Journal of Health and Biomedical Informatics 2024; 11(1): 83-95
doi: 10.34172/jhbmi.2024.15
Medical Informatics Research Center, Kerman University of Medical Sciences

Original article

Prediction of 30-day Mortality for Heart Failure Patients with Cardiogenic Shock Using Random Forest Hyperparameter Tuning

Shiva Kanani[1]*, Iraj Mahdavi[2], Naghmeh Ziaie[3], Bagher Rahimpour Cami[4]

1. Ph.D. Student in Industrial Engineering, Mazandaran University of Science and Technology, Babol, Iran
2. Professor, Department of Industrial Engineering, University of Science and Technology of Babol, Babol, Iran
3. Assistant Professor, Department of Cardiology, Fellowship of Heart Failure and Transplantation, Babol University of Medical Sciences, Rouhani Hospital Research Center, Babol, Iran
4. Assistant Professor, Department of Computer Engineering, University of Science and Technology of Babol, Babol, Iran

ARTICLE INFO:

Article History:

Received: 3 May 2024

Accepted: 8 Jun 2024

Published: 30 Jun 2024

Corresponding Author:

Shiva Kanani

Email:
shiva.kanani400@gmail.com

Citation: Kanani S, Mahdavi I, Ziaie N, Rahimpour Cami B. Prediction of 30-day Mortality for Heart Failure Patients with Cardiogenic Shock Using Random Forest Hyperparameter Tuning. Journal of Health and Biomedical Informatics 2024; 11(1): 83-95. [In Persian]

Abstract

Introduction: Heart failure is a clinical syndrome resulting from structural or functional abnormalities of the heart, leading to reduced cardiac output or increased intracardiac pressure. When combined with cardiogenic shock, it becomes an emergency condition with a high mortality rate, necessitating immediate diagnosis and treatment. Accurate prediction of 30-day mortality in these patients is vital for timely care and patient survival. This study aimed to optimize the Random Forest algorithm by adjusting hyperparameters to more accurately predict 30-day mortality in heart failure patients with cardiogenic shock.

Method: In this research, data from 201 cardiac patients aged over 18 years who experienced cardiogenic shock at Rouhani Hospital in Babol in 2020, were used. Thirty-four selected features such as age, history of cardiac surgery, pH, lactate levels, diabetes, etc., were examined, and their one-month mortality was tracked through telephone follow-ups.

Results: The results showed that increasing age (above 57 years), decreasing pH (below 7.3), and elevating lactate levels (above 2) significantly increased the risk of 30-day mortality. By optimizing the hyperparameters of the Random Forest algorithm (ntree=1000 and mtry=14), prediction accuracy improved from 66.0% to 71.8%.

Conclusion: This study demonstrates that the accuracy of the Random Forest algorithm depends on its input hyperparameters and that optimizing these parameters can lead to a more precise prediction of mortality in heart failure patients with cardiogenic shock. With appropriate optimization, this algorithm can serve as an effective tool for the early detection of high-risk patients and timely provision.

Keywords: Heart Failure, Cardiogenic Shock, Random Forest

© 2024 The Author(s); Published by Kerman University of Medical Sciences. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cite

[Downloaded from jhbmi.ir on 2026-11-23]
[DOI: 10.34172/jhbmi.2024.15]

مجله انفورماتیک سلامت و زیست پزشکی دوره یازدهم، شماره اول، سال ۱۴۰۳ ، ص ۹۵-۸۳
doi: 10.34172/jhbmi.2024.15
مرکز تحقیقات انفورماتیک پزشکی، دانشگاه علوم پزشکی کرمان

مقاله پژوهشی

پیش‌بینی مرگ‌ومیر ۳۰ روزه در بیماران نارسایی قلبی با شوک کاردیوژنیک با استفاده از تنظیم هایپرپارامترهای جنگل تصادفی

شیوا کنعانی[۱*]، ایرج مهدوی[۲]، نغمه ضیایی[۳]، باقر رحیم پورکامی[۴]

۱. دانشجوی دکتری مهندسی صنایع، دانشگاه علوم و فنون بابل، بابل، ایران
۲. استاد، گروه مهندسی صنایع، دانشگاه علوم و فنون بابل، بابل، ایران
۳. استادیار، گروه قلب فلوشیپ نارسایی قلب و پیوند، دانشگاه علوم پزشکی بابل، مرکز تحقیقات بیمارستان روحانی بابل، بابل، ایران
۴. استادیار، گروه مهندسی کامپیوتر، دانشگاه علوم و فنون بابل، بابل، ایران

اطلاعات مقاله:

سابقه مقاله:

دریافت: ۱۴۰۳/۲/۱۴
پذیرش: ۱۴۰۳/۳/۱۹
انتشار برخط: ۱۴۰۳/۳/۳۱

نویسنده مسئول:
شیوا کنعانی

ایمیل:
shiva.kanani400@gmail.com

ارجاع: کنعانی شیوا، مهدوی ایرج، ضیایی نغمه، رحیم پورکامی باقر. پیش‌بینی مرگ‌ومیر ۳۰ روزه در بیماران نارسایی قلبی با شوک کاردیوژنیک با استفاده از تنظیم هایپرپارامترهای جنگل تصادفی. مجله انفورماتیک سلامت و زیست پزشکی ۱۴۰۳؛ ۱۱(۱): ۹۵-۸۳.

چکیده

مقدمه: نارسایی قلبی یک سندرم کلینیکی است که به دنبال ناهنجاری ساختاری یا عملکردی قلب ایجاد شده و باعث کاهش خون پمپاژ شده یا افزایش فشار داخل قلبی می‌شود. نارسایی قلبی با شوک کاردیوژنیک یک وضعیت اورژانسی با نرخ مرگ‌ومیر بالا است که نیازمند تشخیص و درمان فوری است. پیش‌بینی دقیق مرگ‌ومیر ۳۰ روزه در این بیماران برای ارائه مراقبت‌های به‌موقع و نجات جان بیماران حیاتی است. این مطالعه به بهینه‌سازی الگوریتم جنگل تصادفی با تنظیم هایپرپارامترها برای پیش‌بینی دقیق‌تر مرگ‌ومیر ۳۰ روزه بیماران نارسایی قلبی با شوک کاردیوژنیک می‌پردازد.

روش کار: این تحقیق از داده‌های ۲۰۱ بیمار قلبی بالای ۱۸ سال که در بیمارستان روحانی بابل در سال ۱۳۹۹ دچار شوک کاردیوژنیک شده‌اند، استفاده می‌کند. ۳۴ ویژگی مانند سن، سابقه جراحی قلب باز، pH، لاکتات، دیابت و غیره مورد بررسی قرار گرفت و مرگ یک‌ماهه آن‌ها از طریق تماس تلفنی پیگیری شد.

یافته‌ها: نتایج نشان داد با افزایش سن (بیش از ۵۷ سال)، کاهش pH (کمتر از ۷/۳) و افزایش لاکتات (بیش از ۲)، خطر مرگ ۳۰ روزه افزایش می‌یابد. با تنظیم بهینه هایپرپارامترهای الگوریتم جنگل تصادفی (ntree= ۱۰۰۰ و mtry= ۱۴)، دقت پیش‌بینی از ۰/۶۶٪ به ۷۱/۸٪ ارتقاء یافت.

نتیجه‌گیری: این مطالعه نشان می‌دهد که دقت الگوریتم جنگل تصادفی وابسته به هایپرپارامترها است و با بهینه‌کردن این پارامترها می‌توان پیش‌بینی دقیق‌تری از مرگ‌ومیر بیماران نارسایی قلبی با شوک کاردیوژنیک داشت. این الگوریتم با بهینه‌سازی مناسب، می‌تواند ابزاری کارآمد برای تشخیص زودهنگام بیماران در معرض خطر و ارائه خدمات درمانی به‌موقع باشد.

