مجله مدیریت اطلاعات سلامت
سال بیست و سوم/ شماره اول/ بهار ۱۴۰۵

Published by *Isfahan University of Medical Sciences*
DOI: 10.48305/him.2026.45810.1350

تحلیل و پیش بینی فرایندهای درمانی: رویکردی مبتنی بر علم داده فرایندگرا

محمدامین،پیریان [۱]، توکتم خطیبی [۲]

مقاله پژوهشی

چکیده

مقدمه: فرآیندهای مراقبت‌های بهداشتی به دلیل پیچیدگی، نیازمند رویکردهای داده‌محور برای مدیریت اطلاعات سلامت هستند. این مطالعه با تلفیق فرایندکاوی و یادگیری عمیق، پیش‌بینی فعالیت‌های بعدی در بخش اورژانس عفونی را مبتنی بر نگاره‌های رویداد هدف قرار داده تا مدیریت بهینه‌تر را تسهیل کند.

روش بررسی: نگاره رویداد استفاده شده شامل ۱۰۵۰بیمار که با نشانه‌های عفونت و علایم سپسیس به بخش اورژانس بیمارستان مراجعه کردند می باشد. که یک سری از درمان‌های خاص برای آن‌ها انتظار می‌رود. در ابتدا برای شفافیت جریان بیماران به کمک ابزارهای فرآیندکاوی از نگاره رویداد نقشه برداری صورت گرفت. سپس به کمک داده‌های فرایندی تجزیه و تحلیل فرایندی انجام شد در گام بعدی مدل پیشنهادی با استفاده ازیادگیری عمیق و به طور مشخص معماری شبکه‌های ترانسفورمرگام بعدی هر بیمار در مجموعه پیش‌بینی شد . مدل پیشنهادی با استفاده از معیارهایی همچون صحت (Accuracy) مورد ارزیابی قرارگرفت.

یافته‌ها: فرایندکاوی، گلوگاه‌هایی چون تاخیر در آزمایش‌های تشخیصی و ترافیک واحدهای آزمایش خون را شناسایی کرد. مدل ترانسفورمر با دقت میانگین۸۳/۵ درصد در پیش‌بینی فعالیت بعدی، نسبت به روش‌های موجود ۱۳درصد بهبود نشان داد و خطای میانگین مطلق (MAE) برای زمان رویداد بعدی (۰/۹۳ روز) و زمان باقی‌مانده (۵/۴۳ روز) را کاهش داد.

نتیجه‌گیری: مدل ارایه شده، ابزاری قدرتمند برای تحلیل و بهبود فرایندهای درمانی است. این رویکرد می‌تواند به کاهش زمان انتظار بیماران، بهبود تخصیص منابع و ارتقای تصمیم‌گیری مبتنی بر داده در محیط‌های درمانی کمک کند. یافته‌های این پژوهش گامی مؤثر در توسعه سیستم‌های مدیریت اطلاعات سلامت و بهبود خدمات درمانی محسوب می‌شود.

واژه‌های کلیدی: علم داده؛ فرایندکاوی؛ یادگیری عمیق؛ مدیریت اطلاعات سلامت

پیام کلیدی: یافته‌های این پژوهش به مدیران بیمارستان‌ها کمک می‌کند تا با درک بهتر فرایندها، تصمیم‌گیری دقیق‌تری برای تخصیص منابع و کاهش زمان انتظار بیماران داشته باشند. این ابزارها، بدون نیاز به تغییرات فوری، اطلاعات ارزشمندی برای بهبود مدیریت اطلاعات سلامت فراهم می‌کنند و کیفیت خدمات بیمارستانی را ارتقا می‌دهند.

دریافت مقاله:۱۴۰۴/۷/۲۳ پذیرش مقاله:۱۴۰۴/۱۱/۲۷ تاریخ انتشار: ۱۴۰۵/۱/۱۰

ارجاع: پیریان محمدامین، خطیبی توکتم. **تحلیل و پیش بینی فرایندهای درمانی: رویکردی مبتنی بر علم داده فرایندگرا.** مدیریت اطلاعات سلامت۱۴۰۵؛۲۳ (۱): ۳۱-۲۳.

مقدمه

مدل‌های فرایندی چارچوبی حیاتی برای مستندسازی، توصیف و تحلیل جریان‌های کاری هستند و به مدیران بیمارستان امکان می‌دهند مسیرهای درمانی پیچیده را به‌صورت تصویری مشاهده و بهینه‌سازی کنند. (۱). این مدل‌ها بر پایه سامانه‌های اطلاعاتی بنا شده‌اند که داده‌هایی مانند پذیرش، آزمایش‌های تشخیصی و تخصیص کارکنان را ثبت کرده و ستون فقرات عملیات در سیستم‌های نوین درمانی هستند. (۲). سامانه‌های آگاه از فرایند (Process-Aware Information Systems) ، گزارش‌های دقیق و جامعی از رویدادها تولید می‌کنند که منبعی غنی برای تحلیل‌های عملیاتی فراهم می‌آورد (۳). فرایندکاوی با استفاده از این گزارش‌ها، نحوه اجرای واقعی فرایندها را نشان داده و دیدگاه‌های کاربردی درباره ناکارآمدی‌هایی مانند تاخیر یا مصرف بیش ازحد منابع در اختیار مدیران قرار می‌دهد (۴). استخراج دانش از این گزارش‌ها، به بهبود مدیریت فرایندهای کسب‌وکار کمک کرده و بیمارستان‌ها را در تصمیم‌گیری بهتر و تخصیص بهینه منابع یاری می‌رساند (۵).

در حوزه سلامت، فرایندکاوی در مسیرهای بالینی انجام شده اما با چالش‌هایی همراه است، زیرا محیط‌های بیمارستانی پیوسته در حال تغییر بوده و وضعیت بیماران و تصمیمات کادر درمان تنوع بالایی دارند (۶). مطالعات اخیر نقش فزاینده آن را در مدیریت اطلاعات سلامت و کمک به پایبندی به دستورالعمل‌های بالینی و بهینه‌سازی درمان نشان داده‌اند (۷). پژوهش Kurniati و همکاران (۸) نیز ترسیم مسیرهای بیماران را بدون نیاز به تغییرات عملیاتی فوری، ابزاری برای برنامه‌ریزی راهبردی معرفی کرده است.

