HANOI UNIVERSITY OF SCIENCE AND TECHNOLOGY

GRADUATION THESIS

Prediction of In-hospital Mortality in Mechanically Ventilated Patients with Congestive Heart Failure

NGUYEN MINH HUYEN
huyen.nm206652@sis.hust.edu.vn

Major: Information Technology
Specialization: Global ICT

Supervisor: Dr Nguyen Hong Quang ______________

Signature

Department: Computer Engineering

School: School of Information and Communications Technology

HANOI, 12/2024

ACKNOWLEDGMENT

During the time I worked on this project, I have learned so much and gained valuable experience. These achievements would not have been possible without the support and help of many people.

First and foremost, I would like to express my deepest gratitude to my advisor, Dr. Nguyễn Hồng Quang, for his guidance, support, and encouragement. From my third year, he has given me numerous opportunities to participate in various projects, helping me realize the value of my hard work and feel that I am growing and improving every day. His careful, patient, and gentle mentoring has been invaluable to me. I consider myself incredibly fortunate to have had the opportunity to learn from him.

I would also like to extend my heartfelt thanks to my family, especially my parents and my sister. Thank you for always being my source of motivation, for encouraging me to strive and work harder, and for allowing me to pursue what I love while believing in my abilities. Your unwavering trust and support mean the world to me.

Additionally, I am deeply thankful to my beloved partner for his care, love, and support throughout the journey of completing my thesis. I couldn’t have achieved this without you by my side, being my comfort zone. Thank you for always being there for me.

Finally, I would like to thank my friends for all the precious moments we shared during our time together, the unforgettable memories in class, and for always being willing to help each other so that we can all grow and succeed.

Thank you all for being a part of this journey and for making it meaningful and fulfilling.

PLEDGE

Student's Full Name: Nguyen Minh Huyen

Student ID: 20206652

Contact Phone Number: 0835996812

Email: huyen.nm206652@sis.hust.edu.vn

Class: ICT-02 Cohort 65

Program: Global ICT

I, Nguyen Minh Huyen, commit that the Graduation Thesis (GT) is my own research work under the guidance of Doctor Nguyen Hong Quang. The results presented in the GT are truthful and are my own achievements, not copied from any other works. All references in the GT, including images, tables, data, and quotations, are clearly and fully cited in the bibliography. I take full responsibility for any violations of the school's regulations concerning plagiarism.

Hanoi, 2024

Student

(Signature and full name)

ABSTRACT

In-hospital mortality among mechanically ventilated congestive heart failure CHF patients is a pressing concern, requiring accurate risk prediction to enhance clinical care. This study aimed to develop and validate a predictive model for mortality using advanced machine learning techniques and feature selection strategies.

Patient data from the MIMIC-IV database were analyzed, starting with 90 features narrowed to 60 using backward elimination. Seven machine learning models, including GATE, were evaluated through nested k-fold cross-validation. To improve performance, three GATE models trained on first, last, and mean value data were combined into an ensemble. SHAP analysis was employed to identify 20 key features, and a compact model was developed and optimized based on this streamlined feature set. Performance metrics, including AUROC, were used to evaluate model accuracy.

Data from 8,517 patient stays were analyzed. The GATE model outperformed the other six models, with the ensemble demonstrating enhanced generalizability. SHAP analysis identified critical features such as age, MV duration, SAPSII and LODS scores, lactate levels, and markers of organ dysfunction. The compact GATE model, built on 20 essential features, achieved an AUROC of 0.87, closely matching the full model’s 0.88, while improving computational efficiency and interpretability.

The GATE-based prediction model provides robust and accurate identification of mortality risk in mechanically ventilated CHF patients. The compact model offers a practical balance between performance and efficiency, making it a valuable tool for clinical decision-making and resource allocation.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation	Definition
AUROC	Area Under the Receiver Operating Characteristic
CHF	Congestive Heart Failure
DL	Deep learning
GATE	Gated Additive Tree Ensemble
HF	Heart Failure
ICU	Intensive care unit
kNN	k-Nearest Neighbors
LODS	Logistic Organ Dysfunction System
MIMIC-IV	Medical Information Mart for Intensive Care IV
ML	Machine learning
MV	Mechanical ventilation
NODE	Neural Oblivious Decision Ensembles
OASIS	Oxford Acute Severity of Illness Score
SAPS II	Simplified Acute Physiology Score II
SHAP	SHapley Additive Explanations
SOFA	Sequential Organ Failure Assessment
XGBoost	Extreme Gradient Boosting

CHAPTER 1. INTRODUCTION

1.1 Problem Statement

Heart failure (HF) is a clinical syndrome resulting from structural or functional cardiac abnormalities, leading to symptoms such as shortness of breath, fatigue, and fluid retention. In 2021, HF was universally defined as a condition characterized by elevated levels of natriuretic peptides or evidence of pulmonary or systemic congestion [1]. Congestive heart failure (CHF), a more specific and advanced stage of HF, is characterized by significant fluid buildup and congestion. Due to its severity and complexity, CHF often requires intensive care unit (ICU) admission for management [2].

Mechanical ventilation, a critical life-support intervention, is frequently used in the ICU to manage CHF patients experiencing respiratory failure. Although life-saving, this intervention carries substantial risks, including hemodynamic instability, alveolar inflammation, and complications caused by altered intrathoracic pressures [3]. These challenges highlight the urgent need for advanced predictive tools to assess patient outcomes, aiding more effective clinical decision-making and improving management for this high-risk group.

Predictive modeling is an indispensable tool in healthcare, offering the ability to forecast patient outcomes based on clinical data, thereby enhancing decision-making. For mechanically ventilated CHF patients, predictive models assist in identifying individuals who might benefit from early interventions, intensive monitoring, or alternative therapeutic strategies. However, predicting early mortality in this population remains challenging due to the intricate interplay of systemic organ dysfunction, comorbidities, and rapidly changing clinical conditions.

Traditional scoring systems, such as SOFA, APACHE-II, LODS, MODS, and SAPS-II, are significantly associated with mortality in critically ill patients and have demonstrated utility in mortality prediction [4]. While the SOFA and GWTG-HF risk scores provide valuable insights, they have notable limitations. The SOFA score, for example, evaluates organ failure but lacks specificity for heart failure, whereas the GWTG-HF score captures HF-specific risks but insufficiently addresses systemic dysfunction. Additionally, these tools face challenges such as incomplete clinical data, including missing arterial blood gas measurements, and their static nature, which does not accommodate the dynamic changes in a patient's condition [5]. Population variability and the need for interpretable, actionable predictions