کلیدواژه‌ها: نارسایی قلبی، شوک کاردیوژنیک، جنگل تصادفی، تنظیم هایپرپارامتر، پیش‌بینی مرگ‌ومیر

[Downloaded from jhbmi.ir on 2026-11-23]
[DOI: 10.34172/jhbmi.2024.15]

مقدمه

بیماری‌های قلبی یکی از علل اصلی مرگ‌ومیر افراد می‌باشد. پیش‌بینی می‌شود تا سال ۲۰۳۰ مرگ‌ومیر ناشی از بیماری‌های قلبی-عروقی به ۲۳/۳ میلیون نفر افزایش یابد [۱]. طبق گزارش سازمان جهانی بهداشت، سالانه حدود ۱۲ میلیون نفر قربانی بیماری قلبی می‌شوند و ۸۰ درصد از مردم به دلیل بیماری قلبی جان خود را از دست می‌دهند [۲]. یکی از انواع بیماری‌های قلبی، نارسایی قلبی می‌باشد. نارسایی قلبی یک سندرم کلینیکی است که به دنبال ناهنجاری ساختاری و یا عملکردی قلب ایجاد می‌شود و باعث کاهش خون پمپاژ شده از قلب یا افزایش فشار داخل قلبی می‌گردد و با علائمی نظیر تنگی نفس، ارتوپنه، تورم اندام تحتانی مشخص می‌شود و در معاینه بالینی احتقان ریوی و افزایش فشار ورید ژوگولر یافت می‌شود[۳]. در شوک کاردیوژنیک که یکی از عوارض جدی نارسایی قلبی می‌باشد، ایسکمی (کمبود خون‌رسانی)، کاهش قدرت انقباضی ماهیچه قلب (میوکارد) و خروجی خون قلبی (دِبی قلبی) همگی منجر به یک شوک گسترده می‌شوند که نیازمند تشخیص و درمان فوری است؛ لذا یک وضعیت اورژانسی پزشکی محسوب می‌شود [۴]. پیش‌بینی دقیق مرگ و میر بیماران مبتلا به نارسایی قلبی با شوک کاردیوژنیک، نقشی حیاتی در ارجاع و اقدامات درمانی مناسب ایفا می‌کند. این امر به پزشکان کمک می‌کند تا با درک بهتر ریسک مرگ و میر بیمار، بهترین تصمیمات را برای مراقبت از او اتخاذ کنند.

الگوریتم‌های یادگیری ماشین، ابزاری قدرتمند برای پیش‌بینی در زمینه‌های مختلف، از جمله پزشکی، هستند. این الگوریتم‌ها می‌توانند از حجم عظیمی از داده‌ها برای یادگیری الگوها و پیش‌بینی نتایج آینده استفاده کنند [۵]. در این مطالعه، از الگوریتم جنگل تصادفی، نوعی الگوریتم یادگیری ماشین که از مجموعه‌ای از درختان تصمیم برای پیش‌بینی احتمال یک رویداد استفاده می‌کند، برای پیش‌بینی مرگ و میر بیماران مبتلا به نارسایی قلبی با شوک کاردیوژنیک استفاده شده است. الگوریتم جنگل تصادفی دارای چندین پارامتر است که به‌طور قابل توجهی بر عملکرد آن تأثیر می‌گذارد. در اکثر مدل‌هایی که از الگوریتم جنگل تصادفی استفاده کرده‌اند، پارامترهای آن به صورت پیش فرض و یا به صورت آزمون و خطا توسط مدیر تنظیم شده‌اند. درک این پارامترها به ما این امکان را می‌دهد که الگوریتم را برای نتایج بهینه در کاربردهای مختلف تنظیم دقیق کنیم [۶].

در این راستا، مطالعاتی در زمینه بهینه کردن و تنظیم هایپرپارامترهای الگوریتم جنگل تصادفی انجام پذیرفته است. مطالعه König و همکاران به بررسی یادگیری ماشین (Machine Learning) ML برای پیش‌بینی مرگ و میر در بیمارستان در بیماران با نارسایی قلبی (Heart Failure) HF پرداخت. با تحلیل داده‌های اداری از بیمارستان‌های آلمان، آن‌ها تحلیل رگرسیون را با الگوریتم‌های مختلف ML مقایسه کردند. تمامی مدل‌های ML، از جمله جنگل تصادفی با پارامترهای بهینه‌سازی شده از طریق جستجوی شبکه‌ای، در پیش‌بینی مرگ و میر عملکرد بهتری نسبت به رگرسیون داشتند که نشان دهنده پتانسیل قابل توجه ابرپارامترهای بهینه‌شده برای ارتقای دقت پیش‌بینی مرگ و میر در بیمارستان در بیماران مبتلا به (Heart Failure) HF می‌باشد [۷].

Moubayed و همکاران یک مدل جنگل تصادفی بهینه‌سازی شده برای شناسایی بات‌نت‌ها یا وظایف طبقه‌بندی پیشنهاد می‌کنند. روش آن‌ها از کسب اطلاعات برای انتخاب ویژگی‌ها و از الگوریتم ژنتیک برای تنظیم دقیق هایپرپارامترهای مدل جنگل تصادفی استفاده می‌کند. این تمرکز بر بهینه‌سازی هایپرپارامترها فراتر از استفاده سنتی از جنگل تصادفی است و هدف آن بهبود دقت عملکرد کلی مدل می‌باشد [۸].

در مطالعه‌ای دیگر، Atiya و Kamal به بررسی یادگیری ماشین برای پیش‌بینی مرگ‌ومیر ناشی از COVID-19 پرداختند. آن‌ها الگوریتم‌های مختلفی از جمله شبکه‌های عصبی، SVM (Support Vector Machine) و جنگل تصادفی را با روش‌های موجود مقایسه کرده‌اند. به طور خاص، از جستجوی شبکه‌ای با اعتبارسنجی متقاطع ۱۰ مرحله‌ای استفاده کرده‌اند که نشان‌دهنده تنظیم هایپرپارامترها، به‌ویژه برای جنگل تصادفی است. این موضوع بر تمرکز آن‌ها بر بهینه‌سازی عملکرد مدل برای بهبود دقت پیش‌بینی مرگ‌ومیر تأکید دارد [۹].

Bischl و همکاران، مدلی را برای شناسایی بیمارانی که در معرض خطر ابتلا به درد شدید پس از عمل هستند، توسعه داده‌اند که این مدل از الگوریتم جنگل تصادفی استفاده می‌کند و به تنظیمات هایپرپارامتر مدل به بهینه‌سازی پارامترهایی مانند تعداد درختان و عمق

[Downloaded from jhbmi.ir on 2026-11-23]
[DOI: 10.34172/jhbmi.2024.15]

درختان می‌پردازد. این پژوهش بر اثربخشی یادگیری ماشین و تنظیم هایپرپارامتر در افزایش دقت پیش‌بینی نتایج بالینی تأکید می‌کند [۱۰].

Iwase و همکاران در مطالعه‌ای با عنوان الگوریتم پیش‌بینی مرگ و میر و مدت بستری در ICU (Intensive Care Unit) با استفاده از یادگیری ماشین از مدل‌های یادگیری ماشین مانند جنگل تصادفی، ماشین‌های تقویت گرادیان و شبکه‌های عصبی برای پیش‌بینی مرگ و میر و مدت بستری در بخش مراقبت‌های ویژه (ICU) استفاده کردند. پس از تنظیم هایپرپارامترها، دقت پیش‌بینی مدل‌های مختلف از جمله جنگل تصادفی، ماشین تقویت گرادیان و شبکه‌های عصبی با استفاده از معیارهایی مانند «مساحت زیر منحنی راهیاب عملیاتی گیرنده (Area Under the Curve Receiver Operating Characteristic) AUC-ROC ارزیابی شدند و مدل جنگل تصادفی بهینه شده، بالاترین ارزش پیش‌بینی را در بین رویکردهای یادگیری ماشین نشان داد [۱۱].

درحالی‌که Breiman در مطالعه خود بر خلاف مطالعه حاضر به تأثیر افزایش تعداد درختان بر کاهش بازدهی تأکید دارد، Oshiro و همکاران این موضوع را با استفاده از AUC کمّی‌سازی می‌کنند و یک آستانه را پیشنهاد می‌کنند که تحت‌تأثیر منابع محاسباتی قرار دارد؛ اما هر دو مطالعه بر اهمیت تنظیم هایپرپارامترها برای دستیابی به نتایج بهینه تأکید دارند [۱۲،۱۳].

در این مطالعه به دنبال ارائه یک مدل پیش‌بینی بادقت بالاتر نسبت به مدل‌های موجود که می‌تواند به‌نظام سلامت و اقدامات درمانی ارائه شده به بیماران نارسایی قلبی با پیش‌بینی دقیق مرگ‌ومیر بیماران نارسایی قلبی با شوک کاردیوژنیک کمک کند هستیم. تنظیم هایپرپارامترهای الگوریتم جنگل تصادفی برای دستیابی به بهترین دقت در پیش‌بینی مرگ‌ومیر بیماران مبتلا به نارسایی قلبی با شوک کاردیوژنیک (مقدار بهینه دو پارامتر mtry و ntree که بر مدل جنگل تصادفی ما تأثیر می‌گذارند)، نیز از جمله این پژوهش به شمار می‌رود. انتظار می‌رود نتایج این مطالعه به درک عمیق‌تر از پتانسیل الگوریتم جنگل تصادفی در پیش‌بینی مرگ‌ومیر بیماران مبتلا به نارسایی قلبی با شوک کاردیوژنیک و بهبود کیفیت مراقبت از این بیماران منجر شود.