۱. کارشناس ارشد، مهندسی سیستم های سلامت، دانشکده مهندسی صنایع و سیستم ها ،دانشگاه تربیت مدرس،تهران،ایران.
۲. دانشیار، رشته مهندسی صنایع ، دانشکده مهندسی صنایع و سیستم ها، دانشگاه تربیت مدرس، تهران، ایران
نویسنده طرف مکاتبه: توکتم خطیبی ، دانشیار، رشته مهندسی صنایع ، دانشکده مهندسی صنایع و سیستم ها، دانشگاه تربیت مدرس، تهران، ایران
Email: toktam.khatibi@modares.ac.ir

This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 Unported License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

یادگیری عمیق نیز در پایش پیش‌بینانه فرایندها (پیش‌بینی مراحل آتی) نقش مهمی دارد. اگرچه شبکه‌های عصبی در پردازش زبان طبیعی شناخته شده‌اند، کاربرد آن‌ها در پیش‌بینی فرایندهای درمانی به دلیل نیاز به مدیریت پیشگیرانه به سرعت در حال رشد است (۹).

پژوهش Zilker و همکاران ظرفیت این رویکرد را در پیش‌بینی جریان بیماران و تخصیص بهینه منابع تایید کرده است (۱۰). این قابلیت‌ها ابزارهای داده‌محوری در اختیار مدیران می‌گذارند تا چالش‌های عملیاتی را پیش از وقوع مدیریت کنند.

مدل‌های سنتی یادگیری ماشین و یادگیری عمیق به‌صورت نظارت شده با گزارش‌های تاریخی آموزش می‌بینند. شبکه‌ها Long short-term memory (LSTM) در شناسایی الگوهای متوالی موثرند، اما در وابستگی‌های بلندمدت فرایندهای پیچیده درمانی دچار کاستی هستند (۱۱). در مقابل، مدل‌های ترنسفورمر که نخستین‌بار در پژوهش Vaswani و همکاران (۱۲) معرفی و در سال‌های اخیر بهبود یافته‌اند، با بهره‌گیری از مکانیزم توجه، وابستگی‌های بلندمدت و پردازش موازی را ممکن ساخته‌اند. ترکیب رمزگذاری موقعیتی و شبکه‌های باقی‌مانده، دقت بالایی در پیش‌بینی‌های متوالی ایجاد کرده و این مدل‌ها را برای داده‌های دنباله‌دار سلامت مناسب ساخته است(۱۳). ساختارهای مختلفی برای این مدل‌ها بررسی شده که جزییات آن در بخش روش بررسی بیان خواهد شد.

در این مطالعه از داده‌های واقعی بخش اورژانس یک بیمارستان شامل ۱۰۵۰ بیمار با علائم سپسیس استفاده شد (۱۴). فعالیت‌های متعدد طی اقامت، مسیرهای متنوعی برای هر بیمار ایجاد می‌کند. داده‌ها شامل ویژگی‌های ثابت و متوالی (جمعیت‌شناختی، آزمایش‌ها، درمان‌ها و تشخیص‌ها) برای پیش‌بینی مسیرهای فردی است. این داده‌ها با فرایندکاوی پیشرفته تحلیل شدند و مدل ترنسفورمر برای پیش‌بینی گام بعدی درمان به کار رفت.

این پژوهش با هدف به‌دست آوردن بینش‌هایی درباره مسیرهای مراقبت بیماران انجام شده است. همچنین، کاربرد فرآیندکاوی در حوزه سلامت را مورد بررسی قرار می‌دهد و در انتها، دقت مدل‌های ترنسفورمر را در پیش‌بینی گام بعدی درمان می‌آزماید.تلفیق این دو رویکرد، ابزارهای داده‌محوری برای بهبود تصمیم‌گیری، تخصیص منابع و کیفیت خدمات در مدیریت اطلاعات سلامت فراهم می‌کند.

روش بررسی

این پژوهش با رویکردی گذشته‌نگر و اکتشافی به بررسی فرآیندهای درمانی در بخش اورژانس بیمارانی پرداخته و تلاش کرده است تا با ترکیب دو رویکرد داده‌محور یعنی فرایندکاوی و یادگیری عمیق، دیدگاه‌های جدیدی در خصوص جریان‌های کاری مراقبت‌های درمانی ارائه دهد.

در این پژوهش از روش‌شناسی کریسپ استفاده شده است. روش‌شناسی کریسپ یک مدل فرایندمحور برای انجام پروژه‌های علم داده است که راهکاری نظام‌مند، مفید و کاربردی برای این پروژه‌ها ارایه می‌کند (۱۵). با توجه به روش‌شناسی کریسپ مراحل به ترتیبی که در ادامه آمده، انجام شده است.

گام اول (درک مسیله): مسئله پژوهش، تحلیل رفتار و پیش‌بینی فرایندها با راهبرد پایش پیش‌بینانه است. از فرایندکاوی برای ترسیم مسیرهای درمانی و شناسایی گلوگاه‌ها استفاده شد. سپس مدل ترنسفورمر برای پیش‌بینی فعالیت بعدی به کار رفت. هدف، ارائه ابزارهای تصمیم‌گیری داده‌محور برای مدیران بدون نیاز به مداخله فوری است. این رویکرد ترکیبی در راستای پیشرفت انفورماتیک سلامت نوین می‌باشد.

گام دوم(شناخت مجموعه‌داده): داده‌های این مطالعه مربوط به یک مجموعه داده‌ی عمومی از بیماران یک مرکز درمانی در هلند بوده و فرایندهای بخش اورژانس را در بازه‌ی زمانی یک‌ساله پوشش می‌دهد. این اطلاعات از سیستم برنامه ریزی منابع سازمانی بیمارستان گرفته شده است و شامل ۱۵۲۱۴ رویداد مربوط به ۱۰۵۰ مورد سپسیس ، از سال ۲۰۱۳ تا ۲۰۱۵ می‌باشد. علاوه بر این، رویدادها به ۱۶ فعالیت بیمارستانی یا کلاس تقسیم شده‌اند که هر مورد نمایانگر مسیر درمانی یک بیمار در بیمارستان است. محدوده‌ی داده‌ها از لحظه‌ی پذیرش در تریاژ اورژانس تا ترخیص یا فوت یا انتقال به بخش دیگر را در بر می‌گیرد. این داده ها از مخزن TU.Center۴ استخراج شده‌اند (۱۴). داده‌ها ترکیبی از ویژگی‌های ایستا (مانند سن و جنسیت بیماران) و ویژگی‌های ترتیبی (نظیر زمان‌سنجی رویدادها، درمان‌ها و تشخیص‌ها) بوده‌اند. این ویژگی‌ها بستری مناسب برای تحلیل مسیرهای گوناگون درمانی از طریق فرآیندکاوی و همچنین مدل‌سازی پیش‌بینانه فراهم کردند. تمامی داده‌ها به‌صورت ناشناس بوده و به دلیل عمومی بودن آن‌ها، نیاز به اخذ تاییدیه اخلاقی اضافی وجود نداشته است. پیچیدگی و تنوع این مجموعه داده، آن را به گزینه‌ای مناسب برای ارزیابی ابزارهای تحلیل فرآیند و پیش‌بینی در محیط‌های درمانی واقعی تبدیل کرده است .در درجدول ۱ ویژگی‌های آماری مجموعه داده به صورت خلاصه نشان داده شده است... ..

