

Machine Learning Models for Differentiating Causes of Dyspnea in Emergency Department Patients: A Retrospective Study

Justus Constantin Baarts^{1,2}, Joerg Daniel Leuppi^{2,3}, Maria Boesing^{2,3,*}, Giorgia Lüthi-Corridori^{2-4,*}

¹Infectiology and Hospital Hygiene, USB University Hospital Basel, Basel, BS, Switzerland; ²University Institute of Internal Medicine, Kantonsspital Baselland, Liestal, BL, Switzerland; ³Faculty of Medicine, University of Basel, Basel, BS, Switzerland; ⁴Harvard T.H. Chan School of Public Health – ECPE, Boston, MA, USA

*These authors contributed equally to this work

Correspondence: Giorgia Lüthi-Corridori, Email g.luethi@unibas.ch

Background: Dyspnea is a frequent symptom in emergency departments (ED) with multifactorial causes, including cardiac and pulmonary conditions. Accurate and timely diagnosis is crucial to guide appropriate management and improve patient outcomes. Machine learning (ML) may aid this process.

Methods: This retrospective study analyzed 787 adult patients presenting with dyspnea to the ED at Kantonsspital Baselland (Switzerland) in 2022. Clinical, laboratory, diagnostic data and final diagnoses were collected. ML models including decision trees, random forest and boosted decision trees were trained to classify dyspnea etiologies (cardiac vs respiratory) and predict final diagnoses. Performance metrics included accuracy, sensitivity, and specificity.

Results: The most common diagnoses were decompensated heart failure (28.4%), pneumonia (26.4%), and COVID-19 (17%). Binary classification into cardiac vs respiratory causes achieved the highest performance (accuracy: 89.6% with boosted trees). Multiclass prediction of specific diagnoses yielded lower performance (accuracy: 36.8%). CRP, BNP, and cough emerged as key predictive features consistent with established clinical knowledge, supporting model interpretability. Comorbidities, though clinically relevant, showed limited predictive value.

Conclusion: ML algorithms show promise in supporting triage by distinguishing broad etiological categories of dyspnea, such as respiratory versus cardiac origins. While the models demonstrate useful classification performance, their limited sensitivity for specific diagnoses underscores the need for larger, more diverse datasets and advanced modeling approaches.

Plain Language Summary: People often come to the emergency department with shortness of breath, a symptom also called dyspnea. This can be caused by different health conditions, especially those affecting the heart or lungs. Quickly understanding the cause of dyspnea is important to ensure patients receive the right treatment.

Our research team wanted to find out whether computer tools known as machine learning models could help doctors to better identify what is causing a person's breathing problem. We analyze real-life data from 787 adults who came to the emergency department at Kantonsspital Baselland in Switzerland during 2022. For each person, we collected information like symptoms, lab results, and final diagnoses.

We tested different types of machine learning models to classify whether the cause of a patient's shortness of breath was related to the heart (for example, heart failure) or to the lungs (such as pneumonia or COVID-19). One model, called a boosted decision tree, was able to do this correctly almost 90% of the time. However, predicting the exact diagnosis was more challenging, with a lower accuracy. The most helpful indicators for the models were blood markers of inflammation, a heart-related hormone, and whether the patient had a cough.

These results suggest that machine learning tools could support emergency teams by helping to rapidly distinguish between heart- and lung-related causes of shortness of breath. Further research using larger and more diverse patient groups will be needed to improve the accuracy and usefulness of these tools in clinical practice.

Keywords: dyspnea, emergency department, ED, machine learning, ML, diagnosis prediction, boosted decision tree

Introduction

Dyspnea, or difficulty breathing, is a frequent reason for emergency department (ED) visits, accounting for nearly six million visits in the United States in 2021 alone.¹ Globally, it represents 5% of ED visits, 10% of general ward admissions, and 20% of ICU admissions, with an overall in-hospital mortality of 5%.^{2,3} The causes of dyspnea are diverse, ranging from common conditions such as lower respiratory tract infections, heart failure, COPD exacerbations, and asthma to rarer causes including pulmonary embolism, anaphylaxis, or pneumothorax.^{3–9}

Timely and accurate diagnosis is essential for initiating effective treatment. However, studies report diagnostic discordance rates between 15% and 30%, which may delay treatment and worsen outcomes.^{10–13} Emergency physicians often face significant time pressure, incomplete information, and reliance on heuristics or initial impressions, which can contribute to diagnostic errors.^{12,14–16} Additionally, the broad differential diagnosis and variability in patient presentations (eg, age, comorbidities) further complicate the clinical assessment.

Cardiac and pulmonary causes often overlap, as both can present with similar symptoms such as dyspnea and chest discomfort. Moreover, coexisting conditions are frequent, making diagnosis even more challenging.^{17–21}

Clinicians rely on a combination of medical history, physical examination, vital signs, laboratory tests, and imaging to guide diagnosis. However, comprehensive workups for all patients can increase ED length of stay, resource use, and patient risk.²²

A more targeted approach could improve diagnostic efficiency. Many useful parameters such as triage vitals, presenting symptoms, medical history, and basic lab results are available shortly after patient arrival and could support early risk stratification and decision-making.

Given the complexity and time pressure in the ED, artificial intelligence (AI), particularly machine learning (ML), offers promising support for early diagnostic decision-making. ML can analyze large clinical datasets to predict likely diagnoses based on patterns in routinely collected variables.^{23,24} Two common approaches are multilabel (allowing multiple concurrent diagnoses) and multiclass classification (assigning a single most likely diagnosis), each with advantages and trade-offs in clinical applicability.^{25,26}

Multilabel methods are more faithful to real-world complexity but computationally demanding, whereas multiclass methods simplify decision-making at the risk of overlooking comorbid conditions.

Among ML techniques, decision trees provide interpretability and direct mapping to clinical decision pathways.²⁷ Ensemble approaches, such as random forests and boosting, enhance accuracy by aggregating multiple trees and correcting errors iteratively.²⁸ These methods are well suited to healthcare applications, as they manage mixed-type data, tolerate missingness, and highlight clinically meaningful thresholds.

