

Original article



An Intelligent Framework Based on Health Informatics for Diagnosing Liver Diseases Using Generative Adversarial Network-Based Data Augmentation and Ensemble Learning

Nazila Valizadeh jahatloo¹, Omid R. B. Speily^{2*}, Parviz Ghorbanzadeh³

1. M.Sc. Student, Department of Information Technology, Faculty of Modern Science & Technology, Urmia University of Technology, Urmia, Iran

2. Ph.D in Information Technology Engineering, Assistant Professor, Department of Information Technology, Faculty of Modern Science & Technology, Urmia University of Technology, Urmia, Iran

3. Ph. D in Artificial Intelligence, Urmia University of Technology, Urmia, Iran

ARTICLE INFO:

Article History:

Received: 3 Aug 2025

Accepted: 30 Mar 2026

Published: 21 Jun 2026

*Corresponding Author:

Omid R. B. Speily

Email: Speily@uut.ac.ir

Citation: Valizadeh jahatloo N, B. Speily OR, Ghorbanzadeh P. An Intelligent Framework Based on Health Informatics for Diagnosing Liver Diseases Using Generative Adversarial Network-Based Data Augmentation and Ensemble Learning. Journal of Health and Biomedical Informatics 2026; 13(1): 12-23. [In Persian]

Abstract

Introduction: Health is one of the fundamental aspects of human life, and chronic diseases pose a serious threat in this realm, entailing far-reaching consequences. With rapid advancements in artificial intelligence, intelligent medical diagnostic systems have seen significant development, facilitating early diagnosis, cost reduction, and more accurate clinical decision-making. Despite the application of various data mining techniques, the complexity and high dimensionality of medical data remain challenging. In this context, deep learning methods and generative models serve as powerful tools for analyzing such data.

Methods: This study proposes a novel method for improving the diagnosis of liver diseases, based on data preprocessing and augmentation using Generative Adversarial Networks (GANs) alongside a stacking ensemble learning approach. The Indian Liver Patient Dataset (ILPD) was used for evaluation. In the initial phase, the ensemble model achieved an accuracy of 77% using 10-fold cross-validation. Subsequently, the model was retrained using a GAN-based data augmentation technique, yielding an impressive accuracy of 99%.

Results: The results indicate that the appropriate application of preprocessing and data augmentation methods has a significant impact on improving model performance and enhancing the accuracy of early liver disease detection.

Conclusion: The proposed intelligent framework successfully overcomes the challenges of complexity and data imbalance in medical datasets, delivering a highly accurate and reliable model. This approach can serve as an efficient clinical decision-support system within the field of health informatics, offering significant assistance to healthcare professionals in the early diagnosis and improved management of liver diseases.

Keywords: Early Liver Disease Diagnosis, Deep learning, Adversarial Generating Network, Machine Learning



CrossMark

مقاله پژوهشی

یک چارچوب هوشمند مبتنی بر انفورماتیک سلامت برای تشخیص بیماری‌های کبدی با استفاده از افزایش داده مبتنی بر شبکه مولد تخصصی و یادگیری جمعی

نازیلا ولیزاده جهتلو^۱، امیدرضا بلوکی اسپیلی^{۲*}، پرویز قربانزاده^۳

۱. دانشجوی کارشناسی ارشد، گروه مهندسی فناوری اطلاعات، دانشکده علوم و فناوری های نوین، دانشگاه صنعتی ارومیه، ارومیه، ایران

۲. دکتری مهندسی فناوری اطلاعات، استادیار، گروه مهندسی فناوری اطلاعات، دانشکده علوم و فناوری های نوین، دانشگاه صنعتی ارومیه، ارومیه، ایران

۳. دکتری هوش مصنوعی، دانشگاه صنعتی ارومیه، ارومیه، ایران

چکیده

مقدمه: سلامت، یکی از بنیادی‌ترین ابعاد زندگی انسان است و بیماری‌های مزمن تهدیدی جدی در این حوزه محسوب می‌شوند که عوارض گسترده‌ای به دنبال دارند. با پیشرفت‌های سریع هوش مصنوعی، سامانه‌های تشخیص پزشکی هوشمند در زمینه تشخیص زودهنگام، کاهش هزینه‌ها و کمک به تصمیم‌گیری دقیق‌تر پزشکان توسعه چشم‌گیری یافته‌اند. با وجود کاربرد روش‌های گوناگون داده‌کاوی، پیچیدگی و ابعاد بالای داده‌های پزشکی همچنان چالش برانگیز است. در این میان، روش‌های یادگیری عمیق و مدل‌های مولد ابزارهایی توانمند برای تحلیل این داده‌ها هستند.