جدول ۱: آمارتوصیفی نگاره رویداد (ویژگی های زمانی برحسب روز است.)

تعداد موارد	تعداد رویداد	تعداد فعالیت	متوسط طول دنباله	بیشترین طول دنباله	متوسط زمان یک مورد	بیشترین زمان یک مورد	تعداد شاخه‌ها
۱۰۵۰	۱۵۲۱۴	۱۶	۱۴/۴۸	۱۸۵	۲۸/۴۸	۴۲۲	۸۴۵

در این پژوهش از فرآیندکاوی در دو سطح استفاده شده است: نخست برای تحلیل اکتشافی داده و درک مسایل بالینی و شناسایی گلوگاه‌ها ، سپس نتایج حاصل از تحلیل فرایندکاوی به عنوان دانش پیشین برای انتخاب ویژگی‌های کلیدی ورودی به مدل ترنسفورمر به کار گرفته شد.

برای شناسایی گلوگاه‌ها و فعالیت‌های پرتردد در مسیر درمان ، از ابزارهای مختلف فرایندکاوی مانند الگوریتم Heuristics Miner، الگوریتم کشف استخراج کننده بصری القایی و ابزار InterPretA در نرم‌افزار ProM استفاده شد. انتخاب الگوریتم Heuristics Miner به دلیل توانایی آن در مقابله با نویزهای موجود در داده‌های سلامت و تولید مدل‌های قابل تفسیر بود. خروجی این تحلیل به صورت نمودارهای گراف جهت‌دار ارابه شد که در آن توالی فعالیت‌ها، فراوانی انتقال‌ها و زمان‌های انتظار به‌تصویر کشیده شدند. این نقشه‌های فرآیند به مدیران بیمارستان کمک می‌کند تا تاخیر های احتمالی مانند تاخیر در تصویربرداری تشخیصی را شناسایی کرده و تصمیمات بهتری برای تخصیص منابع اتخاذ کنند.

در این تحلیل معیارهایی همچون فراوانی فعالیت‌ها، میانگین زمان انجام هر فعالیت و تنوع نسخه‌های فرآیند (process variants) در نظر گرفته شد. برای اطمینان از دقت مدل، آستانه برازش (fitness threshold) برابر با ۰/۹ تعیین شد. خروجی‌ها شامل نقشه‌های فرآیند، آمار فراوانی فعالیت‌ها و زمان انتقال بین رویدادها بودند. این یافته‌ها به مدیران در درک بهتر ناکارآمدی‌های موجود و تخصیص بهینه منابع کمک کرد.

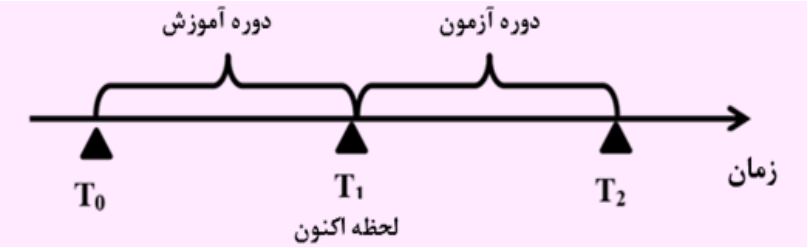
گام سوم (پیش پردازش داده): داده‌های خام ابتدا به فرمت استاندارد XES (فرمت رایج فرایندکاوی) تبدیل شدند. برای این کار از پایتون، کتابخانه‌های PM4PY وOpenXES ، همچنین نرم‌افزار ProM استفاده گردید. ویژگی‌های ایستا (مانند دموگرافیک) از ویژگی‌های ترتیبی (مانند توالی رویدادها) تفکیک شدند تا هم تحلیل فرایند و هم مدل‌سازی پیش‌بینانه ممکن شود. این مرحله شامل زیربخش‌هایی است که در ادامه آمده است.

• پاک‌سازی مجموعه داده: مدیریت داده‌های گم‌شده و رکوردهای ناقص، حذف ستون‌های کم ارزش

• تبدیل مجموعه داده: یکسان‌سازی قالب و نوع داده‌ها، نرمال‌سازی داده‌ها، استانداردسازی طول دنباله‌ها

• مدیریت داده‌های نامتوازن: در این پژوهش به جای oversampling یا undersampling از یادگیری حساس به هزینه استفاده شد. در این روش، به هر کلاس وزنی اختصاص می‌یابد تا خطای پیش‌بینی کلاس‌های کم‌تعداد بیشتر جریمه شود. این کار حساسیت مدل را به کلاس‌های نادر افزایش داده و دقت پیش‌بینی آن‌ها را بهبود می‌بخشد.

• تقسیم مجموعه‌داده: داده‌ها بر اساس ترتیب زمانی مرتب شدند. ۷۰ درصد اول برای آموزش و ۳۰ درصد باقی‌مانده برای اعتبارسنجی و آزمون استفاده گردید...



شکل ۱: تقسیم زمانی مجموعه های آموزشی و آزمایشی

• استخراج ویژگی: برای پیش‌بینی فعالیت بعدی در فرایند، ویژگی‌های اصلی مورد استفاده شامل شناسه مورد، نوع فعالیت، و زمان آغاز فعالیت‌ها است. حال برای استخراج ویژگی‌های عمیق‌تر در این پژوهش از تکنیک‌های مختلفی استفاده شد که در ادامه به دو مورد از آن‌ها اشاره می‌شود:

• نمایش دنباله فعالیت‌ها به شکل برداری (Trace Embedding): در این روش، دنباله فعالیت‌های هر بیمار به یک بردار عددی با طول ثابت تبدیل می‌شود تا مدل بتواند الگوهای مشابه را در مسیرهای درمانی مختلف تشخیص دهد.