Recent studies highlight the growing role of ML in acute care. Hale et al provided a foundational overview of ED dyspnea differentials, highlighting the prevalence of serious but initially subtle presentations and underscoring the need for structured, early decision support.⁹

Recent studies have demonstrated the growing role of AI across different medical domains, such as liver disease detection using ML²⁸ and apnea/hypopnea classification through fuzzy expert systems.²⁹ While these examples underscore the broad importance of developing AI-based diagnostic tools, their focus on diseases outside acute dyspnea further highlights the unmet need for similar approaches in the ED context.

Although recent work has demonstrated the potential of deep learning networks for dyspnea classification in the ED,³⁰ these methods typically require large, longitudinal datasets and complex architectures that may not be available in many healthcare systems. This underscores the importance of parallel research employing simpler and more interpretable ML models, which can be trained on routinely collected ED data, provide clinically transparent outputs, and therefore have a higher likelihood of real-world integration across diverse clinical environments.

Despite the promise of AI in acute care, the complexity and data requirements of many approaches limit their widespread adoption. The Swiss healthcare setting, characterized by universal coverage and structured cantonal hospital networks, provides a unique context that may enhance prompt access to diagnostic testing. There is a clear need for simpler, interpretable ML models that leverage routinely collected ED data to support clinicians in diagnosing dyspnea accurately and efficiently. This study aims to evaluate the potential of such ML tools to improve the timeliness and

accuracy of diagnoses in ED patients with dyspnea. Additionally, it seeks to provide insight into real-world patient characteristics and diagnostic challenges in a Swiss cantonal hospital.

Methods

Study Design and Setting

This retrospective observational study analyzed routine clinical data from adult patients presenting with dyspnea at the ED of the Cantonal Hospital of Baselland (KSBL) between January and December 2022. The hospital serves a population of 280,000 in northwest Switzerland.³¹

Eligibility Criteria and Data Source

Patients hospitalized at KSBL and discharged in 2022 with dyspnea as a primary complaint were eligible (n = 822). Cases were identified via DaWiMed software (ID Suisse AG), which searched medical records for mentions of dyspnea and synonyms.³² Manual review supplemented the automated process to include missed cases (false negatives) and exclude false positives or cases lacking a clear cause. Cases were excluded if the patient presented with dyspnea but was in fact hospitalized and treated for a primary diagnosis unrelated to the dyspnea episode. In other words, dyspnea was a concomitant or incidental symptom rather than the chief complaint driving the clinical encounter. For multiclass classification, only patients with a single final diagnosis were included.

Data Collection and Variables

Patient records were manually reviewed by trained investigators (E.D., G.L-C., M.K.M., M.M.S.) using the KISIM[®] clinical information system.). Data were recorded in REDCap[®], a secure platform for clinical research.^{33,34}

Extracted variables included:

- Demographics and comorbidities: age, sex, smoking status, cardiovascular and pulmonary comorbidities.
- Symptoms and severity: acute vs chronic dyspnea, classified according to the New York Heart Association (NYHA) functional classification.³⁵
- Vital signs: heart rate, blood pressure, respiratory rate, oxygen saturation, temperature.
- Laboratory and diagnostic data: routine labs, ECG, imaging, and other tests performed in the ED.
- Outcomes: discharge diagnosis, hospital length of stay, post-discharge procedures, and 1-year rehospitalization.

To ensure data quality, a second investigator independently reviewed a subset of records. Extreme values (top and bottom five) were checked for plausibility.

Statistical Analysis and Machine Learning Models

Feature Selection and Engineering

All extracted clinical and laboratory variables considered clinically relevant were initially included. No automated feature selection, dimensionality reduction, or creation of composite features was performed; instead, we prioritized including all available clinically meaningful predictors. Descriptive statistics were computed as counts and percentages for categorical variables and medians with interquartile ranges for continuous variables. For machine learning (ML), variables with $\leq 40\%$ missing data were imputed using k-nearest neighbor.^{36,37} Classification models were trained on a 70/30 split (training/test sets) using R packages including rpart, randomForest, and xgboost.

Multiclass classification algorithms were built for the following target variables:

1. Final diagnosis (12 classes as seen in [Table 1](#))
2. Diagnosis type: (4 classes: “respiratory”, “cardiac”, “both”, other)
3. Exclusive diagnosis type: (2 classes: “respiratory”, “cardiac”)

Table 1 Final Diagnosis/Outcome

	All (n=787)
Final Diagnosis n (%)	
Pneumonia	208 (26.4)
Pulmonary embolism	51 (6.5)
Asthma exacerbation	28 (3.6)
COPD exacerbation	97 (12.3)
Acute coronary syndrome	91 (11.6)
Decompensated heart failure	244 (28.4)
COVID-19	134 (17)
Respiratory tract infection	51 (6.5)
Arrhythmia	72 (9.1)
Other	127 (16.1)
Outcome n (%)	
Discharged home/Care facility	646 (82)
Other hospital/inpatient rehabilitation	94 (11.9)
Death	48 (6.1)
Length of stay median [IQR] in days	7 [4–10]

The classification for the target variables “Diagnosis type” and “Exclusive diagnosis type” was done as follows:

- Respiratory diagnoses: pneumonia, asthma exacerbation, COPD exacerbation, COVID-19, respiratory tract infections
- Cardiac diagnoses: pulmonary embolism, acute coronary syndrome, decompensated heart failure, arrhythmia
- Other: anaphylactic reaction, panic attack or anxiety, other

In the models built for “Final diagnosis (12 classes)”, cases with more than one final diagnosis were excluded from the analysis. In the models built for “Exclusive diagnosis type (2 classes)”, patients with more than one diagnosis of different types, or diagnosis of type “other” were excluded from the analysis. This step was necessary to allow unambiguous classification during model training and to avoid noise in the outcome labels. However, these exclusions reflect a “best-case” diagnostic scenario. All ML models were built on the imputed data set. For the ML models, data was split into training (70% training set, 30% test set). Data imputation and analysis were performed with R version 4.1.0 using the packages “bnstruct”, “rpart”, “randomForest”, and “xgboost”.³⁸ Class imbalance was present, as some diagnoses (eg, decompensated heart failure, pneumonia) were much more frequent than others (eg, pulmonary embolism, asthma). No resampling was applied; instead, performance metrics were reported per class to highlight this limitation.