روش کار

این مطالعه از نوع کاربردی – توصیفی بود و از داده‌های بیماران مبتلا به نارسایی قلبی بالای ۱۸ سال که در بدو مراجعه در بیمارستان روحانی شهر بابل در کشور ایران دچار تابلوی شوک کاردیوژنیک شده‌اند، استفاده شد. اطلاعات دموگرافیک و بالینی و پاراکلینیکی بیماران در پرسشنامه‌هایی طراحی و تکمیل شد و مرگ‌ومیر یک‌ماهه آن‌ها از طریق تماس تلفنی بررسی گردید. روش نمونه‌گیری این پژوهش تمام شماری بود.

مدل پیشنهادی در این پژوهش، یک رویکرد یادگیری با نظارت جهت پیش‌بینی مرگ بیماران نارسایی قلبی دارای شوک کاردیوژنیک می‌باشد. در این مطالعه مطابق با پیوست ۱ مجموعه داده به‌عنوان ورودی، وارد الگوریتم شد. در این مطالعه ابتدا داده‌ها به دو قسمت آموزشی (Train) و آزمایشی (Test) تقسیم گردید که در این جا به نسبت ۸۰ به ۲۰ است و سپس تابع موردنظر بر روی داده‌ها اجرا شد. سپس مدل RF جهت پیش‌بینی مرگ بیماران نارسایی قلبی دارای شوک کاردیوژنیک ایجاد گردید و پارامترهای مهم آن به‌منظور بالابردن دقت سیستم تنظیم شد. سپس در مرحله آخر، به‌منظور سنجش دقت الگوریتم موردنظر، مدل آموزش‌دیده با مجموعه‌داده‌های آزمایش، مطابق با دو معیار، مورد آزمایش و بررسی قرار گرفت. در واقع، در این پژوهش سعی شد مقدار بهینه پارامترهای الگوریتم جنگل تصادفی به‌منظور بالابردن دقت سیستم جهت پیش‌بینی مرگ بیماران نارسایی قلبی دارای شوک کاردیوژنیک تعیین شود، برای این منظور ابتدا الگوریتم جنگل تصادفی و پارامترهای مهم این الگوریتم معرفی می‌شود.

الگوریتم جنگل تصادفی (Random Forest)

جنگل تصادفی یک الگوریتم یادگیری ماشین تحت نظارت است. الگوریتم‌های طبقه‌بندی و پیش‌بینی باهدف افزایش دقت طبقه‌بندی با یکدیگر ترکیب می‌شوند و به عنوان یک مدل طبقه‌بندی ترکیبی یا گروهی استفاده می‌شوند که الگوریتم جنگل تصادفی یکی از این الگوریتم‌های گروهی می‌باشد. جنگل تصادفی چندین درخت تصمیم ساخته و آن‌ها را با یکدیگر ادغام می‌کند تا پیش‌بینی‌های صحیح‌تر و پایدارتری حاصل شوند. در واقع، جنگل تصادفی، چند درخت تصمیم ایجاد می‌کند و زیرمجموعه‌هایی از داده‌ها که در اختیار این درخت‌ها قرار می‌گیرند تصادفی هستند. مدل چندین بار روی نمونه‌های تصادفی از مجموعه‌داده آموزش می‌بیند تا به عملکرد خوبی در پیش‌بینی

[Downloaded from jhbmi.ir on 2026-11-23]
[DOI: 10.34172/jhbmi.2024.15]

دست پیدا کند. در این روش یادگیری گروهی، خروجی همه درخت‌های تصمیم موجود در جنگل تصادفی، با هم ترکیب می‌شوند تا پیش‌بینی نهایی را تولید کنند. پیش‌بینی نهایی الگوریتم جنگل تصادفی یا با نمونه‌گیری نتایج هرکدام از درخت‌ها و یا با انتخاب یکی از پیش‌بینی‌ها که بیشتر از همه‌بین درخت‌ها تکرار شده است، به دست می‌آید. در این روش، در مرحله آموزش، تعداد زیادی درخت تصمیم‌گیری برای مجموعه‌های داده مختلف ساخته می‌شود. در مرحله آزمون، هر درخت از این مجموعه درخت‌ها به هر داده یک برچسب کلاس اختصاص می‌دهد. پس از پیش‌بینی برچسب کلاس توسط هر یک از درخت‌ها، تصمیم نهایی برای هر داده با رأی‌گیری اکثریت اعضای مجموعه درخت‌ها اتخاذ می‌شود. برچسب کلاسی که بیشترین تعداد رأی را دریافت کند، به‌عنوان برچسب صحیح برای داده آزمایش در نظر گرفته می‌شود تا دقت پیش‌بینی آن افزایش یابد. این روش برای هر داده در مجموعهٔ داده انجام می‌شود[۱۶].

پارامترهای مهم جنگل تصادفی

الگوریتم جنگل تصادفی دارای چندین پارامتر است که به طور قابل‌توجهی بر عملکرد آن تأثیر می‌گذارد. تنظیم فراپارامتر، فرآیندی برای جستجوی بهترین پارامترها برای ایجاد یک مدل ایده‌آل است. تنظیم هایپرپارامتر مدل بسیار مهم است؛ زیرا مستقیماً بر رفتار مدل آموزشی ما تأثیر می‌گذارد که تأثیر قابل توجهی بر مجموعه داده‌های آزمایشی دارد. تنظیم بهینه این پارامترها به ما این امکان را می‌دهد که الگوریتم را برای پیش بینی دقیق‌تر در میزان مرگ بیماران شوک کاردیوژنیک مورد استفاده قرار دهیم. دو پارامتر کلیدی که قصد پیدا کردن بهینه آن‌ها را داریم به قرار زیر می‌باشند:

تعداد درختان (ntree): این پارامتر تعداد درختان تصمیم را در جنگل تصادفی تعیین می‌کند. افزایش تعداد درختان به‌طورکلی منجر به عملکرد بهتر می‌شود؛ اما زمان محاسبه را نیز افزایش می‌دهد.

تعداد ویژگی‌ها/ متغیرها (mtry): تعداد ویژگی‌های در نظر گرفته شده را هنگام یافتن بهترین تقسیم در هر گره تعیین می‌کند. محدودکردن تعداد ویژگی‌ها می‌تواند به کاهش پیچیدگی مدل و بهبود تعمیم کمک کند. با انتخاب و تنظیم دقیق این پارامترها، می‌توان عملکرد الگوریتم جنگل تصادفی را برای این دسته از داده‌ها بهینه نمود.

کدهای برنامه‌نویسی استفاده شده در این مطالعه، از زبان برنامه‌نویسی R نسخه ۴.۲.۲ در محیط RStudio که به‌عنوان یک ابزار قدرتمند در فرآیندهای یادگیری ماشین است، می‌باشد. ماتریس اغتشاش و همه شکل‌ها با استفاده از نرم فرار R رسم گردید.

در این پژوهش، ۳۴ ویژگی جهت پیش‌بینی مرگ استفاده شد که تمام این ویژگی‌ها در پیوست ۱ بیان شده است. برخی از این ویژگی‌ها مانند سن، سابقه جراحی قلب باز، pH، لاکتات، دیابت و غیره مورد بررسی قرار گرفت و مرگ یک‌ماهه آن‌ها از طریق تماس تلفنی پیگیری شد.

معیارهای مختلفی نظیر دقت و کاپا برای ارزیابی عملکرد پیش‌بینی استفاده شد[۲۰-۱۷]. در ماتریس در هم‌ریختگی (جدول ۱) چهار عامل اصلی شامل مثبت واقعیTP ، مثبت کاذبFP ، منفی واقعی TN و منفی کاذبFN شناسایی شدند. در صورتی که مدل بتواند شناسایی مناسب سطح خطر بالای مرگ ۳۰ روزه را افزایش دهد، نشان دهنده دقت مدل می‌باشد.

TP: نشان‌دهنده تعداد رکوردهایی است که دسته واقعی آن‌ها مثبت بوده و الگوریتم دسته‌بندی نیز دسته آن‌ها را به‌درستی مثبت تشخیص داده است (پیش‌بینی مرگ است و بیمار فوت شده است).

TN: نشان‌دهنده تعداد رکوردهایی است که دسته واقعی آن‌ها منفی بوده و الگوریتم دسته‌بندی نیز دسته آن‌ها را به‌درستی منفی تشخیص داده است (پیش‌بینی بقا است و بیمار زنده است).

FP: نشان‌دهنده تعداد رکوردهایی است که دسته واقعی آن‌ها منفی بوده و الگوریتم دسته‌بندی نیز دسته آن‌ها را به‌اشتباه مثبت تشخیص داده است (پیش‌بینی مرگ است و بیمار زنده است).

