Multimorbidity in Heart Failure Patients: Application of Machine Learning Algorithms to Predict Imminent Health Outcomes

Jorge Cerejo[1,2][a], Rui Lopes Baeta[2][b], Simão Gonçalves[1][c], Bernardo Neves[1,3][d], Pedro Morais Sarmento[3][e], José Maria Moreira[1][f], Nuno André da Silva[1][g], Francisca Leite[1][h], Bruno Martins[2][i] and Mário J. Silva[2][j]

[1]*Hospital da Luz Learning Health, Luz Saúde, Lisboa, Portugal*

[2]*INESC-ID, Instituto Superior Técnico, Universidade de Lisboa, Lisboa, Portugal*

[3]*Hospital da Luz Lisboa, Luz Saúde, Lisboa, Portugal*

fi

{*bernardo.neves, psarmento*}*@hospitaldaluz.pt,* {*jose.maria.moreira, nuno.asilva, francisca.leite*}*@luzsaude.pt,*

Keywords: Multimorbidity, Heart Failure, Laboratory Tests, Health Outcomes Prediction.

Abstract: As populations age and life expectancy increases, multimorbidity, which is the simultaneous presence of two or more chronic conditions, has become increasingly common, especially among older adults. Heart failure, a widespread and heterogeneous syndrome, has sparked research into multimorbidity to deepen our understanding of its pathophysiology and improve clinical management approaches. This paper offers a detailed characterization of a heart failure patient cohort, utilizing clinical data from a Portuguese tertiary hospital. Based on this characterization, we developed a clinical tool for identification of high-risk patients and prediction of imminent hospital admissions based on laboratory tests. Our models for predicting imminent hospitalization showed reasonable effectiveness (AUROC of 0.79 with lab test prescriptions and 0.72 with lab test results). These findings emphasize the significant predictive value of laboratory tests in the context of HF. Additionally, we investigated the explainability of our models using SHAP values, in collaboration with clinical experts, providing insights into factors influencing the models' predictions. These results highlight the importance of secondary clinical data analysis assisting healthcare professionals in identifying patients at high risk of adverse events, and improving patient care and outcomes.

1 INTRODUCTION

As life expectancy increases and populations age, multimorbidity, defined as the presence of two or more chronic conditions, has become increasingly prevalent in healthcare systems worldwide (WHO, 2016). Multimorbidity poses significant challenges for patients, clinicians, and healthcare systems, underscoring the urgent need for innovative tools to better understand and improve clinical outcomes (Majnarić et al., 2021). In Portugal, the prevalence of multimorbidity is especially high, affecting 78.3% of the elderly population (aged 65 and older) and 38.3% of those between the ages of 24 and 75 (Rodrigues et al., 2018; Romana et al., 2019).

Heart Failure (HF), a clinical syndrome characterized by structural and/or functional cardiac abnormalities, is particularly associated with multimorbidity (Bozkurt et al., 2021). Among HF patients, 86% suffer from at least two chronic conditions, and 42% live with five or more (Chamberlain et al., 2015). The coexistence of conditions complicates treatment decisions and increases the risk of adverse outcomes, emphasizing the need for improved risk stratification methods (Navickas et al., 2016).

Electronic Health Records (EHRs) offer a valuable opportunity for advanced multimorbidity char-

[a] https://orcid.org/0009-0004-5221-4964
[b] https://orcid.org/0009-0006-8350-8302
[c] https://orcid.org/0009-0004-6565-8599
[d] https://orcid.org/0000-0002-1559-7482
[e] https://orcid.org/0000-0002-5970-2707
[f] https://orcid.org/0000-0003-2420-7930
[g] https://orcid.org/0000-0003-4216-2107
[h] https://orcid.org/0000-0003-2550-2616
[i] https://orcid.org/0000-0002-3856-2936
[j] https://orcid.org/0000-0002-5452-6185

Cerejo, J., Baeta, R. L., Gonçalves, S., Neves, B., Sarmento, P. M., Moreira, J. M., da Silva, N. A., Leite, F., Martins, B. and Silva, M. J.
Multimorbidity in Heart Failure Patients: Application of Machine Learning Algorithms to Predict Imminent Health Outcomes.
DOI: 10.5220/0013381800003911

In *Proceedings of the 18th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2025) - Volume 2: HEALTHINF*, pages 330-339
ISBN: 978-989-758-731-3; ISSN: 2184-4305

acterization and enhanced patient outcome predictions (Williams et al., 2022). Prior studies suggest that the secondary use of EHR data can improve care quality, reduce medical errors, and generate cost savings (CMS, 2021). However, translating this data into actionable insights remains a critical challenge.

In this work, we focus on the development of a predictive tool designed to forecast imminent hospitalizations among HF patients with multimorbidity using pseudonymized EHR data from a Portuguese tertiary hospital. As part of the broader IntelligentCare project, this study aims to identify high-risk patients and support clinicians in making timely and informed decisions in the emergency department (ED). We introduce the ICIHO (IntelligentCare Imminent Health Outcomes) predictive tool, which leverages classification algorithms to predict imminent health outcomes (IHO) using routine laboratory test results collected during ED visits. By utilizing observational data, the ICIHO tool enhances the early identification of HF patients at heightened risk for hospital admission, thereby facilitating personalized treatment planning and potentially improving patient outcomes.

The paper is organized as follows: Section 2 provides an overview of key concepts and related work relevant to the study. Section 3 describes the methodology used for data processing and prediction modeling. The results are presented in Section 4, followed by a discussion in Section 5. Finally, Section 6 offers concluding remarks and highlights the clinical implications of the study.