Multiclass Classification

Simple Decision Trees

Simple decision trees were built on the training set and plotted. Accuracy (with 95% confidence interval), sensitivity and specificity were evaluated for the test set. To preserve interpretability, no hyperparameter tuning was applied.

Random Forest Models

Random forest models were optimized for the number of trees and the number of features considered at each split node (*mtry*). Hyperparameter tuning was performed using the *tuneRF()* function from the *randomForest* package. Starting from the default *mtry* value (number of predictors/3), candidate values were iteratively adjusted with *stepFactor* = 1.5. At each step, a random forest with *nTreeTry* = 100 trees was grown, and the out-of-bag (OOB) error rate was computed. The search stopped when the relative improvement in OOB error fell below the predefined threshold (improve = 0.01). The *mtry* value that minimized OOB

error was selected, and the final model was trained with $n_{tree} = 150$ trees and the optimized $mtry$ value. Variable importance was assessed via mean decrease in Gini index, and model performance was evaluated on the test set.

Decision Tree Boosting

Boosted decision trees were built with a maximum tree depth of 3, learning rate of 0.1, subsampling rate of 0.6 for rows, and 0.8 for features.

Multilabel Classification

Multilabel classification was performed using a problem transformation approach, where the multilabel problem was converted into multiple binary classification tasks using the “Binary Relevance” method. Model training was conducted on the training dataset, with separate binary classifiers trained for each label. Predictions were made on the test dataset, and performance was evaluated for each label using accuracy, sensitivity, and specificity. Metrics were not averaged across labels. To assess whether the model’s accuracy significantly exceeded that of the No Information Rate (NIR), McNemar’s test was conducted for each label at a significant level of 0.05.

Results

Baseline Characteristics

In 2022, 822 patients were hospitalized via the ED with dyspnea as the primary complaint. After excluding 35 cases with no identified etiology, 787 cases were included. The baseline characteristics are presented in [Table 2](#).

The median age was 76 years (range 15–99), and 56.5% were male. Most cases involved acute dyspnea (85%), while 14% had acute on chronic dyspnea. Chronic dyspnea without exacerbation was rare (0.4%).

Common comorbidities included hypertension (67.6%), heart failure (32.4%), diabetes mellitus (28.1%), coronary artery disease (29.2%), arrhythmia (30.5%), COPD (19.9%), and depression (10.8%). Pulmonary embolism (8.4%), cerebrovascular disease (11.9%), and prior COVID-19 infection (10.7%) were also observed.

Presenting symptoms included cough (42.3%), sputum production (19.3%), chest pain (18.4%), and fever (14.7%). Peripheral edema was noted in 7.5%, and calf pain in 1.1%. Smoking status was documented in 42% of patients; of these, 44% were current smokers, 45% former, and 11% non-smokers.

Vital signs on admission showed that 27% of patients had a systolic blood pressure <120 mmHg, 31% had values between 120–140 mmHg, and 42% exceeded 140 mmHg. Tachycardia (heart rate >100 bpm) was observed in 29%, and oxygen saturation without supplementation was <92% in 27%. The median oxygen flow rate was 3 L/min, and fever $\geq 38^{\circ}\text{C}$ was documented in 15.6% of cases.

The most common diagnostic tests performed included EKG (87.9%), chest X-ray (40%), and thoracic CT (37.8%). Lung ultrasound (4%), echocardiography (5%), and emergency coronary angiography (4.3%) were infrequently used. Lung auscultation was documented in 93% of patients, revealing crackles in 36%, wheezing in 18%, and absent breath sounds in 25%. One-third of patients also had documentation of clinical signs, most commonly peripheral edema (87%) and jugular venous distension (41%).

Availability of diagnostic tests varied: CRP was available in 95% of patients, BNP in 70%, and chest imaging in 40%. Symptoms and clinical findings were less systematically documented, reflecting variability in routine ED practice.

The final diagnosis was most commonly decompensated heart failure (28.4%), followed by pneumonia (26.4%) and COVID-19 (17%). Other frequent diagnoses included COPD exacerbation (12.3%), acute coronary syndrome (11.6%), arrhythmia (9.1%), and pulmonary embolism (6.5%). Notably, 253 patients received multiple diagnoses considered relevant for their dyspnea: 198 had two, 45 had three, and 10 had four diagnoses. Details about final diagnosis can be found in [Table 1](#).

As for outcomes, 82% of patients were discharged home or to a care facility, 12% were transferred to another hospital or inpatient rehabilitation, and 6% died. The median length of stay was 7 days (IQR 4–10). Details about outcomes can be found in [Table 1](#).

Table 2 Baseline Characteristics

	All (n=787)	Missing n (%)
Demographic / Baseline		
Age median [IQR] in years	76 [66;84]	
Gender (males)	445 (56.5)	
BMI median [IQR]	26.9 [23.7;31.6]	380 (48.3)
Type of Dyspnea		
Acute Dyspnea	664 (85.5)	
Acute on chronic Dyspnea	110 (14.2)	
Chronic Dyspnea	3 (0.4)	
Comorbidities n (%)		
Heart failure	255 (32.4)	
Coronary heart disease	230 (29.2)	
Arterial hypertension	533 (67.6)	
Dyslipidemia	251 (31.9)	
Arrhythmia	240 (30.5)	
PAVK	82 (10.4)	
Cerebrovascular disease	94 (11.9)	
COPD	157 (19.9)	
Asthma bronchial	63 (8)	
Pulmonary embolism	66 (8.4)	
Status post COVID-19	84 (10.7)	
Diabetes mellitus	221 (28.1)	
OSAS	83 (10.5)	
Depression	85 (10.8)	
Tumor	163 (20.7)	
Symptoms		
Cough	333 (42.3)	
Sputum production	152 (19.3)	
Nasal congestion	24 (3.1)	
Chest pain	145 (18.4)	
Peripheral edema	59 (7.5)	
Calves pain	9 (1.1)	
Fever	116 (14.7)	
Other	260 (33)	
Smoking status		
Reported	330 (41.9)	457 (58.1)
Current smoking	147 (44.4)	
Former smoking	148 (44.8)	
Nonsmokers	35 (10.6)	
Vital signs at admission		
Systolic blood pressure (in mmHg)		48 (6.1)
< 120 mmHg	199 (26.9)	
120–140 mmHg	232 (31.2)	
> 140 mmHg	308 (41.7)	

(Continued)

Table 2 (Continued).