• تعبیه‌یابی‌های موقعیتی (Positional Embeddings): مدل ترنسفورمر برخلاف شبکه‌های بازگشتی، ترتیب فعالیت‌ها را ذاتا درک نمی‌کند. بنابراین با این روش، جایگاه هر فعالیت در توالی به صورت عددی به مدل آموزش داده می‌شود تا ترتیب زمانی رویدادها حفظ گردد.

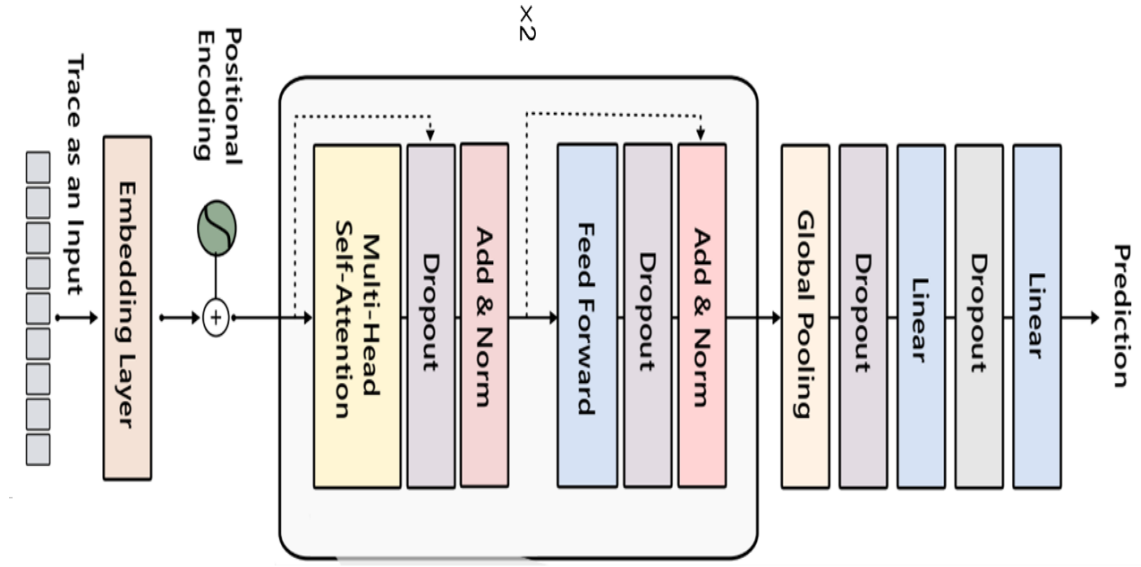
گام چهارم (مدل سازی): با توجه به نتایج رضایت‌بخش مدل‌های ترنسفورمر در پردازش زبان طبیعی و ماهیت دنباله‌ای فرایندها، مدل ترنسفورمر Vaswani و همکاران به عنوان پایه انتخاب شد. (۱۲) در ادامه تلاش شد با تغییر معماری و تنظیم هایپر پارامترها مدل بهبود یابد. خروجی فرایندکاوی شامل گلوگاه‌ها (مانند تاخیر در آزمایش CRP) و فعالیت‌های پرتکرار (مانند تزریق آنتی‌بیوتیک وریدی) به عنوان دانش پیشین برای انتخاب ویژگی‌های ورودی مدل ترنسفورمر به کار رفت. فعالیت‌های با بیشترین فراوانی یا زمان انتظار در نقشه‌های فرایند، در تعبیه‌ی توالی وزن بالاتری دریافت کردند.

• معماری مدل ترنسفورمر

مدل پیشنهادی مبتنی بر معماری ترنسفورمر است که معماری کلی مدل در شکل۲ نشان داده شده است.

ورودی مدل، توالی رمزگذاری‌شده فعالیت‌ها به همراه اطلاعات موقعیتی است.

http://him.mui.ac.ir



شکل ۲ : معماری شبکه پایه ترنسفورمر (۱۶)

لایه‌های اصلی مدل پیشنهادی به ترتیب عبارتند از:

۱ .لایه تعبیه‌سازی توالی (Trace Embedding): هر فعالیت به یک بردار عددی با طول ثابت نگاشت می‌شود.

۲ .رمزنگار موقعیتی (Positional Encoding): یک بردار ۳۶ بعدی به هر موقعیت اختصاص می‌یابد تا ترتیب فعالیت‌ها در توالی حفظ شود.

۳ .بلوک‌های توجه (Attention Blocks): شامل ۲ بلوک متوالی که هر بلوک از زیرلایه‌های زیر تشکیل شده است:

- توجه چندسری (Multi-Head Attention): با ۴ سر برای یادگیری همزمان جنبه‌های مختلف ارتباط بین فعالیت‌ها.
- لایه Dropout: با نرخ ۰.۱برای کاهش بیش‌برازش.
- لایه Normalization : برای پایدارسازی و سرعت یادگیری.
- لایه Feed-Forward (Position-wise) : شامل دو لایه چگال با ۳۲ و ۱۲۸ نرون و تابع فعال‌سازی ReLU
- اتصالات باقی‌مانده: (Residual Connections): برای جلوگیری از محو شدن گرادیان.

۴ .لایه تجمیع سراسری (Global Max-Pooling): برای کاهش ابعاد و استخراج ویژگی‌های مهم.

۵ .لایه خروجی برای پیش‌بینی فعالیت بعدی: لایه Fully Connected با ۱۶ نرون (به تعداد فعالیت‌ها) و تابع Softmax همراه با تابع خطای Categorical Cross-Entropy

- بهینه‌سازی مدل

هایپر پارامترها با روش جستجوی شبکه‌ای (Grid Search) تنظیم شدند. نرخ یادگیری بهینه ۰/۰۰۱، بهینه‌سازAdam، اندازه دسته (Batch Size) برابر ۳۲، و تعداد دوره‌های آموزشی ۱۰۰ با توقف زودهنگام (Early Stopping) در نظر گرفته شد، همچنین برای منظم‌سازی از Dropout و نرمال‌سازی L2 استفاده گردید.

گام پنجم (معیارهای ارزیابی و اعتبارسنجی): صحت (Accuracy) معیاری رایج برای پیش‌بینی فعالیت بعدی است (۱۷)، اما در داده‌های نامتوازن می‌تواند گمراه‌کننده باشد. به همین دلیل، در این پژوهش از صحت وزنی (Weighted Accuracy) و F-Score وزنی استفاده شد .F-Score وزنی با در نظر گرفتن عدم توازن کلاس‌ها، ترکیبی از دقت (Precision) و بازخوانی (Recall) است. دقت، صحت پیش‌بینی‌های مدل و بازخوانی، کامل بودن آن را نشان می‌دهد.