2 RELATED WORK

EHRs serve as a rich source of patient data, enabling applications in learning healthcare systems and precision medicine (Aronson and Rehm, 2015). However, analyzing EHR data presents significant challenges, including missing data, inaccuracies, and data heterogeneity (Hripcsak et al., 2011). To address these issues, data standardization initiatives like the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) have emerged, providing a framework for uniform data extraction and analysis (Sciences and Informatics(OHDSI), 2023). The OMOP CDM facilitates global data sharing for comparative longitudinal studies, making it an essential tool for analyzing complex patient data (Dixon et al., 2020; Liyanage et al., 2018).

Laboratory tests are essential in healthcare, serving roles in diagnosis, monitoring, screening, and research. They play a crucial role in reducing diagnostic errors and facilitating informed clinical decisions (Wians, 2009; Plebani and Lippi, 2016). Abnormal test values are often early indicators of adverse events, such as increased morbidity and mortality (Asadollahi et al., 2007). Despite their critical importance, analyzing laboratory data is challenging due to its inherent heterogeneity. To ensure consistency and improve interoperability, the healthcare industry employs the LOINC (Logical Observation Identifiers, Names, and Codes) system, which standardizes the identification and representation of laboratory measurements (Loinc® Indianapolis, IN: Regenstrief Institute, Inc,).

In recent years, the use of supervised machine learning (ML) models to predict adverse clinical outcomes from EHR data has grown substantially (Lee et al., 2020; Nwanosike et al., 2022). Logistic Regression (LR) models remain widely used due to their simplicity, interpretability, and effectiveness in predicting key outcomes such as mortality and hospital admissions, thereby influencing clinical decisions and improving healthcare delivery (Alanazi, 2022). Meanwhile, more advanced techniques, such as deep learning models, have gained prominence for their ability to handle large datasets and extract complex patterns (Shamout et al., 2020).

An important application of ML in healthcare is the use of laboratory data to predict IHO. For example, Loekito et al. developed a multivariate LR model utilizing 30 laboratory variables to predict IHOs (Loekito et al., 2013). Their model demonstrated its effectiveness by accurately identifying key outcomes such as Medical Emergency Team (MET) calls (AUROC = 0.69), ICU admissions (AUROC = 0.82), and in-hospital mortality (AUROC = 0.90). Similarly, Mueller et al. employed a comparable approach to predict in-hospital mortality, achieving high performance (AUROC = 0.88). Another model integrating both demographic and laboratory data also demonstrated strong predictive ability for hospitalizations (AUROC = 0.80) (Mueller et al., 2021).

3 MATERIAL AND METHODS

Using a combination of data mining and ML techniques, we developed a pipeline for characterization of HF patterns with multimorbidity from EHRs. The pipeline enables the stratification of HF patients based on their risk of adverse events, such as imminent ($\leq$ 24h) hospitalizations. With the developed pipeline, we processed the clinical records of Hospital da Luz Lisboa (HLL), an institution that provides comprehensive medical services across all medical specialties.

Figure 1 provides a visual overview of the developed workflow, which processes the clinical data in four main stages:

1. Observational Data Extraction
2. HF Cohort Selection
3. Patient Data Loading
4. Imminent Health Outcomes Prediction

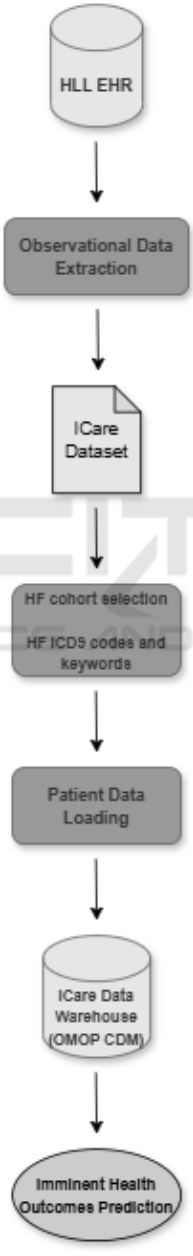


Figure 1: Overview of the developed workflow.

3.1 Observational Data Extraction

We utilized the IntelligentCare (ICare) dataset from Hospital da Luz Lisboa (HLL), containing anonymized medical histories of 834,529 patients, spanning January 2007 to August 2021. This dataset was approved by the hospital's Institutional Review Board (IRB) for research on multimorbidity. Data included patient visits, diagnoses, and laboratory results, essential for predicting imminent health outcomes (IHO).

3.2 HF and Comorbidities Phenotyping

We utilized a locally validated phenotyping algorithm that combined ICD-9 codes and HF-related keywords in clinical text, as detailed in a prior publication by our research group (Martins et al., 2024). Comorbidities were selected based on their prevalence and clinical significance, guided by clinical expertise. The phenotyping rules applied included checks for both ICD-9 codes and relevant clinical text to ensure accurate cohort identification.

3.3 Patient Data Loading

To ensure interoperability, we harmonized the ICare dataset and uploaded it into an OMOP CDM (version 5.3) data warehouse using tools from the Observational Health Data Sciences and Informatics (OHDSI) community. The data warehouse serves as the backbone of our analytical processes, enabling standardized analysis through tools like White Rabbit for ETL design (OHDSI, 2021), Athena for concept mapping (Athena, 2022), and Achilles for generating standard metrics (OHDSI, 2014).