FN: نشان‌دهنده تعداد رکوردهایی است که دسته واقعی آن‌ها مثبت بوده و الگوریتم دسته‌بندی نیز دسته آن‌ها را به‌درستی منفی تشخیص داده است (پیش‌بینی بقا است و بیمار فوت شده است).

[Downloaded from jhbmi.ir on 2026-11-23]
[DOI: 10.34172/jhbmi.2024.15]

مجله انفورماتیک سلامت و زیست پزشکی دوره یازدهم، شماره اول، سال ۱۴۰۳ ، ص ۹۵-۸۳
doi: 10.34172/jhbmi.2024.15
مرکز تحقیقات انفورماتیک پزشکی، دانشگاه علوم پزشکی کرمان

جدول ۱: ماتریس درهم ریختگی

برچسب شناخته شده		ماتریس درهم‌ریختگی	
منفی(N)	مثبت(P)		
False negative (FN)	True positive (TP)	مثبت(P)	برچسب پیش‌بینی‌شده
True negative (TN)	False positive (FP)	منفی(N)	

دقت (Accuracy) را می‌توان توسط فرمول زیر محاسبه نمود:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN(all\ data)}$$

آماره Kappa:

آماره Kappa یک پارامتر برجسته برای قضاوت‌کردن در مورد سازگاری مدل است. این آماره، نتیجه مدل پیشنهادی را با یک نتیجه تولید شده توسط روش کلاس‌بندی تصادفی، مقایسه می‌کند. مقدار آماره Kappa بین ۰ و ۱ است. مقدار نزدیک‌تر به ۱، اثر مورد انتظار مدل را نشان می‌دهد درحالی‌که صفر به معنای نامعتبر بودن مدل است. این معیار توسط روابط زیر محاسبه می‌گردد:

$$K = \frac{[P(A) - P(E)]}{[1 - P(E)]}$$

$$P(A) = \frac{(TP + TN)}{N}$$

$$P(E) = \frac{[(TP + FN) \times (TP + FP) \times (TN + FN)]}{N \times N}$$

نتایج

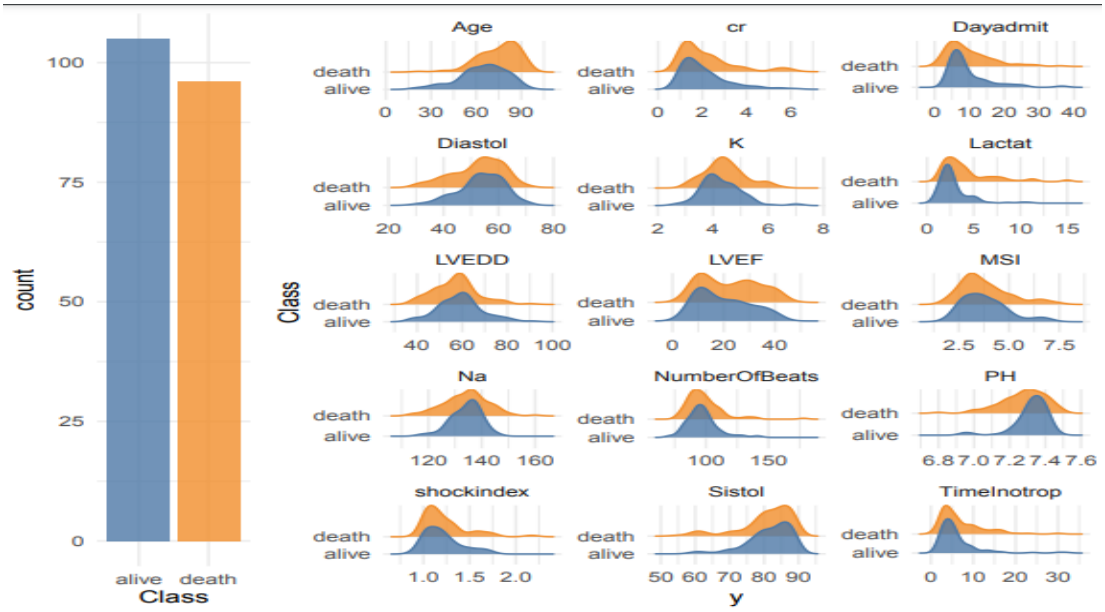
یافته‌ها نشان می‌دهد در مطالعه پیش‌بینی مرگ بیماران نارسایی قلبی با شوک کاردیوژنیک در مجموعه‌داده موردنظر، از ۲۰۱ بیمار، میانگین سن بیماران ۱۵/۷۱±۶۹/۴۴ سال و ۵۷/۳ درصد از بیماران مرد بودند که از این میزان ۴۵/۲ درصد فوت و ۵۴/۸ درصد زنده بودند (جدول ۲).

جدول ۲: درصد فراوانی بیماران فوت شده و زنده مانده

جنسیت	زنده	فوت شده	جمع کل
مرد	٪۵۴/۸	٪۴۵/۲	٪۵۷/۳
زن	٪۴۸/۸۳	٪۵۱/۱۷	٪۴۲/۷

[Downloaded from jhbmi.ir on 2026-11-23]
[DOI: 10.34172/jhbmi.2024.15]

شکل ۱ وابستگی جزئی متغیرهای عددی با مرگ بیماران نارسایی قلبی با شوک کاردیوژنیک را نشان می‌دهد که با افزایش سن (بالاتر از ۵۷ سال) خطر مرگ ۳۰ روزه در بیماران نارسایی قلبی با شوک کاردیوژنیک افزایش می‌یابد. خطر مرگ ۳۰ روزه فشارخون سیستولی بالاتر از ۷۵ بیشتر از فشارخون سیستولی کمتر از ۶۵ می‌باشد. همچنین خطر مرگ ۳۰ روزه برای فشارخون دیاستول بالای ۵۰ بیشتر از فشارخون دیاستول کمتر از ۴۰ می‌باشد. با کاهش میزان pH کمتر از ۳/۷ خطر مرگ ۳۰ روزه در بیماران نارسایی قلبی با تابلوی شوک کاردیوژنیک نیز افزایش می‌یابد. افزایش Lactate بالاتر از ۲ ریسک مرگ ۳۰ روزه در بیماران نیز افزایش می‌یابد. با استفاده از این شکل می‌توان اطلاعاتی حیاتی در مورد خطر مرگ ۳۰ روزه در اختیار کارکنان و کادر درمان قرار داد.



شکل ۱: وابستگی جزئی متغیرهای عددی با مرگ بیماران نارسایی قلبی با شوک کاردیوژنیک

در پیش‌بینی مرگ نارسایی قلبی از تکنیک جنگل تصادفی استفاده شد، این روش یادگیری تجمعی معمولاً برای وظایف تشخیص الگو به کار می‌رود و شامل چندین درخت تصمیم‌گیری است که آن را در مواجهه با داده‌های پیچیده بسیار کارآمد می‌سازد. جنگل تصادفی همچنین به نام bagging یا bootstrap aggregation نیز شناخته می‌شود. این روش، خروجی‌های درختان مجزا را ترکیب کرده و پیش‌بینی نهایی را تولید می‌کند. برخلاف یک درخت تصمیم‌گیری تکی که ممکن است بایاس پایین، ولی واریانس بالا داشته باشد، جنگل تصادفی با تجمیع پیش‌بینی‌های چندین درخت، واریانس را کاهش داده و عملکرد پیش‌بینی کلی را بهبود می‌بخشد [۲۱].

روش جنگل تصادفی در نرم‌افزار R دارای تابعی با دو پارامتر mtry و ntree که به ترتیب تعداد متغیرها و تعداد درخت می‌باشد که قابل‌تغییر بوده و توسط کاربر تعیین می‌شود. طبقه‌بندی جنگل تصادفی ابتدا بر روی داده آموزشی و سپس روی داده‌های آزمایشی اجرا می‌گردد. در این مطالعه ابتدا داده‌ها به دو قسمت آموزشی (train) و آزمایشی (test) تقسیم گردیده که در این جا به نسبت ۸۰ به ۲۰ می‌باشد و سپس تابع موردنظر بر روی داده‌ها اجرا شده است.

در این مطالعه، به تنظیم دو پارامتر، یعنی پارامترهای mtry و ntree که بر مدل جنگل تصادفی تأثیر می‌گذارند، پرداخته شد. بسیاری از پارامترهای دیگر نیز وجود دارد، اما این دو پارامتر احتمالاً بیشترین تأثیر را بر دقت نهایی دارند. برای به‌دست‌آوردن بهترین هایپر پارامترهای جنگل تصادفی، حالت‌های مختلف پارامترها را در کنار هم قرار داده تا بتوان الگوریتم جنگل تصادفی بهینه شود. جدول ۳ بیانگر مقادیر بهینه حاصل شده از اجرای این برنامه برای پارامترهای mtry و ntree است.