روش کار: در این پژوهش، روشی نوین برای بهبود تشخیص بیماری‌های کبدی پیشنهاد شده است که بر پایه پیش‌پردازش و افزایش داده‌ها با استفاده از شبکه‌های مولد تخصصی (GAN) و یادگیری جمعی از نوع استکینگ (Stacking) قرار دارد. برای ارزیابی، از مجموعه‌داده استاندارد بیماران کبدی هند (ILPD) استفاده شد. در مرحله نخست، مدل جمعی با اعتبارسنجی متقابل ۱۰-مرحله‌ای دقت ۷۷ درصد را به دست آورد. سپس، با استفاده از تکنیک افزایش داده مبتنی بر GAN، مجدداً مدل آموزش داده شد که این بار دقت چشمگیر ۹۹ درصد حاصل گردید.

یافته‌ها: نتایج نشان می‌دهد که به کارگیری مناسب روش‌های پیش‌پردازش و افزایش داده، تأثیر بسزایی در بهبود عملکرد مدل و ارتقای دقت تشخیص زودهنگام بیماری کبد دارد.

نتیجه‌گیری: چارچوب هوشمند پیشنهادی با غلبه بر چالش پیچیدگی و عدم توازن داده‌های پزشکی، مدلی بسیار دقیق و قابل اتکا ارائه می‌دهد. این رویکرد می‌تواند به عنوان یک سیستم پشتیبان تصمیم‌گیری بالینی کارآمد در حوزه انفورماتیک سلامت، به متخصصان در تشخیص زودهنگام و مدیریت بهتر بیماری‌های کبدی کمک شایانی نماید.

کلیدواژه‌ها: شبکه مولد متخصص، بیماری کبد، یادگیری ماشین، یادگیری عمیق، هوش مصنوعی، تشخیص بیماری

اطلاعات مقاله

سابقه مقاله

دریافت: ۱۴۰۴/۵/۱۲

پذیرش: ۱۴۰۵/۱/۱۰

انتشار برخط: ۱۴۰۵/۳/۳۱

*نویسنده مسئول:

امیدرضا بلوکی اسپیلی

ایمیل:

speily@uut.ac.ir

ارجاع:

ولیزاده جهتلو نازیلا، بلوکی اسپیلی امیدرضا، قربانزاده پرویز. یک چارچوب هوشمند مبتنی بر انفورماتیک سلامت برای تشخیص بیماری‌های کبدی با استفاده از افزایش داده مبتنی بر شبکه مولد تخصصی و یادگیری جمعی. مجله انفورماتیک سلامت و زیست پزشکی ۱۴۰۵؛ ۱۳(۱): ۲۳-۱۲.

مقدمه

هوش مصنوعی به عنوان یکی از فناوری های تحول آفرین و پیشرو در قرن بیست و یکم، نقش کلیدی در پیشرفت های علمی و صنعتی ایفا کرده است. در حوزه پزشکی و سلامت، کاربردهای این فناوری با سرعت چشمگیری در حال گسترش است و توانسته فرآیندهای تشخیصی و درمانی را به شکل قابل توجهی بهبود بخشد [۱-۳]. از مهم ترین کاربردهای هوش مصنوعی در این حوزه می توان به ثبت و تحلیل داده های بالینی، پیش بینی روند بیماری ها، تشخیص زودهنگام، پردازش تصاویر پزشکی از تصویربرداری تشدید مغناطیسی (MRI (Magnetic Resonance Imaging و توموگرافی کامپیوتری (CT (Computerized Tomography)، پایش علائم حیاتی، و مدیریت درمان اشاره کرد [۴-۶]. همچنین، پزشکی از راه دور و ربات های جراحی از دیگر زمینه های مهم کاربرد این فناوری به شمار می روند. با وجود این پیشرفت ها، چالش هایی نظیر کمبود داده های باکیفیت، حفظ حریم خصوصی بیماران، و پیچیدگی ذاتی داده های پزشکی، توسعه مدل های دقیق را با محدودیت مواجه کرده اند [۷].

در زمینه تشخیص بیماری ها، هوش مصنوعی نقش برجسته ای ایفا می کند؛ به ویژه در مورد بیماری های کبدی که از جمله بیماری های مزمن و تهدیدکننده سلامت جهانی محسوب می شوند. بیماری هایی نظیر سیروز و هپاتیت مزمن، در صورت عدم تشخیص به موقع، می توانند به عوارض جدی و حتی مرگ منجر شوند. الگوریتم های یادگیری ماشین و یادگیری عمیق با تحلیل داده های بالینی و شناسایی الگوهای پنهان، امکان تشخیص سریع تر و دقیق تر این بیماری ها را فراهم ساخته اند [۸،۹].