We populated six OMOP CDM tables: Person (demographics), Visit Occurrence (patient visits), Condition Occurrence (diagnoses and comorbidities), Measurement (laboratory values), Death (mortality), and Observation Period (timeline of observations). We extracted the medical histories of HF cohort patients, preserving the chronological order of diagnoses, lab tests, and hospital interactions.

Conditions coded using ICD-9 were translated to standardized concept IDs via the Condition Relationship table to ensure accurate mapping within the Condition Occurrence table. Laboratory data required extensive preprocessing due to non-uniform coding. For unmatched lab codes, we applied heuristics such as removing the last digit for mapping LOINC standards. Of the nearly 40 million lab records, 11.4% could not be mapped and were excluded.

Lab results, which varied from quantitative to

nominal or free-text entries, were harmonized using regular expressions and keywords. Numerical data were standardized, and categorical results (e.g., "Positive", "POS") were unified under consistent labels.

Finally, we mapped the processed data to OMOP CDM fields and loaded them into the data warehouse using Pentaho Data Integration (Hitachi Ventara,), completing the ETL process and enabling robust, standardized analysis for subsequent predictive modeling.

3.4 Imminent Health Outcome Prediction

We developed a methodology to predict imminent hospitalizations of HF patients admitted to the ED using laboratory data from the ICare dataset. The primary aim was to evaluate the predictive power of lab tests in forecasting IHO for HF patients. Only patients with at least one recorded laboratory test were included in the analysis.

To predict imminent hospitalizations, we linked laboratory measurements to subsequent clinical episodes in each patient's history. Clinical episodes, defined as healthcare visits (e.g., hospitalizations, ED visits, consultations), were recorded in the Visit Occurrence table. The algorithm identified the next clinical episode following each lab measurement and calculated the time difference. Episodes followed by hospitalization within 24 hours were labeled as positive (1), while those leading to discharge were labeled as negative (0).

We first trained a multivariate LR model using binary indicators of lab test prescriptions (1 if prescribed, 0 otherwise). This approach eliminated the need for actual lab results, mitigating issues with missing data. By applying the Chi-square test, we identified statistically significant lab tests for predicting imminent hospitalizations. The transformed dataset, in categorical format, enabled the LR model to capture patterns in physician decision-making.

To evaluate the predictive value of actual lab test results, we compared LR and neural networks (NN) models. LR was chosen for interpretability, while NNs were evaluated for potential performance improvements. Statistically significant lab tests were used as features, and NT-proBNP, a critical biomarker for HF management (Bozkurt et al., 2021), was imputed where missing. Class imbalance was addressed through weight correction. Elastic Net regularization prevented overfitting in the LR model, with hyperparameters tuned via 5-fold cross-validation (training set size = 80%), optimizing the F1-score. The NN architecture consisted of two hidden layers with four neurons each (Dervishi, 2020), trained using binary cross-entropy loss and stochastic gradient descent with backpropagation.

To address the trade-off between the number of features and the size of the training dataset, we designed variations of LR and NN models with different numbers of laboratory test features. Missing values in the lab test results, which are not missing at random, made imputation not recommended. Instead, rows with missing values were excluded, meaning models with more features had smaller training datasets. This approach aimed to evaluate how feature count and training data size impacted model performance.

Model performance was assessed using balanced accuracy, precision, recall, F1-score, AUROC, and AUPRC (Han et al., 2012). Confidence intervals were calculated using bootstrapping. To enhance interpretability, we analyzed the LR model's coefficients using odds ratios (ORs) and SHAP (SHapley Additive exPlanations) values (Kasza and Wolfe, 2014; Lundberg and Lee, 2017). We used SHAP summary plots to visualize the overall importance of features and SHAP force plots to analyze individual predictions, offering insights into how specific lab results contributed to imminent hospitalization risk (Lundberg et al., 2018).

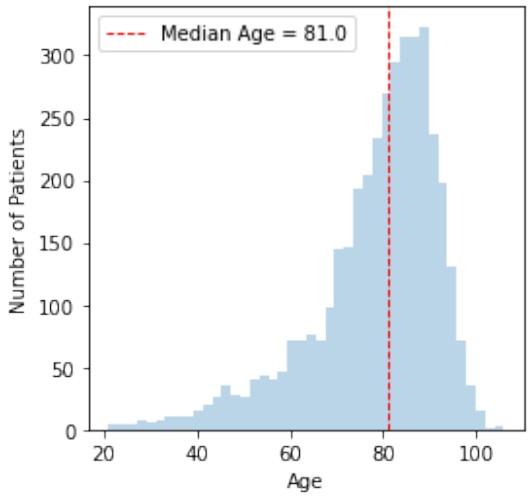
To support real-time interpretation, we developed the interactive ICIHO predictive tool[1]. This tool visualizes predicted hospitalization risk and highlights the influence of key lab tests, enhancing clinicians' ability to make informed decisions based on SHAP-derived insights.

4 RESULTS

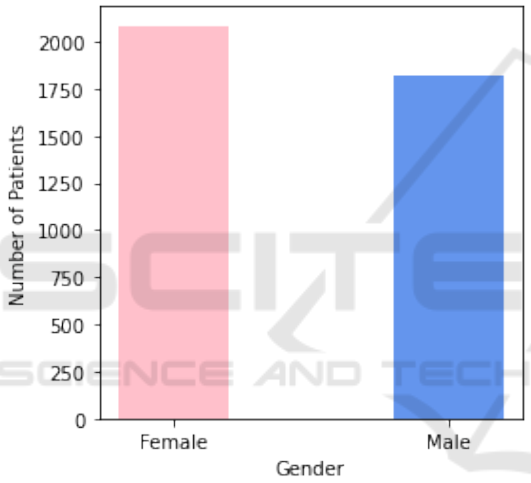
The study population included 3907 patients with HF (53.4% women) with median age of 81 years (interquartile range 72-88 years old) as depicted in Figures 2a and 2b. Comorbidities such as cardiovascular conditions, CKD and Diabetes were highly prevalent (Table 1).