[Downloaded from jhbmi.ir on 2026-11-23]
[DOI: 10.34172/jhbmi.2024.15]

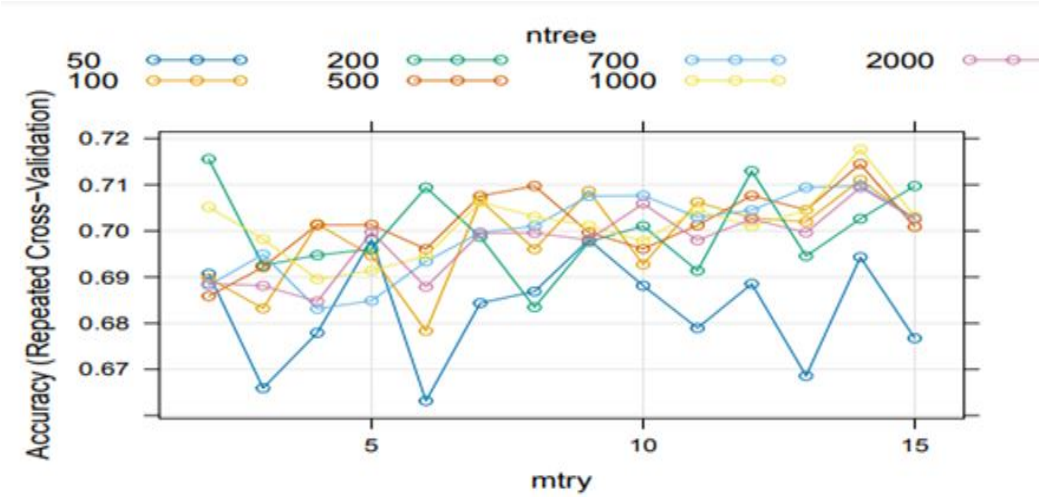
مجله انفورماتیک سلامت و زیست پزشکی دوره یازدهم، شماره اول، سال ۱۴۰۳ ، ص ۹۵-۸۳
doi: 10.34172/jhbmi.2024.15
مرکز تحقیقات انفورماتیک پزشکی، دانشگاه علوم پزشکی کرمان

جدول ۳: حالت‌های متنوع برای تعیین بهترین دقت در پارامترهای جنگل تصادفی

Kappa	Accuracy	Ntree	Mtry	Kappa	Accuracy	Ntree	Mtry
۰/۴۱۲	۰/۷۰۸	۷۰۰	۹	۰/۳۲۸	۰/۶۶۶	۵۰	۳
۰/۳۹۹	۰/۷۰۱	۱۰۰۰	۹	۰/۳۶۲	۰/۶۳۸	۱۰۰	۳
۰/۳۹۳	۰/۶۹۸	۲۰۰۰	۹	۰/۳۷۹	۰/۶۹۲	۲۰۰	۳
۰/۳۷۱	۰/۶۸۸	۵۰	۱۰	۰/۳۷۸	۰/۶۹۲	۵۰۰	۳
۰/۳۸۴	۰/۶۹۳	۱۰۰	۱۰	۰/۳۸۳	۰/۶۹۵	۷۰۰	۳
۰/۳۹۸	۰/۷۰۱	۲۰۰	۱۰	۰/۳۹۰	۰/۶۹۸	۱۰۰۰	۳
۰/۳۸۸	۰/۶۹۶	۵۰۰	۱۰	۰/۳۷۰	۰/۶۸۸	۲۰۰۰	۳
۰/۴۱۳	۰/۷۰۸	۷۰۰	۱۰	۰/۳۴۹	۰/۶۷۸	۵۰	۴
۰/۳۹۲	۰/۶۹۸	۱۰۰۰	۱۰	۰/۳۹۸	۰/۷۰۱	۱۰۰	۴
۰/۴۰۹	۰/۷۰۶	۲۰۰۰	۱۰	۰/۳۸۴	۰/۶۹۵	۲۰۰	۴
۰/۳۵۴	۰/۶۷۹	۵۰	۱۱	۰/۳۹۷	۰/۷۰۱	۵۰۰	۴
۰/۴۰۹	۰/۷۰۶	۱۰۰	۱۱	۰/۳۶۱	۰/۶۸۳	۷۰۰	۴
۰/۳۷۹	۰/۶۹۱	۲۰۰	۱۱	۰/۳۷۴	۰/۶۸۹	۱۰۰۰	۴
۰/۳۹۹	۰/۷۰۱	۵۰۰	۱۱	۰/۳۶۴	۰/۶۸۵	۲۰۰۰	۴
۰/۴۰۳	۰/۷۰۳	۷۰۰	۱۱	۰/۳۹۲	۰/۶۹۸	۵۰	۵
۰/۴۰۶	۰/۷۰۴	۱۰۰۰	۱۱	۰/۳۸۵	۰/۶۹۵	۱۰۰	۵
۰/۳۹۳	۰/۶۹۸	۲۰۰۰	۱۱	۰/۳۸۷	۰/۶۹۶	۲۰۰	۵
۰/۳۷۱	۰/۶۸۸	۵۰	۱۲	۰/۳۹۹	۰/۷۰۱	۵۰۰	۵
۰/۴۰۳	۰/۷۰۳	۱۰۰	۱۲	۰/۳۶۵	۰/۶۸۵	۷۰۰	۵
۰/۴۲۳	۰/۷۱۳	۲۰۰	۱۲	۰/۳۷۸	۰/۶۹۱	۱۰۰۰	۵
۰/۴۱۲	۰/۷۰۸	۵۰۰	۱۲	۰/۳۹۵	۰/۶۹۹	۲۰۰۰	۵
۰/۴۰۵	۰/۷۰۴	۷۰۰	۱۲	۰/۳۲۳	۰/۶۳۳	۵۰	۶
۰/۳۹۹	۰/۷۰۱	۱۰۰۰	۱۲	۰/۳۵۴	۰/۶۷۸	۱۰۰	۶
۰/۴۰۲	۰/۷۰۲	۲۰۰۰	۱۲	۰/۴۱۴	۰/۷۰۹	۲۰۰	۶
۰/۳۳	۰/۶۶۸	۵۰	۱۳	۰/۳۸۸	۰/۶۹۶	۵۰۰	۶
۰/۴۰۱	۰/۷۰۲	۱۰۰	۱۳	۰/۳۸۳	۰/۶۹۳	۷۰۰	۶
۰/۳۸۵	۰/۶۹۴	۲۰۰	۱۳	۰/۳۸۶	۰/۶۹۴	۱۰۰۰	۶
۰/۴۰۶	۰/۷۰۵	۵۰۰	۱۳	۰/۳۷۱	۰/۶۸۸	۲۰۰۰	۶
۰/۴۱۶	۰/۷۰۹	۷۰۰	۱۳	۰/۳۶۵	۰/۶۸۴	۵۰	۷
۰/۴۰۶	۰/۷۰۴	۱۰۰۰	۱۳	۰/۴۰۸	۰/۷۰۶	۱۰۰	۷
۰/۳۹۶	۰/۶۹۹	۲۰۰۰	۱۳	۰/۳۹۵	۰/۶۹۹	۲۰۰	۷
۰/۳۸۶	۰/۶۹۴	۵۰	۱۴	۰/۴۱۱	۰/۷۰۷	۵۰۰	۷
۰/۴۲۰	۰/۷۱۱	۱۰۰	۱۴	۰/۳۹۶	۰/۶۹۹	۷۰۰	۷
۰/۴۰۲	۰/۷۰۳	۲۰۰	۱۴	۰/۴۰۸	۰/۷۰۶	۱۰۰۰	۷
۰/۴۲۶	۰/۷۱۴	۵۰۰	۱۴	۰/۳۹۶	۰/۶۹۹	۲۰۰۰	۷
۰/۴۱۷	۰/۷۱۰	۷۰۰	۱۴	۰/۳۷۱	۰/۶۸۷	۵۰	۸
۰/۴۳۳	۰/۷۱۸	۱۰۰۰	۱۴	۰/۳۸۸	۰/۶۹۶	۱۰۰	۸
۰/۴۱۶	۰/۷۱۰	۲۰۰۰	۱۴	۰/۳۶۴	۰/۶۸۳	۲۰۰	۸
۰/۳۵۰	۰/۶۷۷	۵۰	۱۵	۰/۴۱۶	۰/۷۰۹	۵۰۰	۸
۰/۴۰۲	۰/۷۰۲	۱۰۰	۱۵	۰/۳۹۸	۰/۷۰۱	۷۰۰	۸
۰/۴۱۶	۰/۷۱۰	۲۰۰	۱۵	۰/۴۰۲	۰/۷۰۳	۱۰۰۰	۸
۰/۳۹۹	۰/۷۰۱	۵۰۰	۱۵	۰/۳۹۶	۰/۶۹۹	۲۰۰۰	۸
۰/۴۰۲	۰/۷۰۲	۷۰۰	۱۵	۰/۳۹۳	۰/۶۹۸	۵۰	۹
۰/۴۰۳	۰/۷۰۳	۱۰۰۰	۱۵	۰/۴۱۳	۰/۷۰۸	۱۰۰	۹
۰/۴۰۲	۰/۷۰۳	۲۰۰۰	۱۵	۰/۳۹۳	۰/۶۹۸	۲۰۰	۹
Best Accuracy =۰/۷۱۸				۰/۳۹۶	۰/۶۹۹	۵۰۰	۹

