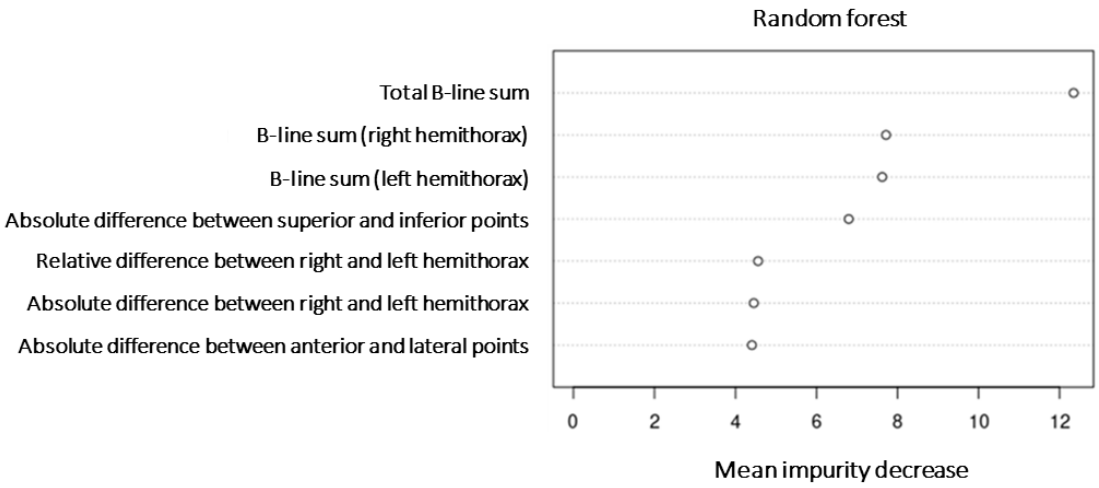
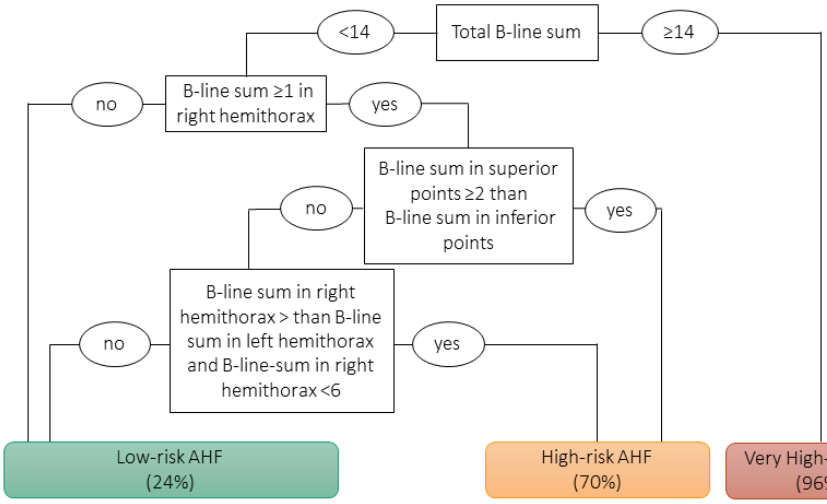