We analyzed 3,407 patients for imminent outcome prediction after excluding those without available laboratory test data. These patients contributed to 46,922 distinct episodes, including 27,744 outpatient admissions, 12,686 ED admissions, and 6,488 hospitalizations. From the ED admissions, 4,693 (37.0% of the total amount of ED admissions) led to hospital admissions within 24 hours, and were labeled as positives. A total of 437,683 laboratory tests were conducted during these ED visits, comprising 252 unique tests.

[1] Available at ICIHO predictive tool



(a)



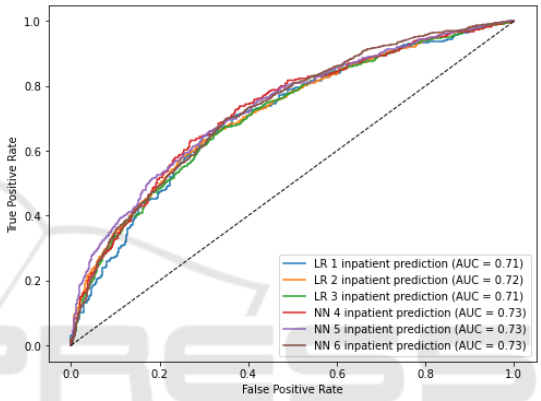
(b)

Figure 2: Gender and Age distributions of the HF cohort.

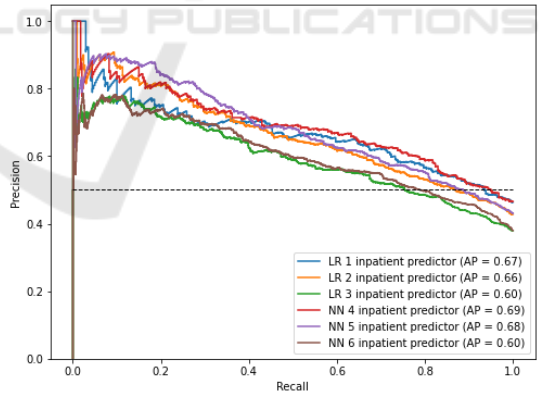
Our multivariate LR model for the prediction of imminent hospitalization using only the lab test prescriptions demonstrated a reasonable performance, with a recall of 0.654, a precision of 0.616, and an F1-score of 0.634, with an AUROC of 0.785 and an AUPRC of 0.707. Models trained on lab test results demonstrated similar performance, summarized in Table 2. A comparative analysis on model performance based, on the ROC curve and precision-recall curve, is provided in Figure 3. Overall, the NN models slightly outperformed the LR models. Models that included more features, namely LR 1 and NN 4, achieved higher performance, while models with fewer features, namely LR 3 and NN 6, exhibited a decrease in performance. This may indicate a lower significance power of these features in predicting imminent hospitalizations among patients admitted to the ED. We selected the LR 2 model, which is highlighted in Table 2, as the optimal approach considering the trade-off between model complexity and performance. This model demonstrated reasonable performance, with AUROC and AUPRC values of 0.718 and 0.663, respectively.

Table 1: Prevalence of the 10 most frequent chronic conditions identified in the population of HF patients.

	Condition	Prevalence
1	Essential hypertension	56%
2	Atrial fibrillation	33%
3	Dyslipidemia	27%
4	Chronic kidney disease	24%
5	Ischemic congestive cardiomyopathy	23%
6	Obesity	18%
7	Heart valve disorder	16%
8	Diabetes mellitus	14%
9	Allergic disposition	13%
10	Bacterial pneumonia	12%



(a)



(b)

Figure 3: Comparison of (a) ROC curves and (b) precision-recall curves for the different models that were trained. All models perform similarly, with AUROC ranging between 0.71 and 0.73 and AUPRC ranging between 0.60 and 0.69.

Table 2: Summary of the performance metrics computed to evaluate the models trained for predicting imminent hospitalizations of HF patients admitted to the ED. The model variations (LR 1, LR 2, LR 3, NN 1, NN 2, NN 3) evaluate the trade-off between the number of features and the size of the training dataset.

Model	**LR 1**	**LR 2**	**LR 3**	**NN 4**	**NN 5**	**NN 6**
Proportion	50%	**75%**	90%	50%	75%	90%
Nr. features	19	**16**	9	19	16	9
Nr. samples	3685	**8599**	12014	3685	8599	12014
Bal_accuracy	0.660	**0.664**	0.659	0.672	0.673	0.664
[95% CI]	[0.627-0.692]	**[0.640-0.688]**	[0.640-0.679]	[0.638-0.704]	[0.652-0.697]	[0.646-0.685]
Precision	0.643	**0.610**	0.555	0.643	0.619	0.558
[95% CI]	[0.593-0.693]	**[0.575-0.646]**	[0.526-0.586]	[0.594-0.692]	[0.578-0.644]	[0.529-0.588]
Recall	0.609	**0.624**	0.624	0.656	0.653	0.636
[95% CI]	[0.556-0.658]	**[0.589-0.659]**	[0.591-0.654]	[0.610-0.704]	[0.630-0.698]	[0.606-0.668]
F1-score	0.625	**0.617**	0.587	0.649	0.635	0.595
[95% CI]	[0.584-0.667]	**[0.575-0.646]**	[0.561-0.613]	[0.610-0.690]	[0.608-0.664]	[0.569-0.620]
AUROC	0.709	**0.718**	0.711	0.726	0.732	0.726
[95% CI]	[0.673-0.743]	**[0.693-0.743]**	[0.691-0.733]	[0.690-0.761]	[0.708-0.756]	[0.706-0.747]
AUPRC	0.668	**0.663**	0.596	0.691	0.682	0.605
[95% CI]	[0.614-0.720]	**[0.630-0.701]**	[0.564-0.631]	[0.634-0.744]	[0.649-0.719]	[0.571-0.640]