	All (n=787)	Missing n (%)
Heart rate		47 (6.0)
< 60 bpm	43 (5.8)	
60–100 bpm	482 (65.1)	
> 100 bpm	215 (29.1)	
Oxygen saturation (without oxygen supplementation, n = 556)		2 / 556 (0.4)
< 92%	150 (27.0)	
≥ 92%	404 (72.7)	
Oxygen saturation (with oxygen supplementation, n = 231)		27 / 231 (11.7)
< 92%	49 (21.2)	32/231 (13.9)
≥92%	155 (67.1)	
Amount of oxygen supplied (in l/min), n = 231	3 (2;5)	
Temperature		107 (13.6)
< 38 °C	557 (70.8)	
≥38 °C	123 (15.6)	

Machine Learning

Three machine learning models were used to predict the etiology of dyspnea: Standard Decision Trees, Random Forests, and Boosted Decision Trees. Among them, Boosted Decision Trees performed best in binary classification tasks, achieving an accuracy of 89.6%. Random Forests followed closely at 88%. In contrast, standard Decision Trees resulted in lower performance, particularly for single final diagnosis prediction, with accuracy dropping to 36.8%. Details about models' accuracy can be found in [Table 3](#).

The decision tree models utilized variables such as clinical symptoms, vital signs, and laboratory results. Parameters like CRP, BNP, and cough emerged as key discriminators between cardiac and respiratory causes. [Figure 1](#) illustrates the decision tree for binary classification (respiratory vs cardiac), which included only cases with clearly assignable causes (n=542). Cough was the most important variable, and the model achieved an accuracy of 84.66% (95% CI: 78.2–89.8%).

[Figure 2A](#) and [B](#) illustrate the expanded model to four diagnostic groups: cardiac, respiratory, combined cardiac and respiratory, and other causes. This multi-class model showed a lower accuracy of 57.8% (95% CI: 51.2–64.2%).

The relative importance of variables in the Random Forest model was assessed using the Mean Decrease Gini. CRP was the most important feature (Gini = 27.6), followed by BNP (23.7), cough (21.2), body temperature (16.6), and creatine kinase (14.9).

A third model attempted to predict specific final diagnoses (n=9 categories) or relevant combinations. Diagnoses were grouped based on pathophysiological mechanisms: respiratory infections (COVID-19 and other RTIs), obstructive lung disease (asthma and COPD exacerbations), and ADHF alone or in combination with other conditions. While conceptually informative, the overall classification accuracy of this approach was only 36.8% (95% CI: 30.4–43.6%), indicating significant overlap in clinical features across diagnostic categories.

Table 3 Comparison of Model Accuracy

	Binary Group	Quadruple Group	Single Diagnosis
Standard Decision Tree	0.85	0.58	0.37
Random Forest	0.88	0.64	0.50
Boosted Decision Tree	0.89	0.69	0.49

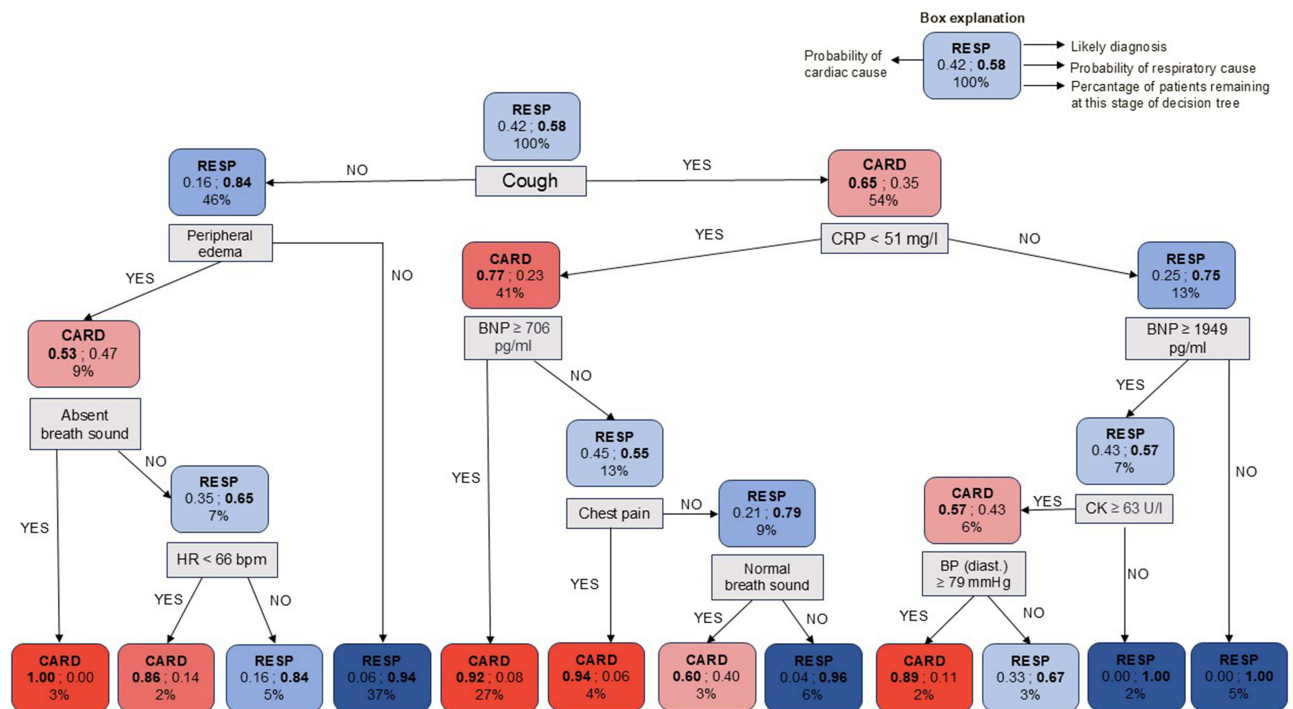


Figure 1 Decision tree for binary classification of dyspnea etiology (cardiac vs respiratory). The bold label (CARD or RESP) indicates the more likely diagnosis at that decision point. The bold probability number corresponds to that likely diagnosis, while the non-bold probability number represents the alternative diagnosis. The percentage shown at the bottom of each box indicates the proportion of patients remaining at that stage of the decision tree.