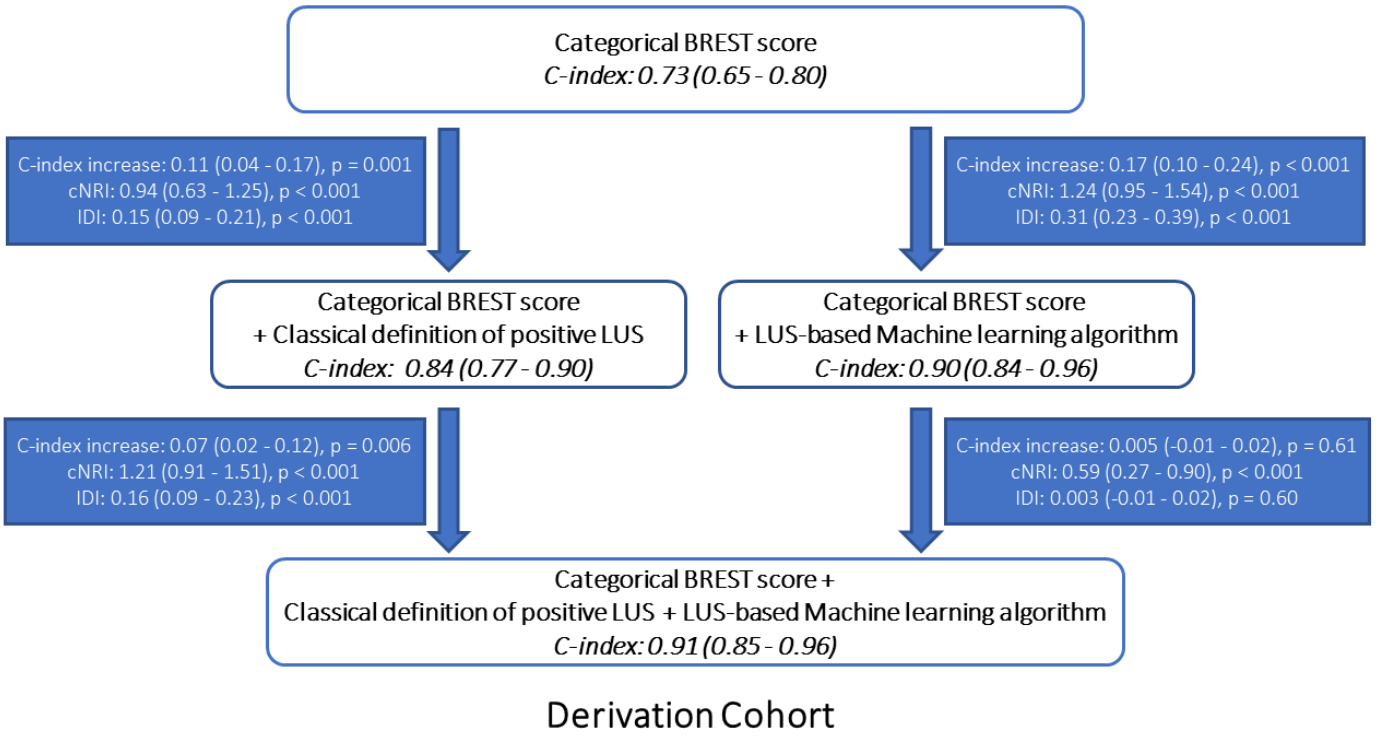
**Figure 2**: Decision tree generated by the CART algorithm from the PLUME dataset

Classification tree model showing relevant variables for the classification of patients in three risk groups (low, high, and very high) for acute heart failure.




| | Low-risk AHF (24%) | | High-risk AHF (70%) | | Very High-risk AHF (96%) | | AUC | Accuracy |
|---|---|---|---|---|---|---|---|---|
| | No AHF | AHF | No AHF | AHF | No AHF | AHF | | |
| Learning dataset (n=117): PLUME Cohort | 34 | 11 | 8 | 19 | 2 | 43 | 0.865 (95% CI 0.798-0.931) | 0.820 |
| Validation dataset (n=130): REMI & Nancy Cohort | 64 | 16 | 16 | 13 | 0 | 21 | 0.906 (95% CI 0.857-0.954) | 0.754 |

**Figure 3**: Description of the incremental value of the LUS-based machine learning algorithm on top of the BREST score and classical definition of positive LUS using C-index, cNRI and IDI.