The odds ratio values corresponding to the coefficients of model LR2 are displayed in Table 3. These values reveal that C-reactive protein (CRP) and NT-proBNP are the laboratory tests that most significantly influence the prediction of imminent hospitalizations of patients admitted to the ED. The laboratory tests of Erythrocytes and Lymphocytes exhibited lower odds values, indicating a negative contribution to imminent hospitalizations. Despite age being considered a risk factor for multimorbidity, its influence in predicting imminent hospitalizations of HF patients is reduced.

Table 3 lists the P-values associated with each coefficient of the multivariate model. Certain variables, such as Hematocrit, Creatinine, Sodium, and Monocytes, seem to be statistically insignificant in predicting imminent hospitalization in a multivariate approach.

The SHAP summary plot shown in Figure 4a further reinforces these findings. High values of CRP, Urea, Leukocytes, and NT-proBNP are associated with higher SHAP values, indicating their strong positive influence on the model's predictions. Conversely, low values of Erythrocytes are associated with higher contributions to the prediction of imminent hospitalizations.

Figure 4b showcases a SHAP force plot example, illustrating the prediction of imminent hospitalization of a HF patient from the test dataset who was correctly classified. The model identified that values of Erythrocytes (789-8=2.16 $counts 10^9/L$), Leukocytes (6690-2=16.22 $counts 10^9/L$), and CRP (1988-5=8.65 mg/dL) had a significant impact on predicting imminent hospitalizations, as indicated by the wider red bars in the plot.

5 DISCUSSION

We developed a framework for processing clinical data to gain insights on the multimorbidity population with HF, uncovering patterns and risks associated to this condition. The ability to utilize healthcare data for better characterization of complex patients and the development of clinical strategies represents a step forward in the management of HF. Compared to related works, our model achieves similar performance while uniquely incorporating feature contribution analysis using odds ratios and SHAP. Additionally, we developed a user-friendly web interface to visualize predictions and feature impacts, supporting clinical decision-making.

Firstly, our multivariate analysis, which focused on lab test prescriptions for HF patients admitted to the ED, enables early identification of patients at an increased risk of imminent hospitalization. We believe that by exploring the rationale behind each lab test prescription, we can partially reveal the intricate clinical judgments and organizational factors influencing these decisions. This approach opens up new research avenues for clinical and operational improvements in high-demand settings, such as the ED.

In addition, we have shown the practical utility of commonly available laboratory test results in conducting risk stratification to predict short-term hospital admissions for HF patients. These models were proficient in making reasonably accurate predictions of hospital admission. We are optimistic that integrating additional information like demographics, vital signs, and diagnoses can further enhance the models' discrimination capabilities. Using these data, clini-

Table 3: Imminent Hospitalization odds ratio for each coefficient of the prediction named model LR 2. The p-values for the null hypothesis that a coefficient is equal to zero (i.e., the odds ratio is equal to one), were computed with a Wald test (Wald, 1943).