این پژوهش، تحلیل فرایند و مدل‌سازی پیش‌بین را در چرخه‌ای تکرارشونده تلفیق کرد. خروجی‌های فرایندکاوی (مانند مکان گلوگاه‌ها و فراوانی فعالیت‌ها) مبنای انتخاب ویژگی‌های ورودی مدل ترنسفورمر قرار گرفت و پیش‌بینی‌های مدل با رویدادهای واقعی مقایسه شد.تمامی اصول اخلاقی رعایت شده است. داده‌ها ناشناس بوده و تحلیل‌ها به صورت ثانویه انجام شده است.

یافته‌ها

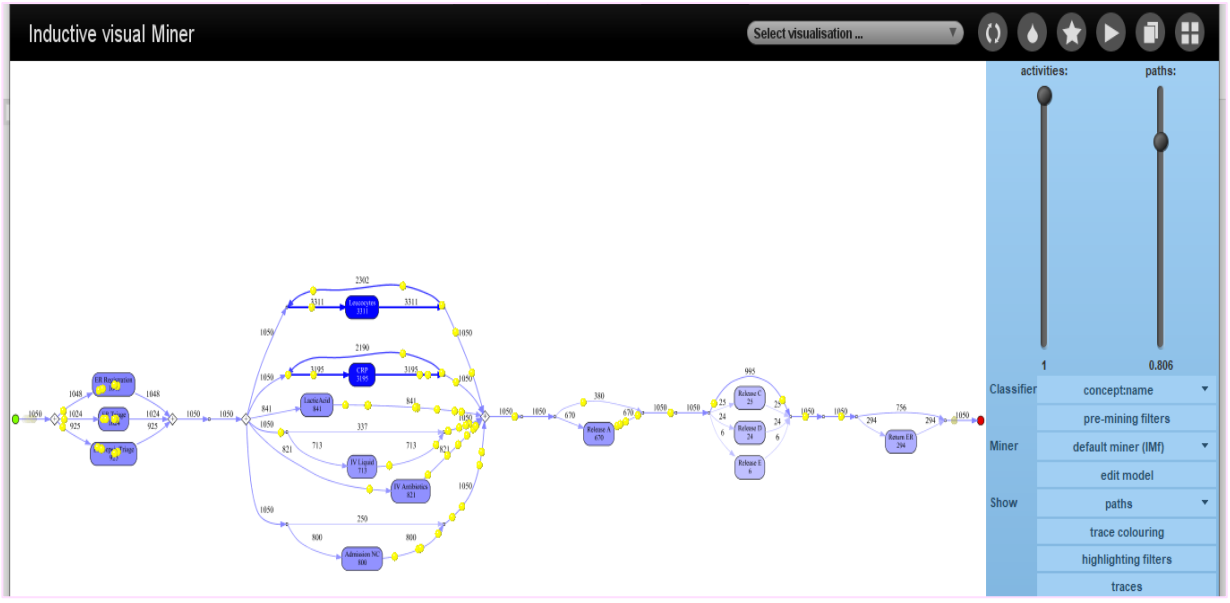
تحلیل فرایندکاوانه :

یکی تجسم‌های کلیدی برای نگاره‌ی رویداد، نمودار نقطه چین است. در شکل۳ یک نوع از این دسته نمودارها را مشاهده می‌کنید که در آن هررنگ نشان دهنده یک فعالیت است و محور طولی هر فعالیت و محور عرضی نشان دهنده تاریخ است.

شکل ۳: نمونه ای از نمودارنقطه ای گزارش رویداد

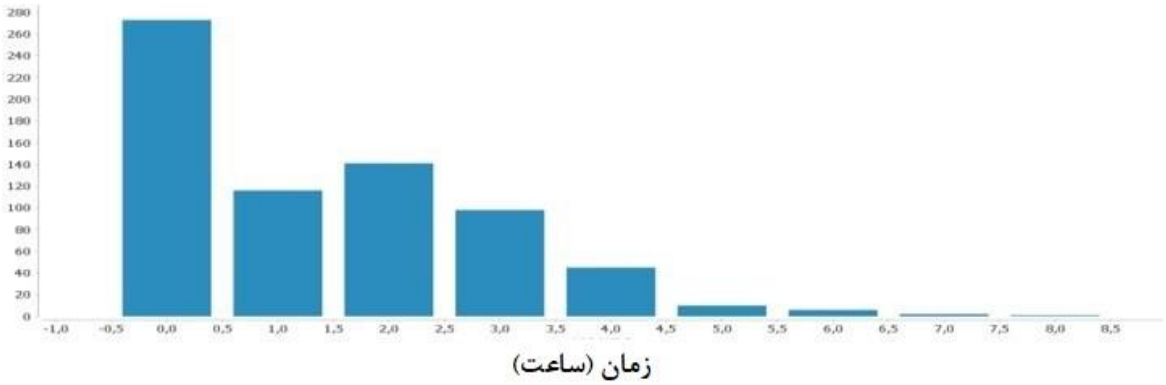
شکل ۳ دنباله وقوع فعالیت های مختلف در بستر زمان را نشان می‌دهد. به عنوان مثال پذیرش بخش عادی وآزمایش CRP به شکل مستمر رخ داده‌اند. همچنین در مقایسه انواع ترخیص مشاهده می‌شود که ترخیص نوع A به شکل روتین اما ترخیص نوع E تنها در چند نمونه با فاصله زمانی زیاد مشاهده شده است .

پس از تجسم داده‌ها، از الگوریتم کشف فرایند برای درک بهتر فرایند ثبت شده استفاده شد. الگوریتم Inductive Visual Miner مدل پویایی از فرایند اورژانس استخراج کرد که به صورت انیمیشن، وضعیت فرایند را در هر لحظه نمایش می‌دهد. در شکل ۴ یک لحظه از این نمودار نمایش داده شده است.