**Supplemental Information**

**Supplementary Figure 1:** Illustrative figure on how to derive the 8-zone scanning protocol from 28 zones.

**Supplementary Figure 2**: ROC curves for random forest and classification tree models in the derivation cohort

**Supplementary Figure 3**: ROC curves for random forest and classification tree models in the validation cohort

**Table 1:** Phenotype description of acute heart failure risk groups in the derivation cohort (PLUME cohort)

| | Overall cohort | | Low-risk (n=45) | | High-risk (n=27) | | Very high-risk (n=45) | | |
|---|---|---|---|---|---|---|---|---|---|
| **Variables** | **N** | **Mean ± SD n(%)** | **N** | **Mean ± SD n(%)** | **N** | **Mean ± SD n(%)** | **N** | **Mean ± SD n(%)** | **P-value** |
| Clinical characteristics | | | | | | | | | |
| Gender (men) | 1 | 52 (44.4 %) | 45 | 16 (35.6 %) | 27 | 17 (63.0 %) | 45 | 19 (42.2 %) | 0.086 |
| Age (years) | 1 | 79.62 ± 11.76 | 45 | 78.58 ± 11.84 | 27 | 77.81 ± 12.98 | 45 | 81.73 ± 10.83 | 0.30 |
| BMI (kg/m2) | 1 | 27.32 ± 6.90 | 40 | 27.62 ± 6.84 | 25 | 27.02 ± 7.73 | 43 | 27.22 ± 6.61 | 0.94 |
| SBP (mmHg) | 1 | 137.09 ± 25.03 | 45 | 135.53 ± 22.24 | 27 | 131.85 ± 26.27 | 45 | 141.78 ± 26.60 | 0.23 |
| DBP (mmHg) | 1 | 73.09 ± 15.51 | 45 | 70.33 ± 12.56 | 27 | 72.04 ± 16.47 | 45 | 76.49 ± 17.20 | 0.16 |
| Heart rate (bpm) | 1 | 93.56 ± 24.58 | 45 | 90.98 ± 20.93 | 27 | 92.93 ± 24.86 | 45 | 96.51 ± 27.81 | 0.56 |
| Respiratory rate | 1 | 27.32 ± 8.64 | 41 | 28.29 ± 9.71 | 24 | 26.79 ± 8.80 | 41 | 26.66 ± 7.44 | 0.66 |
| SpO2 (%) | 1 | 93.48 ± 5.84 | 45 | 92.16 ± 8.17 | 27 | 93.96 ± 3.56 | 45 | 94.51 ± 3.56 | 0.14 |
| NYHA classification | 1 | | 45 | | 27 | | 45 | | 0.071 |
| I | | 3 (2.6 %) | | 2 (4.4 %) | | 1 (3.7 %) | | 0 (0.0 %) | |
| II | | 4 (3.4 %) | | 1 (2.2 %) | | 3 (11.1 %) | | 0 (0.0 %) | |
| III | | 51 (43.6 %) | | 23 (51.1 %) | | 8 (29.6 %) | | 20 (44.4 %) | |
| IV | | 59 (50.4 %) | | 19 (42.2 %) | | 15 (55.6 %) | | 25 (55.6 %) | |
| Jugular venous-distension | 1 | 19 (16.2 %) | 45 | 3 (6.7 %) | 27 | 5 (18.5 %) | 45 | 11 (24.4 %) | 0.053 |
| Hepato-jugular reflux | 1 | 19 (16.2 %) | 45 | 4 (8.9 %) | 27 | 6 (22.2 %) | 45 | 9 (20.0 %) | 0.23 |
| Lower extremity edema (y/n) | 1 | 64 (54.7 %) | 45 | 20 (44.4 %) | 27 | 18 (66.7 %) | 45 | 26 (57.8 %) | 0.17 |
| Crackles | 1 | 52 (44.4 %) | 45 | 7 (15.6 %) | 27 | 15 (55.6 %) | 45 | 30 (66.7 %) | <0.0001 |
| Laboratory | | | | | | | | | |
| Creatinine (mg/L) | 1 | 13.38 ± 10.03 | 45 | 12.15 ± 10.16 | 27 | 13.51 ± 6.58 | 45 | 14.55 ± 11.56 | 0.53 |
| eGFR MDRD | 1 | 60.11 ± 27.51 | 45 | 67.29 ± 28.52 | 26 | 59.65 ± 28.58 | 45 | 53.20 ± 24.46 | 0.051 |
| Sodium (mmol/L) | 1 | 135.90 ± 5.56 | 44 | 135.25 ± 5.96 | 27 | 136.96 ± 4.29 | 44 | 135.89 ± 5.84 | 0.46 |
| BNP (pg/mL) | 8 | 946.43 ± 1016.76 | 31 | 650.74 ± 1258.27 | 18 | 912.83 ± 958.86 | 37 | 1210.51 ± | 0.075 |
| NT-proBNP (pg/mL) | 1 | 2815.47 ± 3740.83 | 6 | 1912.50 ± | 6 | 2613.83 ± | 3 | 5024.67 ± | 0.53 |
| Potassium (mmol/L) | 9 | 4.47 ± 0.76 | 37 | 4.40 ± 0.71 | 20 | 4.51 ± 0.79 | 42 | 4.52 ± 0.80 | 0.78 |
| Hemoglobin (g/dL) | 1 | 12.36 ± 2.03 | 45 | 12.92 ± 1.95 | 27 | 12.29 ± 2.35 | 45 | 11.84 ± 1.79 | 0.040 |
| Radiology | | | | | | | | | |
| Cardiomegaly | 1 | 68 (58.1 %) | 45 | 18 (40.0 %) | 27 | 17 (63.0 %) | 45 | 33 (73.3 %) | 0.005 |
| Pulmonary congestion | 1 | 59 (50.4 %) | 45 | 10 (22.2 %) | 27 | 18 (66.7 %) | 45 | 31 (68.9 %) | <0.0001 |
| Pleural effusion | 1 | 35 (29.9 %) | 45 | 12 (26.7 %) | 27 | 7 (25.9 %) | 45 | 16 (35.6 %) | 0.62 |
| Pneumonia | 1 | 39 (33.3 %) | 45 | 21 (46.7 %) | 27 | 7 (25.9 %) | 45 | 11 (24.4 %) | 0.069 |
| Brest score | | | | | | | | | |
| Continuous | 1 | 5.93 ± 2.66 | 45 | 4.67 ± 2.44 | 27 | 6.59 ± 2.63 | 45 | 6.80 ± 2.44 | 0.0001 |
| Categorized | 1 | | 45 | | 27 | | 45 | | 0.011 |
| (0-3) | | 22 (18.8 %) | | 14 (31.1 %) | | 5 (18.5 %) | | 3 (6.7 %) | |
| (4-8) | | 75 (64.1 %) | | 28 (62.2 %) | | 16 (59.3 %) | | 31 (68.9 %) | |
| (9-15) | | 20 (17.1 %) | | 3 (6.7 %) | | 6 (22.2 %) | | 11 (24.4 %) | |
| Diagnosis of AHF | | | | | | | | | |
| Emergency Department | 1 | 63 (53.8 %) | 45 | 11 (24.4 %) | 27 | 16 (59.3 %) | 45 | 36 (80.0 %) | <0.0001 |
| At hospitalization discharge | 1 | 73 (62.4 %) | 45 | 11 (24.4 %) | 27 | 19 (70.4 %) | 45 | 43 (95.6 %) | <0.0001 |