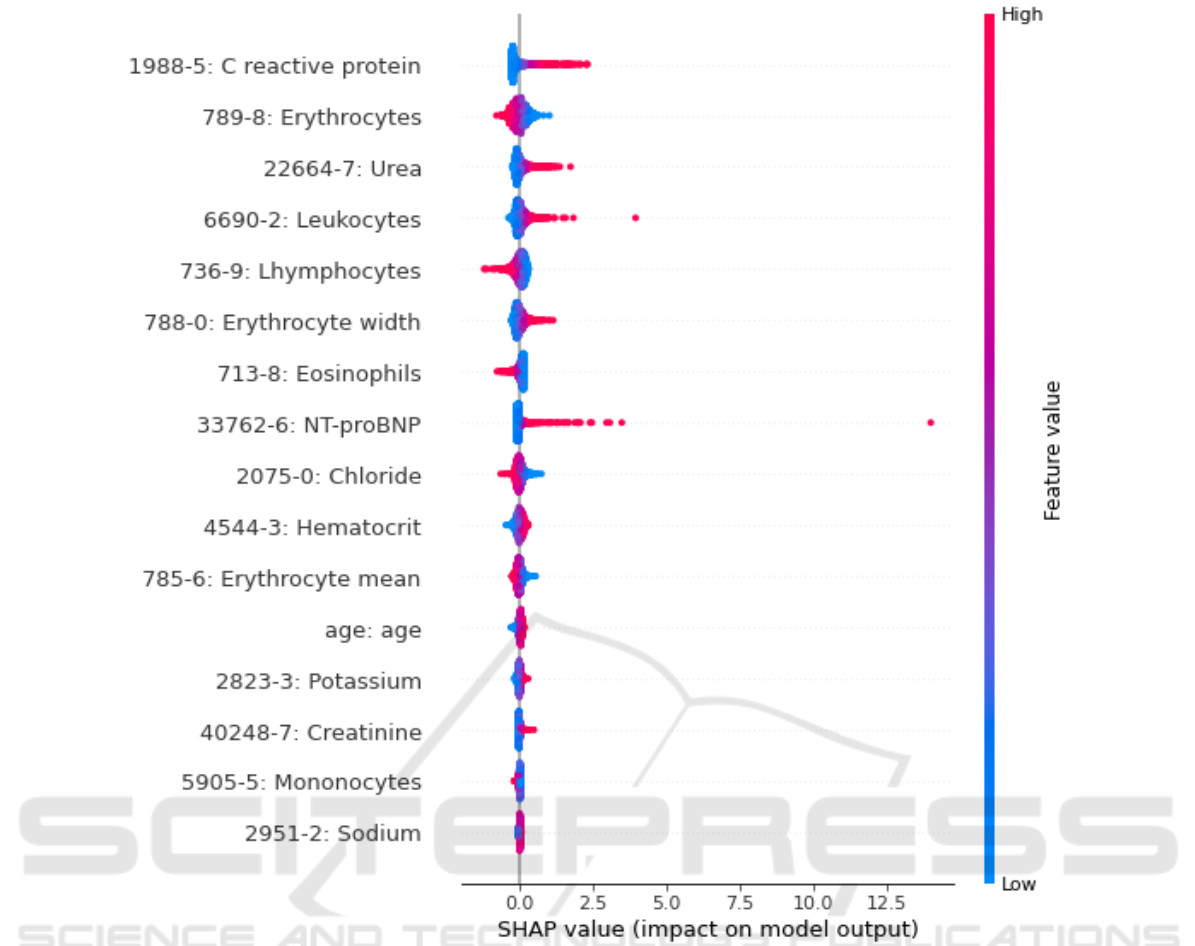
Variable	**Component**	**odds ratio** ($P > \|z\|$)
1988-5	C-reactive protein	1.4689 (< 0.05)
33762-6	NT-proBNP	1.2919 (< 0.05)
6690-2	Leukocytes	1.2359 (< 0.05)
22664-7	Urea	1.2339 (< 0.05)
788-0	Erythrocyte distribution width	1.1745 (< 0.05)
4544-3	Hematocrit	1.1118 (0.068)
Age	Age	1.0597 (< 0.05)
2823-3	Potassium	1.0581 (< 0.05)
40248-7	Creatinineˆbaseline	1.0414 (0.353)
2951-2	Sodium	1.0110 (0.805)
5905-5	Monocytes/100 leukocytes	0.9764 (0.337)
785-6	Erythrocyte mean corpuscular hemoglobin	0.9037 (< 0.05)
713-8	Eosinophils/100 leukocytes	0.8829 (< 0.05)
2075-0	Chloride	0.8816 (< 0.05)
736-9	Lymphocytes/100 leukocytes	0.8463 (< 0.05)
789-8	Erythrocytes	0.7980 (< 0.05)

cians can more accurately gauge the need for hospital admission of these patients, and hospital staff can obtain early estimates of admission rates that can, for instance, lead to improved efficiency in hospital bed planning and resource allocation.

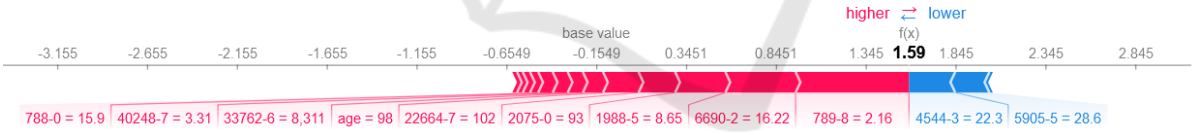
Our machine learning models prioritized interpretability, thereby enhancing trust and clinical applicability. Rigorous evaluation using logistic regression weights and SHAP values ensured transparent and practically relevant outcomes, vital for real-world applications and future research (Lundberg and Lee, 2017). This methodology permits detailed interpretability, allowing clinicians to perform individualized patient risk assessments, significantly improving clinical utility. Our findings indicate that elevated levels of NT-proBNP and CRP are positively correlated with imminent hospital admission, consistent with prior studies (Bozkurt et al., 2021; Anand et al., 2005). NT-proBNP is a well-established marker of HF severity, while the link between CRP and HF prognosis, suggesting possible concurrent infections or unaddressed inflammatory diseases, warrants further investigation. Moreover, we identified a negative correlation between erythrocyte concentration and patient outcomes, reinforcing existing evidence of anemia's adverse impact on HF prognoses, including higher hospitalization rates (Anand et al., 2004). Intriguingly, our analysis revealed that when laboratory data is included, age becomes a less significant predictor of imminent hospitalization.

Furthermore, the study highlights the utility of the SHAP force plot in assessing individual patient risks, offering a detailed insight into the specific impacts of features on model predictions. This analytical tool increases the model's clinical relevance by elucidating not just the direction but also the magnitude of a feature's impact on predictions of imminent hospitalization. Such precise feature-level interpretability is invaluable for predicting heightened risks in scenarios with interacting factors, rendering it a potent instrument for handling complex clinical scenarios like multimorbidity.