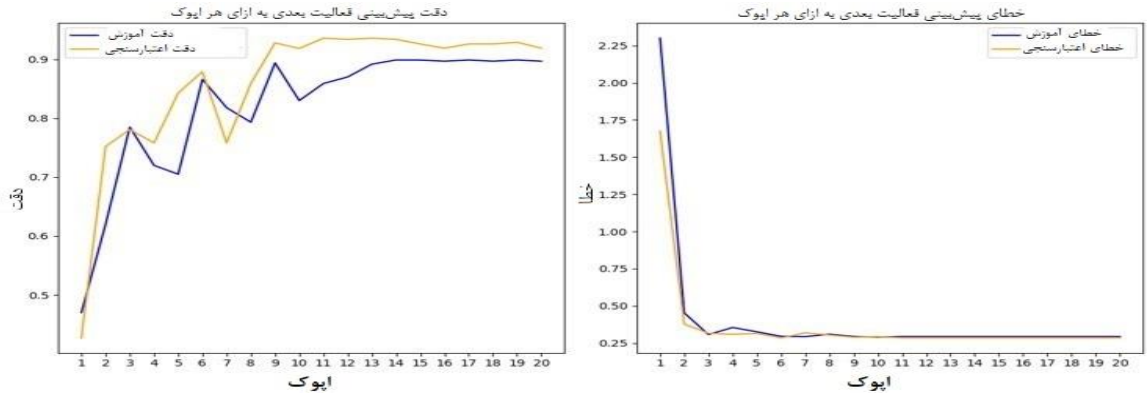
شکل۴ : مدل فرایندبخش اورژانس بیمارستان هلندی که توسط استخراج کننده بصری القایی کشف شده‌است .

در شکل ۴، مسیرهای مربوط به آزمایش CRP و شمارش گلبول سفید تیره‌تر نشان داده شده‌اند که بیانگر تراکم و ترافیک بیشتر روی آن‌هاست. با افزونه InterPretA، زمان بین فعالیت‌ها و تعداد وقوع آن‌ها به صورت نمودار میله‌ای مقایسه شد. در شکل ۵، محور افقی ساعت‌های پس از تریاژ و محور عمودی تعداد موارد را نشان می‌دهد. در بسیاری از موارد، پروتکل تجویز آنتی‌بیوتیک وریدی (IV) ظرف یک ساعت پس از تریاژ نقض شده است.



شکل ۵ : توزیع تعداد بیماران براساس زمان بین فعالیت های تریاژ اتاق اورژانس و اولین استعمال آنتی بیوتیک داخل وریدی

پیش‌بینی یادگیری عمیق: روند آموزش مدل را برای پیش‌بینی فعالیت بعدی در شکل ۶ نشان داده‌شده‌است. در ۴–۵ ایپوک اول، دقت به‌سرعت افزایش می‌یابد که نشان‌دهنده یادگیری الگوهای اولیه است. بین ایپوک‌های ۵ تا ۱۰، نوساناتی ناشی از stochastic gradient descent دیده می‌شود. پس از ایپوک ۱۰، دقت به حدود ۰/۹ می‌رسد. اختلاف کم بین دقت آموزش و اعتبارسنجی نشان می‌دهد مدل به یک تعادل رسیده و بیش‌برازش قابل توجهی رخ نداده است.



شکل ۶ : نمودارهای توابع دقت و زیان برای پیش‌بینی فعالیت بعدی به ازای هر ایپوک

جدول ۲ نتایج پیش‌بینی فعالیت بعدی توسط مدل پیشنهادی را نشان می‌دهد. برای مقایسه، دو مدل دیگر که در پژوهش‌های پیشین (۱۶، ۱۷) بر روی مجموعه‌داده سپسیس مورد استفاده قرار گرفته‌اند، انتخاب شدند.مدل پیشنهادی با دقت ۸۳ درصد عملکرد بهتری نسبت به سایر مدل‌ها دارد.

جدول ۲ : مقایسه عملکرد مدل‌پیشنهادی و مدل‌ها ی دیگر بر روی مجموعه‌داده سپسیس

مدل	فعالیت بعدی Accuracy/%	زمان رویداد بعدی MAE/day	زمان باقی‌مانده MAE/day
مدل پیشنهادی	۸۳/۵	۰/۹۳	۵/۴۳
مدل Process Transformer (۱۶)	۶۴/۶۱	۱/۴۰	۱۴/۷۷
مدل Taxو همکاران (۲۲)	۶۴/۲۲	–	–

بحث

مدل پیشنهادی این مطالعه که ترکیبی از فرایندکاوی و معماری ترنسفورمر است، با دقت ۸۳.۵ درصد در پیش‌بینی فعالیت بعدی، عملکردی برتر نسبت به مدل‌های سنتی نظیر LSTM و CNN داشته و بهبود ۱۳ درصدی را نشان داده است (۶). خطای میانگین مطلق (MAE) برای زمان رویداد بعدی ۰/۹۳ روز (حدود ۲۲ ساعت) و برای زمان باقی‌مانده ۵/۴۳ روز محاسبه شد که نسبت به مدل‌های پیشین کاهش چشمگیری داشته است. از منظر مدیریتی، این رویکرد، ابزارهای پیش‌بینی مناسبی برای برنامه‌ریزی منابع فراهم می‌کند. برای نمونه، پیش‌بینی فعالیت‌های پرترافیک مانند آزمایش CRP به مدیران امکان می‌دهد نیروی انسانی یا تجهیزات را به‌موقع تنظیم کرده و زمان انتظار بیماران را کاهش دهند. همچنین به بهبود برنامه‌ریزی ترخیص و کاهش کمبود تخت‌های بستری کمک می‌کند. این کارکردها در قالب چارچوب CRISP-DM انجام شد که بر تصمیم‌سازی استراتژیک مبتنی بر داده تأکید دارد. علاوه بر این، استفاده از معیارهای دقت وزنی وF-score ، عملکرد مدل را در مواجهه با عدم تعادل کلاس‌ها بهبود بخشید؛ برای نمونه، فعالیت‌های کم‌رخداد ولی حیاتی مانند پذیرش در ICU با دقت قابل اعتمادی پیش‌بینی شدند .

مجموعه داده مورد استفاده در این پژوهش یکی از غنی‌ترین داده‌های حوزه فرایندکاوی سلامت است و پژوهش‌های متعددی بر روی آن انجام شده است. استفاده از الگوریتم Heuristics Miner امکان شناسایی گلوگاه‌های عملیاتی نظیر تأخیر در آزمایش‌های خون و فعالیت‌های پرترافیک مانند آزمایش CRP را فراهم کرد Mannhardt .و Blinde (۱۸) با استفاده از الگوریتم‌های فرایندکاوی، مسیرهای بیماران سپسیس را تحلیل کردند. یافته‌های این پژوهش با نتایج پژوهش های مذکور هم‌راستا است. اما مطالعه ایشان فاقد بخش پیش‌بینی بود. نقشه‌های فرایندی تولیدشده، دیدگاه عملیاتی برای مدیران بیمارستانی فراهم کردند تا بدون نیاز به تغییرات فوری، برنامه‌ریزی استراتژیک منابع را انجام دهند. این یافته با پژوهش De Roock و Martin (۵) درباره نقش فرایندکاوی در بهبود جریان کاری نظام سلامت همسو است .