Discussion

Clinical Findings

This study examined the application and diagnostic accuracy of machine learning tools in identifying the underlying causes of dyspnea in hospitalized patients presenting to the ED. In our cohort, the leading diagnoses were heart failure, pneumonia, and COVID-19. These findings mirror international data on dyspnea etiologies⁸ but also reflect the pandemic context, which limits comparability with pre-pandemic cohorts.³⁹ This underscores the importance of interpreting diagnostic distributions within their temporal and regional context.⁴⁰

Machine Learning and Diagnostic Accuracy

We evaluated ML models to assess their potential in identifying causes of dyspnea. Decision trees achieved 84.66% accuracy (95% CI: 78.2–89.82%) in distinguishing between cardiac and respiratory causes but performed poorly when predicting specific final diagnoses (accuracy: 36.82%). This disparity highlights the models' limited granularity and their reliance on data completeness. This lower accuracy may reflect both the small size of some diagnostic groups and the considerable overlap in clinical features between conditions such as pneumonia, COVID-19, and COPD exacerbations, highlighting the inherent diagnostic complexity of dyspnea presentations. These overlapping pathologies diluted the discriminatory signal of single-class models and contributed to reduced predictive performance. While this complexity limits model accuracy, it reflects the real-world diagnostic challenge in emergency medicine.

Emergency care involves diverse diagnostic processes, and while certain lab tests are quickly and broadly available (eg, CRP, BNP), detailed clinical assessments are time-consuming and not always feasible. Thus, missing data may impair model training.

Despite these limitations, key features, such as CRP, BNP, and the symptom “cough”, showed high discriminatory value. CRP, typically elevated in infections like pneumonia, and BNP, associated with cardiac stress, have both been validated in prior studies. Morrison et al reported BNP's high sensitivity (86%) and specificity (98%) for differentiating cardiac from pulmonary causes.⁴¹ Joffe et al confirmed CRP's role in distinguishing pneumonia from heart failure.⁴² To

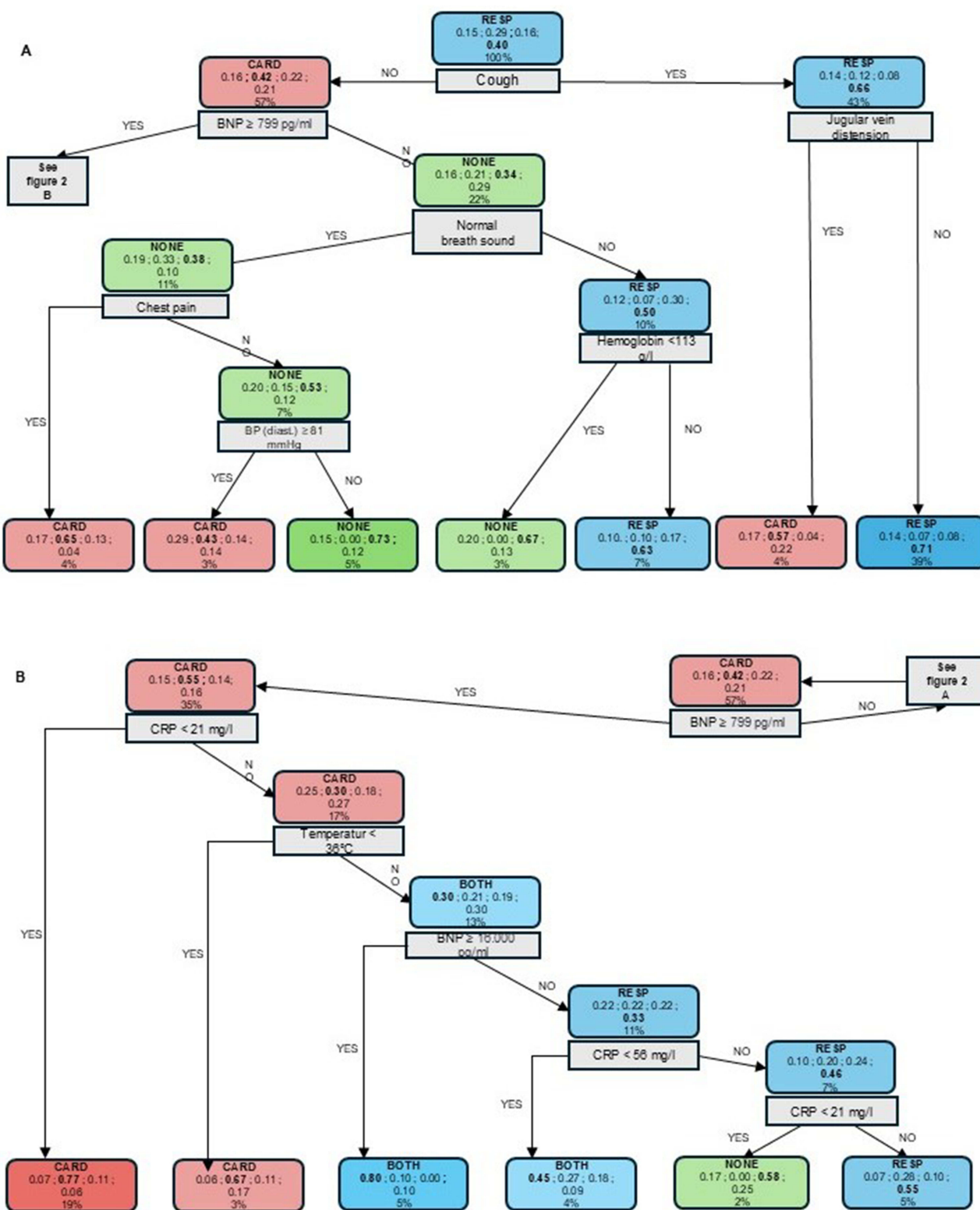


Figure 2 (A) Decision tree for multi-class classification of dyspnea etiology (cardiac, respiratory, combined cardiac–respiratory, or other causes), part A. The bold label indicates the most likely diagnosis at each decision point. The bold probability number corresponds to that diagnosis, while the non-bold probability numbers represent the alternative probabilities. The percentage at the bottom of each box shows the proportion of patients remaining at that stage of the decision tree. **(B)** Decision tree for multi-class classification of dyspnea etiology (cardiac, respiratory, combined cardiac–respiratory, or other causes), part B. The same conventions apply as in **(A)** bold labels and bold numerical values represent the most likely diagnosis; non-bold numbers indicate alternative probabilities; and percentages at the bottom denote the proportion of patients remaining at each stage.

ensure clinical plausibility, we applied feature importance analysis (Gini index) across all models, which confirmed that predictive variables aligned with established pathophysiology.