BMI, body mass index; SBP, systolic blood pressure; DBP, diastolic blood pressure, SpO2, peripheral oxygen saturation; NYHA, New York Heart Association; eGFR, estimated glomerular filtration rate; MDRD, modification of diet in renal disease, BNP, B-type natriuretic peptide; NT-proBNP, N-terminal proBNP; AHF, acute heart failure; SD, standard deviation.

**Table 2:** Phenotype description of risk groups in the validation cohort (REMI and Nancy AHF cohorts)

| | **Overall cohort (n=130)** | | **Low-risk (n=80)** | | **High-risk (n=29)** | | **Very high-risk (n=21)** | | |
|---|---|---|---|---|---|---|---|---|---|
| | **N** | **Mean ±SD n(%)** | **N** | **Mean ±SD n(%)** | **N** | **Mean ±SD n(%)** | **N** | **Mean ±SD n(%)** | **P-value** |
| Clinical characteristics | | | | | | | | | |
| Age (years) | 130 | 63.14 ± 12.65 | 80 | 60.21 ± 11.05 | 29 | 63.79 ± 13.51 | 21 | 73.38 ± 12.26 | <0.0001 |
| Gender (men) | 130 | 43 (33.1 %) | 80 | 21 (26.2 %) | 29 | 7 (24.1 %) | 21 | 15 (71.4 %) | 0.0005 |
| SBP (mmHg) | 103 | 123.83 ± 19.84 | 70 | 125.07 ± 19.23 | 23 | 126.48 ± 17.61 | 10 | 109.00 ± 24.36 | 0.042 |
| DBP (mmHg) | 103 | 75.19 ± 13.33 | 70 | 76.23 ± 12.86 | 23 | 76.26 ± 13.67 | 10 | 65.50 ± 13.22 | 0.052 |
| Hearth rate (bpm) | 103 | 77.78 ± 16.79 | 71 | 74.44 ± 15.15 | 23 | 78.74 ± 15.52 | 9 | 101.67 ± 13.46 | <0.0001 |
| **Laboratory** | | | | | | | | | |
| BNP (pg/ml) | 128 | 635.82 ± 969.31 | 79 | 341.32 ± 461.28 | 28 | 519.07 ± 679.66 | 21 | 1899.38 ± 1573.10 | <0.0001 |
| Sodium (mmol/L) | 125 | 139.11 ± 2.92 | 75 | 139.41 ± 2.58 | 29 | 138.97 ± 2.91 | 21 | 138.24 ± 3.87 | 0.25 |

SBP, systolic blood pressure; DBP, diastolic blood pressure; bpm, beats per minute; BNP, B-type natriuretic peptide; SD, standard deviation.