[Downloaded from jhbmi.ir on 2026-11-23]
[DOI: 10.34172/jhbmi.2024.15]

همان‌طور که در شکل ۲ نشان داده شد برای درک بهتری از بهترین پارامترها و بالاترین دقت، تمام ترکیب‌های ممکن حالت‌های مختلف به تصویر کشیده شد. برای ایجاد حالت‌های ممکن، مقدار mtry از ۳ شروع شد و یکی‌یکی به مقدار آن افزوده شد و برای ntree مقادیر ۵۰ و ۱۰۰ و ۲۰۰ و ۵۰۰ و ۷۰۰ و ۱۰۰۰ و ۲۰۰۰ انتخاب گردید. بر اساس شکل ۲ در ntreeها و mtryهای مختلف مقدار دقت متفاوت است.



شکل ۲: مقایسه دقت الگوریتم جنگل تصادفی در حالت‌های مختلف

بحث و نتیجه‌گیری

نارسایی قلبی همراه با شوک کاردیوژنیک نیازمند تشخیص و درمان فوری است؛ لذا یک وضعیت اورژانسی پزشکی محسوب می‌شود[۴]. پیش‌بینی دقیق مرگ و میر بیماران مبتلا به نارسایی قلبی با شوک کاردیوژنیک، نقشی حیاتی در ارجاع و اقدامات درمانی مناسب ایفا می‌کند. این امر به پزشکان کمک می‌کند تا با درک بهتر ریسک مرگ و میر بیمار، بهترین تصمیمات را برای مراقبت از او اتخاذ کنند. لزوم استفاده از الگوریتم‌های یادگیری ماشین که ابزاری قدرتمند برای پیش‌بینی در زمینه‌های مختلف، از جمله پزشکی، هستند و می‌توانند از حجم عظیمی از داده‌ها برای یادگیری الگوها و پیش‌بینی نتایج آینده استفاده کنند، در این حیطه ضروری به نظر می‌رسد [۵]: لذا در این پژوهش رویکردی با استفاده از الگوریتم یادگیری ماشین جنگل تصادفی و با تنظیم هایپرپارامترهای آن در راستای بهبود دقت پیش‌بینی مرگ و میر بیماران نارسایی قلبی با شوک کاردیوژنیک بر روی داده‌های ۲۰۱ بیمار منتخب در بخش اورژانس قلب بیمارستان روحانی صورت گرفت (شکل ۱). در واقع شکل ۱ اطلاعاتی حیاتی و مهمی همچون ارتباط بین افزایش سن بیمار و افزایش زمان رساندن بیمار به بیمارستان و افزایش لاکتات با افزایش خطر مرگ را در اختیار پزشکان و کادر درمان برای نتیجه گیری موردی از ویژگی‌ها در درمان بیماری، قرار می‌دهد.

تنظیم هایپرپارامترها، به‌ویژه تعداد درختان، برای عملکرد بهینه جنگل تصادفی حیاتی است. مطابق جدول ۴ اگرچه هر دو مطالعه انجام شده توسط Breiman و Oshiro و همکاران بر اهمیت تنظیم هایپرپارامترها برای دستیابی به نتایج بهینه تأکید دارند؛ اما Breiman بر بازدهی کاهش‌یافته با افزایش تعداد درختان تأکید دارد که Oshiro و همکاران این موضوع را با استفاده از AUC کمّی‌سازی می‌کنند و آستانه‌ای تحت‌تأثیر منابع محاسباتی پیشنهاد می‌کنند درحالی‌که در مطالعه حاضر با افزایش تعداد درختان بهبود دقت پیش‌بینی میسر شده است [۱۲،۱۳].

[Downloaded from jhbmi.ir on 2026-11-23]
[DOI: 10.34172/jhbmi.2024.15]

اگرچه مطالعات زیادی در مورد پیش‌بینی مرگ در بیماران نارسایی قلبی با استفاده از یادگیری ماشین انجام پذیرفته است تا دقت پیش‌بینی را بالا ببرد [۷،۱۴،۱۵]. پژوهش Awan مرگ ۳۰ روزه را هم در نظر گرفته بود [۶]. با این حال هنوز جای پیشرفت زیادی در زمینه پیش‌بینی بیماری قلبی –عروقی با استفاده از الگوریتم‌های یادگیری ماشین وجود دارد. همچنین دقت مدل‌های فعلی در برخی موارد می‌تواند به طور قابل توجهی بهبود یابد؛ لذا نیاز به بررسی و مقایسه عملکرد الگوریتم‌های مختلف یادگیری ماشین در وظایف پیش‌بینی بیماری قلبی– عروقی وجود دارد. این مطالعه به طور خاص بر پیش‌بینی مرگ و میر ۳۰ روزه بیماران نارسایی قلبی با شوک کاردیوژنیک تمرکز دارد، در حالی که بسیاری از مطالعات موجود به پیش‌بینی‌های عمومی‌تری پرداخته‌اند؛ استفاده از بهینه‌سازی هایپرپارامترهای مدل جنگل تصادفی در این مطالعه به شکل دقیق‌تری انجام شده است که بهبود دقت پیش‌بینی را به همراه داشته است. تحلیل دقیق‌تری از تأثیر عوامل مختلف بر مرگ و میر بیماران ارائه شده است که در پژوهش‌های قبلی به این سطح از جزئیات پرداخته نشده است. در نهایت، این مطالعه با ارائه مدل بهینه شده و دقیق‌تر می‌تواند به عنوان راهنمایی برای تصمیم‌گیری‌های بالینی و بهبود مدیریت بیماران مورد استفاده قرار گیرد، بنابراین هدف در مطالعه پیش‌رو، بررسی و ارزیابی عملکرد الگوریتم جنگل تصادفی در پیش‌بینی مرگ و میر بیماران مبتلا به نارسایی قلبی با شوک کاردیوژنیک است.

این مطالعه نشان داد که با بهینه‌سازی الگوریتم جنگل تصادفی، دقت پیش‌بینی مرگ و میر ۳۰ روزه در بیماران نارسایی قلبی با شوک کاردیوژنیک به‌طور قابل‌توجهی افزایش می‌یابد که با مطالعه Awan که بر پیش‌بینی مرگ ۳۰ روزه تمرکز دارد، همخوانی دارد [۶]. همچنین در مطالعات انجام شده توسط Atiyah و Kamal و Moubayed و همکاران نیز استفاده از تنظیم هایپر پارامترهای الگوریتم‌های یادگیری ماشین از جمله جنگل تصادفی، به منظور افزایش دقت این روش‌ها و بهبود عملکرد، انجام شده‌اند. در مطالعه König و همکاران که به بررسی مدل‌های مختلف یادگیری ماشین پرداختند، روندهای AUC مشابهی برای تنظیم هایپرپارامتر مدل‌های مختلف که مشمول RF نیز می‌شدند مشاهده شد که نشان‌دهنده بهبود عملکرد مدل‌ها در دقت پیش بینی به وسیله تنظیم فراپارامترها می‌شود [۷]. علاوه بر این مطالعات زیادی با روش‌های متفاوت به منظورتنظیم هایپرپارامترها صورت گرفت که نشان داد مدل‌های بهینه شده نسبت به مدل‌های سنتی بالاترین ارزش پیش‌بینی را در بین رویکردهای یادگیری ماشین می‌باشند [۱۳-۸].