Integrating such markers into a simplified decision tool could support triage. Examples like CURB-65 (for pneumonia) and NT-proBNP-guided heart failure management illustrate the utility of parameter-driven approaches.^{43–46} While our limited feature set precludes development of a standalone score, combining these variables with demographic and clinical data could enhance diagnostic workflows in future applications.

Our cohort also exhibited frequent comorbidities, particularly arterial hypertension, diabetes, and coronary artery disease, conditions that increase the risk of heart failure. Prior studies^{21,30} affirm the relevance of comorbidities in risk stratification. Despite their clinical relevance, comorbidities demonstrated a low Gini index in our machine learning algorithms, indicating limited predictive utility in this specific context. One possible explanation is that while comorbidities provide background information, they might be non-specific and do not directly distinguish between the multifactorial causes of dyspnea. The complexity of dyspnea's etiology, where overlapping cardiac and pulmonary conditions often coexist, might dilute the weight of comorbidities as individual predictive factors. Additionally, the algorithms may prioritize acute clinical parameters, such as vital signs and specific diagnostic tests, over chronic background conditions, as suggested by Heymann et al, who noted the higher diagnostic weight of real-time clinical variables.³⁰

Compared with prior ML applications in ED or respiratory diagnostics, our models achieved similar accuracy in broad differentiation (cardiac vs respiratory causes) but diverged in two important ways: first, by focusing specifically on undifferentiated dyspnea rather than isolated conditions (eg, pneumonia or heart failure alone); and second, by demonstrating how missing or incomplete features in routine ED data directly constrain diagnostic granularity. This focus on dyspnea as a heterogeneous syndrome, rather than a single disease entity, highlights the novelty of our approach and its relevance for real-world ED workflows.

Strengths and Limitations

The low sensitivity of ML models for specific diagnoses can be attributed to limitations that directly affect model training and validation. First, retrospective data collection introduced missing or inconsistent features, particularly for symptoms and physical findings. This limited data quality reduces model generalizability and increases the risk of overfitting to the available features while ignoring real-world variability. Second, overlapping clinical presentations (eg, COVID-19 vs pneumonia) limited the models' ability to learn distinct diagnostic patterns. Third, less common diagnoses were under-represented, leading to class imbalance and skewing predictions toward more prevalent conditions. Additionally, the exclusion of patients with multiple diagnoses in certain models may introduce selection bias and reduce generalizability, as many patients in emergency care present with overlapping cardiac and respiratory conditions. Although multilabel classification was performed to partially address this, the Binary Relevance approach does not capture correlations between diagnoses, which could further limit model performance. These aspects highlight the need for future work using methods that better reflect the complexity of real-world diagnostic overlap.

The study's strengths include using real-world data from a typical hospital population, having very complete lab data, and combining clinical knowledge with algorithm-based analysis. However, the results should still be interpreted carefully because of the study's limitations. In particular, the Swiss healthcare setting, characterized by universal coverage and structured cantonal hospital networks, may limit the direct generalizability of our findings to healthcare systems with different resource availability.

Future Studies

To improve model robustness and diagnostic accuracy, future studies should incorporate larger, multicenter datasets. Broader patient representation can reduce selection bias and improve generalizability. Furthermore, the application of more sophisticated techniques, such as deep learning, could enable better handling of complex, non-linear patterns.

Prospective studies are also needed. Unlike retrospective designs, prospective data collection ensures consistent documentation, reduces bias, and allows for better evaluation of model performance in real-time clinical settings.

While multimodal data (eg, imaging, continuous monitoring, or free-text notes) could further improve accuracy, our study specifically targeted admission-available data to support clinicians in their earliest decision-making; however future research

may explore hybrid models that first provide an admission-based prediction and later update performance as additional modalities (eg, imaging or EHR text) become available, thereby combining timeliness with improved diagnostic accuracy.

Integrating AI Into Clinical Practice: Benefits and Challenges

AI-based decision support could offer clear potential to support clinical workflows, particularly under time constraints. In the context of dyspnea, early differentiation between cardiac and respiratory etiologies could streamline investigations, accelerate treatment, and reduce morbidity. However, ethical challenges must be addressed before widespread implementation. Over-reliance on AI could lead to missed diagnoses, especially if clinicians defer to algorithms without critical assessment. Less experienced staff may be particularly susceptible to this risk. Therefore, AI systems must be designed as decision aids, not replacements for clinical reasoning. Clear guidelines should accompany implementation, outlining the intended use and limitations of AI tools. Validation in diverse clinical environments is essential to ensure reliable performance across different populations and healthcare systems. Future research should also investigate how AI influences clinician behavior, decision-making, and patient outcomes. Understanding these dynamics will be key to ensuring safe, effective, and ethically responsible adoption.

Conclusions

AI is poised to play an increasingly important role in emergency and acute care. In patients with dyspnea, ML tools can aid initial triage by rapidly identifying etiologies, guiding early diagnostics, and supporting resource allocation. However, current models lack sufficient accuracy to guide specific diagnoses reliably and should not replace clinical judgment.

Key parameters such as CRP, BNP, and cough emerged as valuable features for distinguishing between cardiac and respiratory causes, underscoring their potential role in simplified diagnostic tools. Although comorbidities provide important clinical context, they demonstrated limited predictive value in this setting.

To enhance AI integration into emergency care, future efforts should focus on prospective, multicenter studies, the use of advanced modeling techniques, and thorough evaluation of real-world impact. Ensuring that AI tools are used ethically, transparently, and in conjunction with clinical expertise will be essential for delivering safe and effective care. Before clinical implementation, prospective validation and real-time testing will be essential to confirm robustness and applicability in diverse emergency care settings.

Data Sharing Statement

All data generated during this study were analyzed and the results were included in this article. The data presented in this study are available on reasonable request from the corresponding author. The data are not publicly available due to restrictions on data privacy.