همچنین Novak و همکاران (۱۹) نشان دادند که نقشه‌های فرایند می‌توانند به تصمیم‌گیری مدیریتی کمک کنند. در پژوهش حاضر، این گلوگاه‌های شناسایی‌شده به عنوان دانش پیشین برای وزن‌دهی به ویژگی‌های ورودی مدل ترنسفورمر مورد استفاده قرار گرفتند که وجه تمایز این مطالعه با پژوهش‌های پیشین است.

در حوزه پیش‌بینی فعالیت بعدی، Taxو همکاران(۲۰) از مدل LSTM استفاده کردند و به دقت ۶۴/۲۲ درصد دست یافتند. Bukhsh و همکاران (۱۶) مدل Process Transformerرا معرفی کردند که دقت ۶۴/۶۱ درصد را بر روی همین مجموعه‌داده به دست آورد. همچنین Pasquadibisceglie و همکاران (۱۷) از شبکه‌های عصبی کانولوشنال (CNN) برای پیش‌بینی فرایند استفاده کردند که دقتی حدود ۶۲ درصد گزارش شد.

در مقایسه با این پژوهش‌ها، مدل پیشنهادی ما با دقت۸۳/۵ درصد ، بهبود ۱۳ تا ۲۱ درصدی را نشان می‌دهد. این برتری عمدتاً ناشی از دو عامل است : نخست استفاده از معماری ترنسفورمر با مکانیزم خودتوجه که وابستگی‌های بلندمدت بین رویدادها را بهتر از LSTM مدل می‌کند. دوم یادگیری حساس به هزینه-Cost) (Sensitive Learningبا اختصاص وزن‌های بالاتر به خطاهای پیش‌بینی در کلاس‌های کم‌تعداد(مانند انتقال بهICU)، مشکل عدم تعادل داده را تا حد زیادی مرتفع ساخته است. در مجموعه‌داده مورد استفاده، فعالیت‌های نادر و حیاتی با وزن‌دهی مناسب، دقت پیش‌بینی بهتری یافتند (بخش یافته‌ها). ترکیب این دو عامل، مدل پیشنهادی را به ابزاری کارآمد برای محیط‌های درمانی پویا تبدیل کرده است.

این مطالعه با چند محدودیت مواجه است. نخست، استفاده از داده‌های پس‌نگرانه (گذشته‌نگر) و تک مرکزی بودن، تعمیم‌پذیری نتایج را به بیمارستان‌هایی با پروتکل‌ها یا ویژگی‌های جمعیتی متفاوت محدود می‌سازد. دوم، حجم مجموعه‌داده (۱۰۵۰ بیمار) اگرچه برای فرایندکاوی مناسب است، اما برای آموزش مدل‌های عمیق با پارامترهای زیاد نسبتاً محدود محسوب می‌شود. سوم، تعیین آستانه ۰.۹ در الگوریتم Heuristics Miner، هرچند برای کاهش نویز مفید است، اما احتمال ساده‌سازی بیش از حد و نادیده‌گرفتن مسیرهای نادر اما حیاتی (مانند مسیرهای منجر به فوت) را افزایش می‌دهد.

با وجود این محدودیت‌ها، طراحی مطالعه بر پشتیبانی از تصمیم‌گیری‌های راهبردی متمرکز بوده است، نه تغییرات عملیاتی فوری؛ از این رو نتایج همچنان برای مدیران سلامت کاربردی و قابل استفاده است.

نتیجه‌گیری

یافته‌های این پژوهش نشان می‌دهد که تلفیق فرایندکاوی با مدل ترنسفورمر در بهینه‌سازی مدیریت فرایندهای درمانی اثربخش است. شناسایی گلوگاه‌ها و پیش‌بینی دقیق فعالیت‌ها، مدیران را در بهینه‌سازی منابع و کاهش زمان انتظار بیماران یاری می‌کند. با وجود چالش‌هایی مانند تعمیم‌پذیری داده‌ها و نیاز محاسباتی، این مطالعه پایه‌ای برای مدیریت داده‌محور سلامت و بهبود پیامدهای بالینی بیماران سپسیس فراهم می‌کند.

پیشنهادها

در گام‌های آتی، اعتبارسنجی مدل در محیط‌های بالینی برخط (Real-time) و بهره‌گیری از هوش مصنوعی قابل توضیح (مانند تحلیل وزن‌های توجه) برای افزایش شفافیت و اعتماد درمانگران توصیه می‌شود. همچنین، توسعه مدل برای بیماری‌های مزمن و سایر تخصص‌ها، بهینه‌سازی برای کاهش نیاز محاسباتی (به‌ویژه برای بیمارستان‌های کم‌منبع)، و اعتبارسنجی بر روی داده‌های مراکز درمانی ایران از اولویت‌های پژوهشی بعدی است. در نهایت، همکاری با شرکت‌های فناوری اطلاعات سلامت می‌تواند فرایند استقرار را تسهیل کند.

http://him.mui.ac.ir

تشکر و قدردانی

این مقاله برگرفته از پایان‌نامه‌ی کارشناسی ارشد به شماره ۲۳۰۸۲۳۶۹ در دانشگاه تربیت مدرس می‌باشد. نگارندگان مراتب سپاس و قدردانی خود را از اساتید و کارکنان این دانشگاه به‌خاطر همکاری و راهنمایی‌های ارزشمندشان در تمام مراحل انجام این پژوهش ابراز می‌دارند .

تضاد منافع

در انجام پژوهش حاضر، نویسندگان هیچ‌گونه تضاد منافعی نداشتند.