از نقاط قوت این مطالعه می‌توان افزایش دقت پیش‌بینی به بالاترین مقدار با استفاده از بهینه‌سازی فراپارامترها اشاره کرد که نشان‌دهنده کارایی بالای سیستم در تشخیص درست بیماران است. در این پژوهش بر اساس تغییر در پارامترهای الگوریتم جنگل با تعیین پارامترهای mtry و ntree که احتمالاً بالاترین تأثیر را در دقت نهایی دارند به ترتیب برابر ۱۴ و ۱۰۰۰، دقتِ پیش‌بینی، از مقدار ۰/۶۶ به بالاترین مقدار دقت ۰/۷۱ رسید. از جمله نقاط منفی نیز می‌توان به محدودیت‌هایی در دسترسی به داده‌های بیماران از بیمارستان‌های مختلف وجود داشت که می‌تواند بر تعمیم‌پذیری نتایج تأثیرگذار باشد. این مسئله می‌تواند دقت پیش‌بینی‌ها را در موارد مختلف کاهش دهد. نتایج به‌دست‌آمده نشان داد که بهینه‌کردن پارامترهای الگوریتم جنگل تصادفی دقت سیستم را بهبود می‌بخشد و باعث پیش‌بینی دقیق‌تر مرگ در بیماران نارسایی قلبی دارای شوک کاردیوژنیک می‌شود. این بهینه‌سازی به کادر درمان و پزشکان کمک می‌کند تا با تعیین بهترین پارامترهای الگوریتم میزان خطا را کاهش داده و در این زمینه با بهینه سازی نظام ارجاع و یا اقدام درمانی بعدی، باعث نجات جان بیماران بیشتری شوند.

باتوجه‌به این که اقدام درمانی سریع برای این دسته از بیماران می‌تواند به معنای بقای بیمار باشد، می‌توان به‌کارگیری سایر الگوریتم‌های یادگیری ماشین، استفاده از داده‌های بزرگ‌تر و متنوع‌تر برای بهبود دقت مدل و الگوریتم‌های ترکیبی برای این داده‌ها جهت بهبود دقت پیش‌بینی مرگ و ابزاری قوی برای متخصصان و کادر درمان برای این دسته از بیماران، مورد ارزیابی قرارداد. همچنین می‌توان پارامترهای دیگر الگوریتم جنگل تصادفی را برای به‌دست‌آوردن دقت بالاتر در پیش‌بینی مرگ‌ومیر مورد ارزیابی قرارداد. از بهینه‌سازی الگوریتم جنگل تصادفی می‌توان در بیماری‌های دیگر که نیاز به‌دقت بالایی برای پیش‌بینی بقایا مرگ دارند نیز، استفاده کرد.

[Downloaded from jhbmi.ir on 2026-11-23]
[DOI: 10.34172/jhbmi.2024.15]

جدول ۴: مقایسه مطالعات پیشین با مطالعه حاضر

مطالعات انجام شده	روش‌ها	نتایج AUC	مقایسه با مطالعه حاضر
Breiman (۲۰۰۱) [۱۲]	Random Forest	Random Forest: ۰/۸۷	مطابق با Breiman نشان‌دهنده مقادیر AUC مشابه برای RFاست. (به اهمیت تنظیم پارامترها برای بهبود عملکرد مدل و بهبود دقت با تنظیم مناسب هایپرپارامترها در هر دو مطالعه اشاره شده است) بهبود دقت با تنظیم مناسب هایپرپارامترها
Oshiro و همکاران (۲۰۱۲) [۱۳]	Random Forest	Random Forest:۰/۸۸	نتایج مشابه با Oshiro و همکاران، تأییدکننده اثربخشی Random Forest است و اهمیت تنظیم هایپرپارامترها برای بهبود عملکرد مدل است
Awan (۲۰۱۹) [۶]	Gradient Boosting	Gradient Boosting: ۰/۸۵	نتایج قابل‌مقایسه با Awan، نشان‌دهنده مقادیر مشابه AUC برای Gradient Boosting است تمرکز مطالعه بر روی پیش‌بینی مرگ ۳۰ روزه در بیماران نارسایی قلبی با بررسی مدل‌های و عملکردهای مخالف بوده برخلاف مطالعه حاضر که بر RF تمرکز دارد
Kamal و Atiya (۲۰۲۱) [۹]	استفاده از روش‌های ANN و RANDOM FOREST برای پیش‌بینی	ANN: ۰/۹۲, SVM: ۰/۸۸, Random Forest: ۰/۹۰	مطالعه حاضر با Atiya و Kamal مطابقت دارد و نشان‌دهنده عملکرد مشابه ANN و Random Forest است در این مطالعه نیز اهمیت تنظیم هایپرپارامترها برای بهبود عملکرد مدل‌ها پرداخته؛ اما تمرکز اصلی بر روی یک مدل نداشته است.
Moubayed و همکاران (۲۰۲۰) [۸]	GBM, RF, NNET, XGBoost	GBM: ۰/۸۷۱, RF: ۰/۸۷۳, NNET: ۰/۸۶۴, XGBoost: ۰/۸۷۶	مطالعه حاضر نشان‌دهنده مقادیر AUC قابل‌مقایسه با Moubayed است که نشان‌دهنده اعتبار مدل‌های ML است.
König و همکاران (۲۰۲۳) [۷]	GLM, RF, GBM, NNET, XGBoost	GLM: ۰/۸۲۹, RF: ۰/۸۷۵, GBM:۰/۸۸۲, NNET:۰/۸۶۶, XGBoost: ۰/۸۸۲	نتایج مطالعه حاضر با König مطابقت دارد و روندهای مشابهی در مقادیر AUC نشان می‌دهد.

تشکر و قدردانی

بدین‌وسیله از تمامی افرادی که در انجام این پژوهش یاری کردند، قدردانی می‌گردد.

تعارض منافع

نویسندگان اعلام می‌دارند که هیچ‌گونه تعارض منافعی در مطالعه حاضر وجود ندارد.

[Downloaded from jhbmi.ir on 2026-11-23]

[DOI: 10.34172/jhbmi.2024.15]

حمایت مالی

پژوهش حاضر تحت حمایت مالی هیچ سازمانی نبوده است.

کد اخلاق

این مطالعه پس از تصویب در کمیته اخلاق معاونت پژوهشی دانشگاه علوم و فنون بابل با کد ۹۸/۳۴۲/۱ بر روی بیماران مبتلا به نارسایی قلبی با شوک کاردیوژنیک و بالای ۱۸ سال در بیمارستان روحانی شهر بابل انجام شد. همچنین، این پژوهش مستخرج از رساله مصوب معاونت پژوهشی دانشگاه علوم و فنون بابل مورخ ۱۳۹۸/۱۱/۳۰ می‌باشد.

سهم مشارکت نویسندگان

نقش فعال نویسنده اول و مسئول مقاله در طراحی مطالعه پژوهش و نوشتن مقاله، نقش نویسنده دوم در تایید نسخه نهایی مقاله و تجزیه و تحلیل‌های مهندسی آن، نقش نویسنده سوم در گردآوری داده های پزشکی مقاله و تجزیه و تحلیل‌های پزشکی و نقش نویسنده چهارم در تجزیه و تحلیل و تفسیر داده‌های مقاله و درستی الگوریتم‌های آن و بازبینی و اصلاح موشکافانه مقاله می‌باشد.

References

[1]. Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. PLoS Med 2006;3(11):e442. https://doi.org/10.1371/journal.pmed.0030442

[2]. Said IU, Adam AH, Garko AB. Association rule mining on medical data to predict heart disease. International Journal of Science Technology and Management 2015;4(8):26-35.

[3]. Ponikowski P, Voors AA, Anker SD, Bueno H, Cleland JGF, Coats AJS, et al. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: The Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC)Developed with the special contribution of the Heart Failure Association (HFA) of the ESC. Eur Heart J 2016;37(27):2129-200.

[4]. Jentzer JC, Chonde MD, Dezfulian C. Myocardial dysfunction and shock after cardiac arrest. BioMed Research International 2015;2015(1):314796. https://doi.org/10.1155/2015/314796

[5]. Hosseinpoor M. Presenting an Intelligent Algorithm to Predict Heart Attack. Journal of Health and Biomedical Informatics 2024;10(4):346-56. [In Persian] doi:10.34172/jhbmi.2024.03

[6]. Awan SE, Bennamoun M, Sohel F, Sanfilippo FM, Chow BJ, Dwivedi G. Feature selection and transformation by machine learning reduce variable numbers and improve prediction for heart failure readmission or death. PLoS One 2019;14(6):e0218760. doi: 10.1371/journal.pone.0218760

[7]. König S, Pellissier V, Hohenstein S, Bernal A, Ueberham L, Meier-Hellmann A, et al. Machine learning algorithms for claims data-based prediction of in-hospital mortality in patients with heart failure. ESC Heart Fail 2021;8(4):3026-36. doi: 10.1002/ehf2.13398

[8]. Moubayed A, Injadat M, Shami A. Optimized random forest model for botnet detection based on DNS queries. 32nd International Conference on Microelectronics (ICM); 2020 Dec 14-17; Aqaba, Jordan: IEEE; 2020. p. 1-4. doi: 10.1109/ICM50269.2020.9331819

[9]. Atiyah MA, Kamal CA. Predicting Mortality COVID-19 Model and Characterize the High Risk of Dying Based on Machine Learning Models. The Scientific Journal of Cihan University–Sulaimaniya 2023;7(1):61-75. doi: https://doi.org/10.25098/7.1.4

[10]. Bischl B, Binder M, Lang M, Pielok T, Richter J, Coors S, et al. Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges. Wires Data Mining and Knowledge Discovery 2023;13(2): e1484. https://doi.org/10.1002/widm.1484

[11]. Iwase S, Nakada TA, Shimada T, Oami T, Shimazui T, Takahashi N, Yamabe J, Yamao Y, Kawakami E. Prediction algorithm for ICU mortality and length of stay using machine learning. Scientific Reports 2022;12(1):12912.