Ethics Approval

The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board (or Ethics Committee) of Northwest and Central Switzerland (ID 2015-339 amendments were approved 20 November 2023) for studies involving humans. This manuscript was prepared in accordance with the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines.

Consent to Participate

Patients whose general written consent was unavailable but for whom a consent exception was permitted by the Ethics Committee (Art. 34 HFG, Swiss Federal Act on Research Involving Human Beings) were included. This exception applies when patients are unable to provide consent and retrospective analysis poses minimal risk, provided no objection was recorded. Patients who denied the hospital's general consent request were excluded.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically

reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Disclosure

J.D.L. has received unrestricted grants from AstraZeneca AG Switzerland, GSK AG Switzerland, OM Pharma SA Switzerland, and Sanofi AG Switzerland. All other authors declare no potential conflicts of interest in this work.

References

- Ashman JJ, Cairns C, DeFrances CJ, Schwartzman A. Respiratory illness emergency department visits in the national hospital care survey and the national hospital ambulatory medical care survey. *Natl Health Stat Rep.* 2021;(151):1–18.
- Stevens JP, Dechen T, Schwartzstein RM, O'Donnell CR, Baker K, Banzett RB. Association of dyspnoea, mortality and resource use in hospitalised patients. *Eur Respir J.* 2021;58(3):1902107. doi:10.1183/13993003.02107-2019
- Kelly AM, Keijzers G, Klim S, et al. An observational study of dyspnea in emergency departments: the Asia, Australia, and New Zealand Dyspnea in Emergency Departments Study (AANZDEM). *Acad Emerg Med.* 2017;24(3):328–336. doi:10.1111/acem.13118
- Parshall MB, Schwartzstein RM, Adams L, et al. An official American thoracic society statement: update on the mechanisms, assessment, and management of dyspnea. *Am J Respir Crit Care Med.* 2012;185(4):435–452. doi:10.1164/rccm.201111-2042ST
- Laribi S, Keijzers G, van Meer O, et al. Epidemiology of patients presenting with dyspnea to emergency departments in Europe and the Asia-Pacific region. *Eur J Emerg Med.* 2019;26(5):345–349. doi:10.1097/MEJ.0000000000000571
- Wahls SA. Causes and evaluation of chronic dyspnea. *Am Fam Physician.* 2012;86(2):173–182.
- Budhwar N, Syed Z. Chronic dyspnea: diagnosis and evaluation. *Am Fam Physician.* 2020;101(9):542–548.
- Berliner D, Schneider N, Welte T, Bauersachs J. The differential diagnosis of dyspnea. *Dtsch Arzteblatt Int.* 2016;113(49):834–845.
- Hale ZE, Singhal A, Hsia RY. Causes of shortness of breath in the acute patient: a national study. *Acad Emerg Med Off J Soc Acad Emerg Med.* 2018;25(11):1227–1234. doi:10.1111/acem.13448
- Huang YC, Ferry OR, McKenzie SC, et al. Diagnosis of the cause of chronic dyspnoea in primary and tertiary care: characterizing diagnostic confidence. *J Thorac Dis.* 2018;10(6):3745–3756. doi:10.21037/jtd.2018.05.183
- Spörl P, Beckers SK, Rossaint R, Felzen M, Schröder H. Shedding light into the black box of out-of-hospital respiratory distress—A retrospective cohort analysis of discharge diagnoses, prehospital diagnostic accuracy, and predictors of mortality. Veldhuizen RA, editor. *PLoS One.* 2022;17(8):e0271982. doi:10.1371/journal.pone.0271982
- Zwaan L, Thijs A, Wagner C, Timmermans DRM. Does inappropriate selectivity in information use relate to diagnostic errors and patient harm? The diagnosis of patients with dyspnea. *Soc Sci Med.* 2013;91:32–38. doi:10.1016/j.socscimed.2013.05.001
- Patel K, Smith DJ, Huntley CC, et al. Exploring the causes of COPD misdiagnosis in primary care: a mixed methods study. *PLoS One.* 2024;19(3):e0298432. doi:10.1371/journal.pone.0298432
- Joseph MM, Mahajan P, Snow SK, Ku BC, Saidinejad M. American academy of pediatrics committee on pediatric emergency medicine, American college of emergency physicians pediatric emergency medicine committee, emergency nurses association pediatric committee. *Optimizing Pediatric Patient Safety Emergency Care Setting Pediatr.* 2022;150(5):e2022059673.
- Ly DP, Shekelle PG, Song Z. Evidence for anchoring bias during physician decision-making. *JAMA Intern Med.* 2023;183(8):818. doi:10.1001/jamainternmed.2023.2366
- Gartlehner G, Wagner G, Affengruber L, et al. Point-of-care ultrasonography in patients with acute dyspnea: an evidence report for a clinical practice guideline by the American college of physicians. *Ann Intern Med.* 2021;174(7):967–976.
- Celli BR, Fabbri LM, Aaron SD, et al. Differential diagnosis of suspected chronic obstructive pulmonary disease exacerbations in the acute care Setting: best practice. *Am J Respir Crit Care Med.* 2023;207(9):1134–1144. doi:10.1164/rccm.202209-1795CI
- Hui D, Bohlke K, Bao T, et al. Management of dyspnea in advanced cancer: ASCO guideline. *J Clin Oncol Off J Am Soc Clin Oncol.* 2021;39(12):1389–1411. doi:10.1200/JCO.20.03465
- Malas Ö, Çağlayan B, Fidan A, Öcal Z, Özdoğan S, Torun E. Cardiac or pulmonary dyspnea in patients admitted to the emergency department. *Respir Med.* 2003;97(12):1277–1281.
- Wang CS. Does this dyspneic patient in the emergency department have congestive heart failure? *JAMA.* 2005;294(15):1944. doi:10.1001/jama.294.15.1944
- Tinë M, Bazzan E, Semenzato U, et al. Heart failure is highly prevalent and difficult to diagnose in severe exacerbations of COPD presenting to the emergency department. *J Clin Med.* 2020;9(8):2644. doi:10.3390/jcm9082644
- Kocher KE, Meurer WJ, Desmond JS, Nallamothu BK. Effect of testing and treatment on emergency department length of stay using a national database. *Acad Emerg Med.* 2012;19(5):525–534.
- Chadaga K, Prabhu S, Sampathila N, et al. Explainable artificial intelligence approaches for COVID-19 prognosis prediction using clinical markers. *Sci Rep.* 2024;14(1):1783. doi:10.1038/s41598-024-52428-2
- Lüscher TF, Wenzl FA, D'Ascenzo F, Friedman PA, Antoniadis C. Artificial intelligence in cardiovascular medicine: clinical applications. *Eur Heart J.* 2024;45(40):4291–4304. doi:10.1093/eurheartj/ehae465
- Liu W, Wang H, Shen X, Tsang IW. The emerging trends of multi-label learning. *IEEE Trans Pattern Anal Mach Intell.* 2022;44(11):7955–7974.
- Zhou L, Zheng X, Yang D, Wang Y, Bai X, Ye X. Application of multi-label classification models for the diagnosis of diabetic complications. *BMC Med Inform Decis Mak.* 2021;21(1):182. doi:10.1186/s12911-021-01525-7