References

1. Lima RM, Lopes ET, Lopes DC, Gonçalves BS, Cunha PG. Hospital operations management–characterising patients' process flows in emergency departments. Business Process Management Journal. 2024 Dec 16;30(8):207-31.
2. Shojaei P, Vlahu-Gjorgievska E, Chow YW. Security and privacy of technologies in health information systems: A systematic literature review. Computers. 2024 Jan 31;13(2):41. Available from: https://www.mdpi.com/2073-431X/13/2/41
3. Naz TH. Change Mining and Logging in Process-Aware Information Systems: A Literature Review - Student Theses Faculty of Science and Engineering. Ubrugnl [Internet]. 2022 [cited 2025 Apr 18]; Available from: https://fse.studenttheses.ub.rug.nl/28646/
4. Zerbino P, Stefanini A, Aloini D. Process Science in Action: A Literature Review on Process Mining in Business Management. Technological Forecasting and Social Change. 2021 Nov;172:121021.
5. De Roock E, Martin N. Process mining in healthcare – An updated perspective on the state of the art. Journal of Biomedical Informatics. 2022 Mar;127:103995.
6. Dallagassa MR, dos Santos Garcia C, Scalabrin EE, Ioshii SO, Carvalho DR. Opportunities and challenges for applying process mining in healthcare: a systematic mapping study. Journal of ambient intelligence and humanized computing. 2022;13(1):165-82..
7. Gatta R, Vallati M, Fernandez-Llatas C, Martinez-Millana A, Orini S, Sacchi L, et al. Clinical Guidelines: A Crossroad of Many Research Areas. Challenges and Opportunities in Process Mining for Healthcare. Lecture notes in business information processing. 2019 Jan 1;545–56.
8. Kurniati AP, Gede, Kusuma GP, Siti Saadah, Osman NA, None Zulhelmy, et al. Patient Clustering to Improve Process Mining for Disease Trajectory Analysis Using Indonesia Health Insurance Dataset. 2024 May 24;88–93.
9. Rama-Maneiro E, Vidal JC, Lama M. Deep learning for predictive business process monitoring: Review and benchmark. arXiv preprint arXiv:2009.13251. 2020 Sep 24.
10. Zilker S, Weinzierl S, Kraus M, Zschech P, Matzner M. A machine learning framework for interpretable predictions in patient pathways: The case of predicting ICU admission for patients with symptoms of sepsis. Health Care Management Science. 2024;27(2):136-67.
11. Tanikonda A, Pandey BK, Peddinti SR, Katragadda SR. Application of transformer models for advanced process optimization and process mining. Journal of Science & Technology (JST). 2022 Sep;3(5).
12. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is All you Need. Advances in Neural Information Processing Systems. 2017;30:5998–6008.
13. Denecke K, May R, Rivera-Romero O. Transformer models in healthcare: a survey and thematic analysis of potentials, shortcomings and risks. Journal of medical systems. 2024 Feb 17;48(1):23.
14. Boudewijn van Dongen. BPI Challenge 2012. Data4tunl [Internet]. 2025 [cited 2025 Apr 18]; Available from: https://data.4tu.nl/articles/_/12689204/1
15. Schröer C, Kruse F, Gómez JM. A Systematic Literature Review on Applying CRISP-DM Process Model. Procedia Computer Science. 2021;181:526–34.
16. Bukhsh ZA, Saeed A, Dijkman RM. ProcessTransformer: Predictive Business Process Monitoring with Transformer Network. 2021.
17. Pasquadibisceglie V, Appice A, Castellano G, Malerba D. Using Convolutional Neural Networks for Predictive Process Analytics [Internet]. IEEE Xplore. 2019 [cited 2023 Apr 15]. p. 129–36. Available from: https://ieeexplore.ieee.org/abstract/document/8786066/
18. Mannhardt F, Blinde D. Analyzing the trajectories of patients with sepsis using process mining. InRadar+ emisa 2017 2017 (pp. 72-80). CEUR-ws. org. Available from: https://www.semanticscholar.org/paper/Analyzing-the-Trajectories-of-Patients-with-Sepsis-Mannhardt-Blinde/0897a405a97182bb004d865a1f0dd787a0fab0d3
19. Novak C, Pfahlsberger L, Bala S, Revoredo K, Mendling J. Enhancing decision-making of IT demand management with process mining. Business Process Management Journal. 2023 Dec 18;29(8):230-59.
20. Tax N, Verenich I, La Rosa M, Dumas M. Predictive business process monitoring with LSTM neural networks. InInternational conference on advanced information systems engineering 2017 May 27 (pp. 477-492). doi: 10.1007/978-3-319-59536-8_30.

Health Information Management
Vol 23/ No 1/ Spring 2026

Published by *Isfahan University of Medical Sciences*
DOI: 10.48305/him.2026.45810.1350

Analyzing and Predicting Treatment Processes: A Process-Oriented Data Science Approach

Mohammad Amin Pirian[1], Toktam Khatibi[2]

Original Article

Abstract

Introduction: Healthcare processes are inherently complex and demand data-driven strategies for effective health information management. This study integrates process mining with deep learning to predict subsequent activities in infectious disease emergency care based on event logs, aiming to support more efficient management.

Methods: The event log comprised 1,050 patients admitted to the emergency department with infection symptoms and suspected sepsis, for whom specific treatment pathways were expected. First, patient flows were mapped using process mining tools to visualize and clarify care trajectories. Next, process data were analyzed, and the proposed model—built on deep learning with a transformer-based architecture—was employed to predict each patient's next step. Model performance was evaluated using metrics such as accuracy.

Results: Process mining identified bottlenecks, including delays in diagnostic tests and congestion in blood test units. The transformer model achieved an average accuracy of 83.5% in next-activity prediction, representing a 13% improvement over existing approaches. It also reduced mean absolute error (MAE) to 0.93 days for next-event time and 5.43 days for remaining time.

Conclusion: The proposed model offers a robust tool for analyzing and improving clinical processes. This approach can reduce patient waiting times, optimize resource allocation, and enhance data-driven decision-making in healthcare. The findings contribute to advancing health information management systems and improving treatment services.

Keywords: Data Science; Process Mining; Deep Learning; Health Information Management

Received:15 Oct; 2025 Accepted: 17 Jun; 2025 Published: 30 Mar; 2026

Citation: Pirian MA, Khatibi T. **Analyzing and Predicting Treatment Processes: A Process-Oriented Data Science Approach.** Health Inf Manage 2026; 23(1): 23-31.

Article resulted from a Msc thesis with No 23082369

1 Msc, MSc Student, Systems and Industrial Engineering Department, Tarbiat Modares University, Tehran, Iran, Iran
2 .Associate Professor , Faculty of Industrial and Systems Engineering, Tarbiat Modares University; Tehran,, Iran

Corresponding Author: Toktam Khatibi, Associate Professor , Faculty of Industrial and Systems Engineering, Tarbiat Modares University; Tehran,, Iran Email: toktam.khatibi@modares.ac.ir