[12]. Breiman L. Random Forests. Machine Learning 2001;45(1):5-32.

[13]. Oshiro TM, Perez PS, Baranauskas JA. How many trees in a random forest?. Machine Learning and Data Mining in Pattern Recognition. 8th International Conference; 2012 Jul 13-20; Berlin: Berlin Heidelberg: Springer; 2012. p. 154-68.

[14]. Krittanawong C, Hassan Virk H, Hahn J, Al-Azzam F, Greason K, Isath A, et al. Leveraging machine learning models to predict in-hospital mortality after Mitraclip. Journal of the American College of Cardiology 2021;77:940.

[Downloaded from jhbmi.ir on 2026-11-23]

[DOI: 10.34172/jhbmi.2024.15]

[15]. Li X, Sousa-Casasnovas I, Devesa C, Juárez M, Fernández-Avilés F, Martínez-Sellés M. Predictors of in-hospital mortality among cardiogenic shock patients. Prognostic and therapeutic implications. Int J Cardiol 2016:224:114-8. doi: 10.1016/j.ijcard.2016.09.019
[16]. Probst P, Wright MN, Boulesteix AL. Hyperparameters and tuning strategies for random forest. Wires Data Mining and Knowledge Discovery 2019;9(3):e1301. https://doi.org/10.1002/widm.1301
[17]. Fawcett T. Introduction to ROC analysis. Pattern Recognition Letters 2006;27(8):861-74. https://doi.org/10.1016/j.patrec.2005.10.010
[18]. Chawla N. Data Mining for Imbalanced Datasets: An Overview. In: Fayyad U, Piatetsky-Shapiro G, Smyth P, editors. Data Mining and Knowledge Discovery Handbook. 2nd ed. New York: Springer; 2010. p. 853-67.
[19]. Witten IH, Frank E, Hall MA. Data Mining: Practical Machine Learning Tools and Techniques. 3rd ed. Morgan Kaufmann; 2011. https://doi.org/10.1016/C2009-0-19715-5
[20]. Kubat M, Matwin S. Addressing the curse of imbalanced training sets: one-sided selection. 1997. Available from: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=ebc3914181d76c817f0e35f788b7c4c0f80abb07
[21]. Talukder MS, Sulaiman RB, Angon MB. Unleashing the Power of Extra-Tree Feature Selection and Random Forest Classifier for Improved Survival Prediction in Heart Failure Patients. arXiv preprint arXiv:2308.05765. 2023 https://doi.org/10.48550/arXiv.2308.05765

[Downloaded from jhbmi.ir on 2026-11-23]

[DOI: 10.34172/jhbmi.2024.15]

پیوست ۱: مشخصات مجموعه داده مورداستفاده در مدل پیشنهادی

ردیف	نام متغیر	نوع متغیر	مقیاس متغیر	نقش متغیر	واحد اندازه‌گیری	توضیحات
۱	مورتالیتی	کیفی	اسمی	وابسته	بلی/خیر	مرگ در بیمارستان + مرگ ۳۰ روزه
۲	فشارخون سیستول	کمی پیوسته	نسبتی	مستقل	mmHg	Min=55,max=89,mean=81
۳	فشارخون دیاستول	کمی پیوسته	نسبتی	مستقل	mmHg	Min=30,max=71,mean=53.26
۴	تعداد ضربان نبض	کمی گسسته	نسبتی	مستقل		Min=71,max=178,median=97
۵	ضربان قلب	کیفی	اسمی	مستقل	سینوسی/غیر سینوسی	1(N=121),2(N=80)
۶	اندام‌های انتهایی سرد	کیفی	اسمی	مستقل	دارد/ندارد	دارد(تعداد=۱۹۳)–ندارد (تعداد=۸)
۷	ایگوری	کیفی	اسمی	مستقل	دارد/ندارد	دارد (تعداد=۱۱۲)–ندارد (تعداد=۸۹)
۸	کنفیوزیون	کیفی	اسمی	مستقل	دارد/ندارد	دارد (تعداد=۱۲۸)–ندارد (تعداد=۷۳)
۹	LVEF	کمی پیوسته	نسبتی	مستقل	٪	Min=5,max=45,mean=21
۱۰	LVEDD	کمی پیوسته	نسبتی	مستقل	Mm	Min=38,max=91,mean=58
۱۱	Na	کمی گسسته	نسبتی	مستقل	Mmol/L	Min=113,max=160,mean=134
۱۲	K	کمی گسسته	نسبتی	مستقل	Mmol/L	Min=2.6,max=7.1,mean=4.4
۱۳	لاکتات خون شرانی	کمی گسسته	نسبتی	مستقل	Mmol/L	Min=0.8,max=15,mean=3.5
۱۴	PH خون شریانی	کمی پیوسته	نسبتی	مستقل		Min=6.8,max=7.5,mean=7.3
۱۵	Cr	کمی پیوسته	نسبتی	مستقل	Mmol/L	Min=0.9,max=6.2,mean=2.13
۱۶	زمان طول کشیده از شروع علائم بالینی تا رسیدن به بیمارستان	کمی گسسته	نسبتی	مستقل	دقیقه	بستری (۶۴)–۰ تا ۳(۵۸)–۳ تا ۶(۳۲)–۶ تا ۱۲(۲۱) — ۱۲ تا ۲۴(۱۷) - بیشتر از ۲۴(۶)
۱۷	مدت‌زمان حضور در بیمارستان	کمی گسسته	نسبتی	مستقل	روز	کمترین روز بستری (۱ روز)–بیشترین روز بستری (۳۷ روز)
۱۸	نوع اینوتروپ دریافتی	کیفی	اسمی	مستقل	سه دسته	دسته اول (۳۳)– دسته دوم (۱۴۶) – دسته سوم (۲۳)
۱۹	مدت.زمان.دریافت.اینوتروپ	کمی گسسته	نسبتی	مستقل	روز	Min=1,max=9,mean=7
۲۰	MSI	کمی پیوسته	نسبتی	مستقل		Min=1.73,max=7.45,mean=3.80
۲۱	Shock Index	کمی پیوسته	نسبتی	مستقل		Min=0.86,max=2.20,mean=1.22
۲۲	سن	کمی گسسته	نسبتی	سایر	سال	Min=19,max=96,mean=69
۲۳	جنس	کیفی	نسبتی	سایر	زن/مرد	مرد (۱۱۵) – زن (۸۶)
۲۴	سابقه بیماری عروق کرونر	کیفی	اسمی	مستقل	دارد/ندارد	دارد (۱۱۳) – ندارد (۸۸)
۲۵	سابقه انفارکتوس میوکارد	کیفی	اسمی	مستقل	دارد/ندارد	دارد (۷۸) – ندارد (۱۲۳)
۲۶	PCI	کیفی	اسمی	مستقل	دارد/ندارد	دارد (۴۶) – ندارد (۱۵۵)
۲۷	CABG	کیفی	اسمی	مستقل	دارد/ندارد	دارد (۲۹) – ندارد (۱۷۲)
۲۸	فشارخون بالا	کیفی	اسمی	مستقل	دارد/ندارد	دارد (۱۰۵) – ندارد (۹۶)
۲۹	نارسایی کلیوی	کیفی	اسمی	مستقل	دارد/ندارد	دارد (۲۵) – ندارد (۱۷۶)
۳۰	دیابت	کیفی	اسمی	مستقل	دارد/ندارد	دارد (۶۱) – ندارد (۱۴۰)
۳۱	آسم	کیفی	اسمی	مستقل	دارد/ندارد	دارد (۱۶) – ندارد (۱۸۵)
۳۲	فیبرلاسیون دهلیزی	کیفی	اسمی	مستقل	دارد/ندارد	دارد (۷۷) – ندارد (۱۲۴)
۳۳	سکته مغزی	کیفی	اسمی	مستقل	دارد/ندارد	دارد (۲۰) – ندارد (۱۸۱)
۳۴	Ischemic	کیفی	اسمی	مستقل	دارد/ندارد	دارد (۱۱۳) – ندارد (۸۸)

[Downloaded from jhbmi.ir on 2026-11-23]

[DOI: 10.34172/jhbmi.2024.15]