This study also has limitations that warrant discussion. A primary limitation is the analyzed data coming from a single hospital, which may be the source of biases associated with the diversity of diseases treated and the complexity of healthcare delivery at this facility. Although the hospital provides a range of care services, the lack of universal primary healthcare could limit the scope of our analysis, thus restricting the depth of insights into disease complexity and nuances in healthcare provision. Moreover, our focus on laboratory test prescriptions might overlook essential aspects of patient histories and experiences prior to admission to the ED, which are captured in different data types. Integration with alternative data sources, such as clinical notes and drug prescriptions, could bolster confidence in our findings. Our observed correlations between prescription patterns and clinical outcomes do suggest implicit clinical and organizational processes that are interesting research avenues to explore. However, extra caution is advised in interpreting their significance. Not only is further external validation necessary, but there is also a need to be mindful of potential biases that this approach may introduce and perpetuate, such as discrimination



(a) *SHAP summary plot*



(b) *SHAP force plot*

Figure 4: (a) SHAP summary plot computed using laboratory results and age of all episodes of the test dataseet. (b) SHAP force plot of a HF patient correctly classified as imminent hospitalization, in which colour red represents the lab tests results (or age) that are increasing the chance of imminent hospitalization while colour blue represents the negative outcome.

against underrepresented subpopulations. This issue has been increasingly recognized in the literature and merits further investigation before any clinical implementation (Obermeyer et al., 2019).

Additionally, the inclusion of more comprehensive clinical information could enhance data interpretation and improve algorithm performance. For example, ejection fraction, a key factor in heart failure (HF), was not included in our study. Often recorded in free text, this information presents challenges for systematic and reliable extraction and was consequently not utilized. We intend to address this limitation in future work.

Finally, it is essential to emphasize that the tools developed in this study just went through an initial trial stages. The utilization of AI systems to aid clinical decision-making represents a significant innovation in healthcare. However, their adoption requires meticulous evaluation and strict adherence to regulations, particularly in Europe. The tools and methodologies described in our research serve as illustrations of potential approaches to enhance clinical decision-making by leveraging existing clinical data with its inherent limitations. Prior to their implementation in

real-world clinical settings, these tools must undergo comprehensive regulatory processes to ensure their safety, effectiveness, and ethical compliance.

6 CONCLUSION

In conclusion, this study has successfully developed a comprehensive framework for analyzing clinical data, particularly in the context of patients with HF with concurrent multimorbidity. Our approach contributes to more refined risk stratification and informed clinical decision-making. This methodology showcases the potential of healthcare data to improve clinical insight and individualized risk assessment that can eventually lead to better patient outcomes.

Looking ahead, our research lays the groundwork for future investigations to enhance predictive models that make use of laboratory data and delve deeper into the impacts of various comorbidities on HF outcomes on long-term perspective. Prioritizing the expansion of data collection methods, will be essential to enrich the quality and relevance of the data in future studies. Furthermore, the versatility of our analytical framework holds promise for broader applications, extending to diverse patient populations with chronic conditions such as Diabetes Mellitus, Chronic Kidney Disease, and Chronic Obstructive Pulmonary Disease. We intend to address these in future work.

FUNDING

This work was developed under the IntelligentCare project LISBOA-01-0247-FEDER-045948 which is co-financed by the ERDF/LISBOA2020 and by FCT, Portugal, under CMU-Portugal and by FCT, Portugal, through the INESC-ID Research Unit, ref. UIDB/00408/2020 and ref. UIDP/00408/2020.

ACKNOWLEDGEMENTS

We acknowledge Carlos Magalhães and Jaime Machado, from Hospital da Luz, for the data extraction.

REFERENCES

Alanazi, A. (2022). Using machine learning for healthcare challenges and opportunities. *Informatics in Medicine Unlocked*, 30:100924.

Anand, I., McMurray, J. J., Whitmore, J., Warren, M., Pham, A., McCamish, M. A., and Burton, P. B. (2004). Anemia and Its Relationship to Clinical Outcome in Heart Failure. *Circulation*, 110(2):149–154.

Anand, I. S., Latini, R., Florea, V. G., Kuskowski, M. A., Rector, T., Masson, S., Signorini, S., Mocarelli, P., Hester, A., Glazer, R., and Cohn, J. N. (2005). C-Reactive Protein in Heart Failure. *Circulation*, 112(10):1428–1434.

Aronson, S. J. and Rehm, H. L. (2015). Building the foundation for genomics in precision medicine. *Nature*, 526(7573):336–342.

Asadollahi, K., Hastings, I., Beeching, N., and Gill, G. (2007). Laboratory risk factors for hospital mortality in acutely admitted patients. *QJM : monthly journal of the Association of Physicians*, 100:501–7.

Athena (2022). Ohdsi athena. (accessed: 13.01.2023).

Bozkurt, B., Coats, A. J., Tsutsui, H., and et al. (2021). Universal Definition and Classification of Heart Failure: A Report of the Heart Failure Society of America, Heart Failure Association of the European Society of Cardiology, Japanese Heart Failure Society and Writing Committee of the Universal Definition of Heart Failure. *Journal of Cardiac Failure*, 27(4):387–413.

Chamberlain, A. M., Sauver, J. L., Gerber, Y., Manemann, S. M., Boyd, C. M., Dunlay, S. M., Rocca, W. A., Rutten, L. J., Jiang, R., Weston, S. A., and Roger, V. L. (2015). Multimorbidity in Heart Failure: A Community Perspective. *The American journal of medicine*, 128(1):38.

CMS (2021). Electronic Health Records.

Dervishi, A. (2020). A deep learning backcasting approach to the electrolyte, metabolite, and acid-base parameters that predict risk in icu patients. *PLOS ONE*, 15(12):1–19.