27. Podgorelec V, Kokol P, Stiglic B, Rozman I. Decision trees: an overview and their use in medicine. *J Med Syst.* 2002;26(5):445–463. doi:10.1023/A:1016409317640
28. Hassan Mohamed M, Hussein Ali B, Ibrahim Taloba A, et al. Toward an accurate liver disease prediction based on Two-Level ensemble stacking model. *IEEE Access.* 2024;12:180210–180237. doi:10.1109/ACCESS.2024.3459429
29. Sömbül H, Yüzer AH. Design of a fuzzy input expert system visual information interface for classification of apnea and hypopnea. *Multimed Tools Appl.* 2023;83(7):21133–21152. doi:10.1007/s11042-023-16152-9
30. Heyman ET, Ashfaq A, Ekelund U, et al. A novel interpretable deep learning model for diagnosis in emergency department dyspnoea patients based on complete data from an entire health care system. *PLoS One.* 2024;19(12):e0311081. doi:10.1371/journal.pone.0311081
31. Federal Office of Public Health (FOPH). *Hospital statistics*. Bern: FOPH; 2023 [Cited September 11, 2025]. Available from: https://spitalstatistik-id.bagapps.ch/data/download/kzp22_publication.pdf?v=1740817749. Accessed October 06, 2025.
32. ID Suisse AG. *NGWAB – medizin software*. Available from: <https://www.id-suisse-ag.ch/>. Accessed October 06, 2025.
33. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: building an international community of software platform partners. *J Biomed Inf.* 2019;95:103208. doi:10.1016/j.jbi.2019.103208
34. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inf.* 2009;42(2):377–381. doi:10.1016/j.jbi.2008.08.010
35. Dolgin M, New York Heart Association, editors. *Nomenclature and Criteria for Diagnosis of Diseases of the Heart and Great Vessels*. Boston, Mass: Little, Brown and Co; 1994:334.
36. Beretta L, Santaniello A. Nearest neighbor imputation algorithms: a critical evaluation. *BMC Med Inform Decis Mak.* 2016;16(S3):74. doi:10.1186/s12911-016-0318-z
37. Mucherino A, Papajorgji PJ, Pardalos PM. k-Nearest Neighbor Classification. In: *Data Mining in Agriculture*. New York:Springer New York; 2009.
38. R Core Team. 2021 R A language and environment for statistical computing. R foundation for statistical computing, vienna. - References - scientific research publishing. Available from: <https://www.scirp.org/reference/referencespapers?referenceid=3131254>. Accessed October 06, 2025.
39. Lindskou TA, Pilgaard L, Søvsø MB, et al. Symptom, diagnosis and mortality among respiratory emergency medical service patients. *PLoS One.* Lazzeri C, editor. 2019;14(2):e0213145. doi:10.1371/journal.pone.0213145
40. Collins SP, Pang PS, Lindsell CJ, et al. International variations in the clinical, diagnostic, and treatment characteristics of emergency department patients with acute heart failure syndromes. *Eur J Heart Fail.* 2010;12(11):1253–1260. doi:10.1093/eurjhf/hfq133
41. Morrison LK, Harrison A, Krishnaswamy P, Kazanegra R, Clopton P, Maisel A. Utility of a rapid B-natriuretic peptide assay in differentiating congestive heart failure from lung disease in patients presenting with dyspnea. *J Am Coll Cardiol.* 2002;39(2):202–209. doi:10.1016/S0735-1097(01)01744-2
42. Joffe E, Justo D, Mashav N, et al. C-reactive protein to distinguish pneumonia from acute decompensated heart failure. *Clin Biochem.* 2009;42(16–17):1628–1634. doi:10.1016/j.clinbiochem.2009.08.007
43. Siebert U, Milev S, Zou D, et al. Economic evaluation of an N-terminal Pro B-type natriuretic peptide-supported diagnostic strategy among dyspneic patients suspected of acute heart failure in the emergency department. *Am J Cardiol.* 2021;147:61–69. doi:10.1016/j.amjcard.2021.01.036
44. Heidenreich PA, Bozkurt B, Aguilar D, 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).
45. Jones BE, Jones J, Bewick T, et al. CURB-65 pneumonia severity assessment adapted for electronic decision support. *Chest.* 2011;140(1):156–163. doi:10.1378/chest.10-1296
46. Metlay JP, Waterer GW, Long AC, et al. Diagnosis and treatment of adults with community-acquired pneumonia. an official clinical practice guideline of the American thoracic society and infectious diseases society of America. *Am J Respir Crit Care Med.* 2019;200(7):e45–67. doi:10.1164/rccm.201908-1581ST

Journal of Multidisciplinary Healthcare

Publish your work in this journal

The Journal of Multidisciplinary Healthcare is an international, peer-reviewed open-access journal that aims to represent and publish research in healthcare areas delivered by practitioners of different disciplines. This includes studies and reviews conducted by multidisciplinary teams as well as research which evaluates the results or conduct of such teams or healthcare processes in general. The journal covers a very wide range of areas and welcomes submissions from practitioners at all levels, from all over the world. The manuscript management system is completely online and includes a very quick and fair peer-review system. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/journal-of-multidisciplinary-healthcare-journal>

Dovepress
Taylor & Francis Group