Dixon, B. E., Wen, C., French, T., Williams, J. L., Duke, J. D., and Grannis, S. J. (2020). Observational Health Data Science and Informatics (OHDSI). *BMJ Health Care Inform*, 27:100054.

Han, J., Kamber, M., and Pei, J. (2012). *Data mining: Data mining concepts and techniques*. Morgan Kaufmann, 3 edition.

Hitachi Ventara. Pentaho Data Integration & Analytics.

Hripcsak, G., Knirsch, C., Zhou, L., Wilcox, A., and Melton, G. (2011). Bias associated with mining electronic health records. *Journal of Biomedical Discovery and Collaboration*, 6:48–52.

Kasza, J. and Wolfe, R. (2014). Interpretation of commonly used statistical regression models. *Respirology*, 19(1):14–21.

Lee, T. C., Shah, N. H., Haack, A., and Baxter, S. L. (2020). Clinical implementation of predictive models embedded within electronic health record systems: A systematic review. *Informatics*, 7(3):25. Epub 2020 Jul 25.

Liyanage, H., Liaw, S. T., Jonnagaddala, J., Hinton, W., and De Lusignan, S. (2018). Common Data Models (CDMs) to Enhance International Big Data Analytics: A Diabetes Use Case to Compare Three CDMs. *Studies in Health Technology and Informatics*, 255:60–64.

Loekito, E., Bailey, J., Bellomo, R., Hart, G. K., Hegarty, C., Davey, P., Bain, C., Pilcher, D., and Schneider, H. (2013). Common laboratory tests predict imminent medical emergency team calls, intensive care unit admission or death in emergency department patients. *Emergency Medicine Australasia*, 25(2):132–139.

Loinc® Indianapolis, IN: Regenstrief Institute, Inc. Logical Observation Identifiers Names and Codes (LOINC).

Lundberg, S. M. and Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, page 4768–4777, Red Hook, NY, USA. Curran Associates Inc.

Lundberg, S. M., Nair, B., Vavilala, M. S., Horibe, M., Eisses, M. J., Adams, T., Liston, D. E., Low, D. K.-W., Newman, S.-F., Kim, J., et al. (2018). Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nature Biomedical Engineering*, 2(10):749.

Majnarić, L. T., Babič, F., O'Sullivan, S., and Holzinger, A. (2021). AI and big data in healthcare: Towards a more comprehensive research framework for multimorbidity. *Journal of Clinical Medicine*, 10(4):766. Number: 4 Publisher: Multidisciplinary Digital Publishing Institute.

Martins, C., Neves, B., Teixeira, A. S., Froes, M., Sarmento, P., Machado, J., Magalhães, C., Silva, N., Silva, M., and Leite, F. (2024). Identifying subgroups in heart failure patients with multimorbidity by clustering and network analysis. *BMC Medical Informatics and Decision Making*, 24(1):95.

Mueller, O., Rentsch, K., Nickel, C., and Bingisser, R. (2021). Disposition decision support by laboratory based outcome prediction. *Journal of Clinical Medicine*, 10:939.

Navickas, R., Petric, V.-K., Feigl, A. B., and Seychell, M. (2016). Multimorbidity: What Do We Know? What Should We Do? *Journal of Comorbidity*, 6(1):4–11.

Nwanosike, E. M., Conway, B. R., Merchant, H. A., and Hasan, S. S. (2022). Potential applications and performance of machine learning techniques and algorithms in clinical practice: A systematic review. *International Journal of Medical Informatics*, 159:104679.

Obermeyer, Z., Powers, B., Vogeli, C., and Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464):447–453.

OHDSI (2014). Achilles for data characterization. (accessed: 13.01.2023).

OHDSI (2021). Whiterabbit for etl design. (accessed: 13.01.2023).

Plebani, M. and Lippi, G. (2016). Improving diagnosis and reducing diagnostic errors: The next frontier of laboratory medicine. *Clinical Chemistry and Laboratory Medicine*, 54(7):1117–1118.

Rodrigues, A. M., Gregório, M. J., Sousa, R. D., Dias, S. S., Santos, M. J., Mendes, J. M., Coelho, P. S., Branco, J. C., and Canhão, H. (2018). Challenges of ageing in portugal: Data from the EpiDoC cohort. *Acta Medica Portuguesa*, 31(2):80–93.

Romana, G. Q., Kislaya, I., Salvador, M. R., Cunha-Goncalves, S., Nunes, B., and Dias, C. (2019). Multimorbidity in portugal: results from the first national health examination survey. *Acta Médica Portuguesa*, 32(1).

Sciences, O. H. D. and Informatics(OHDSI) (2023). Standardized data: The omop common data model. (accessed: 22.11.2023).

Shamout, F., Zhu, T., and Clifton, D. (2020). Machine learning for clinical outcome prediction. *IEEE Reviews in Biomedical Engineering*, PP:1–1.

Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society*, 54:426–482.

WHO (2016). Multimorbidity Technical Series on Safer Primary Care Multimorbidity: Technical Series on Safer Primary Care. page 28.

Wians, F. H. (2009). Clinical laboratory tests: Which, why, and what do the results mean? *Laboratory Medicine*, 40(2):105–113.

Williams, T. B., Garza, M., Lipchitz, R., Powell, T., Baghal, A., Swindle, T., and Sexton, K. W. (2022). Cultivating informatics capacity for multimorbidity: A learning health systems use case. *Journal of Multimorbidity and Comorbidity*, 12. Publisher: SAGE Publications.