مطالعات متعددی به بررسی عملکرد مدل های مبتنی بر هوش مصنوعی در تشخیص بیماری ها، به ویژه بیماری های کبدی، پرداخته اند. برای نمونه، در مطالعه ای [۲]، مدلی مبتنی بر شبکه عصبی کانولوشنی (Convolutional Neural Network) ارائه شده است که در مقایسه با الگوریتم های سنتی نظیر نایو بیز (Naïve Bayes)، ماشین بردار پشتیبانی (Support Vector Machine)، نزدیک ترین همسایه (K-Nearest Neighbor) و رگرسیون لجستیک (Logestic Regression)، عملکرد قابل قبولی در طبقه بندی بیماری های کبدی نشان داده و صحت ۷۵/۵۵٪ و ۷۲٪ را به ترتیب برای مجموعه داده های شرکت خدمات مالی و مراقبت سلامت بریتانیایی BUPA و مجموعه داده بیماران کبدی هندوستان ILPD (Indian Liver Patient Dataset) به دست آورده است. همچنین، در پژوهش دیگری [۳]، با استفاده از تکنیک های پیش پردازش پیشرفته و الگوریتم های یادگیری تجمعی، موفق به افزایش صحت تشخیص بیماری کبد به بیش از ۹۱٪ شده اند. در مطالعه ای دیگر [۱۰]، با بهره گیری از مدل ترکیبی شبکه عصبی مصنوعی و الگوریتم جنگل تصادفی، صحت طبقه بندی ۹۸/۲۹٪، صحت کلی ۹۷/۴۳٪، نرخ بازخوانی (Recall) ۱۰۰٪ و امتیاز F1 معادل ۹۸/۶۹٪ گزارش شده است.

با این وجود، چالش های قابل توجهی در توسعه و کاربرد عملی سیستم های هوش مصنوعی در حوزه پزشکی همچنان به قوت خود باقی است [۱۱]. از جمله مهم ترین این چالش ها می توان به حفظ حریم خصوصی بیماران و نگرانی های حقوقی و اخلاقی مرتبط با استفاده از داده های حساس اشاره کرد [۱۲]. همچنین، محدودیت در دسترسی به داده های باکیفیت و حجم کافی، یکی دیگر از موانع عمده در آموزش مدل های دقیق و قابل اعتماد محسوب می شود. داده های ناکامل یا با تنوع پایین ممکن است موجب کاهش صحت مدل یا تولید نتایج نادرست گردند [۱۳-۱۵].

از سوی دیگر، اهمیت تشخیص زودهنگام بیماری ها، به ویژه بیماری های کبدی، در بهبود روند درمان و کاهش هزینه های مرتبط، به طور گسترده مورد تأکید قرار گرفته است. با این حال، وجود خطا در تشخیص ها نه تنها موجب افزایش هزینه های درمانی می شود، بلکه به تأخیر در روند درمان نیز منجر می گردد؛ که نیازمند بهبود روش های تشخیص می باشد.

شکاف اصلی پژوهش حاضر به سه مؤلفه اساسی مرتبط می شود: نخست، نبود مطالعه ای جامع و نظام مند در خصوص به کارگیری شبکه های مولد تخاصمی (GAN) تخصصی برای تولید داده های جدولی پزشکی در حوزه بیماری های کبدی؛ دوم، فقدان ارزیابی از تأثیر داده های مصنوعی باکیفیت حاصل از این مدل ها بر بهبود عملکرد روش های یادگیری جمعی، به ویژه در شناسایی کلاس اقلیت؛ و سوم، نبود چارچوبی یکپارچه که بتواند هم زمان سه چالش کمبود داده، نقض حریم خصوصی بیماران، و عدم تعادل شدید میان کلاس ها را برطرف سازد.

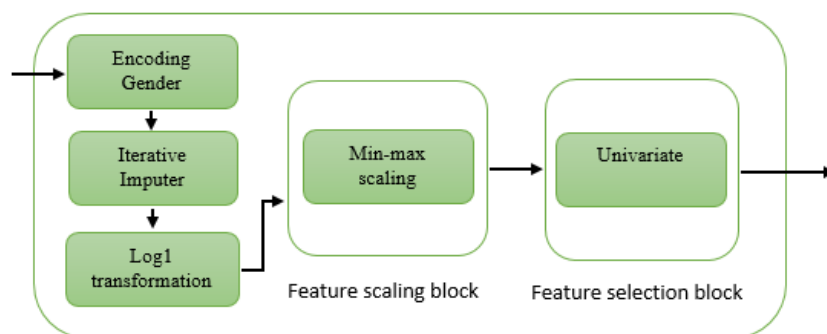
با وجود پیشرفت‌های قابل توجه در استفاده از یادگیری ماشین برای تشخیص بیماری‌ها، بسیاری از مدل‌های موجود به دلیل محدودیت حجم داده و ناتوانی در شناسایی دقیق الگوهای پیچیده، از دقت کافی برخوردار نیستند. همچنین، مسائل مرتبط با حفظ حریم خصوصی و حساسیت داده‌های بالینی، امکان استفاده گسترده از داده‌های واقعی را محدود کرده است [۱۶]. به همین دلیل، به کارگیری روش‌های افزایش داده نظیر شبکه‌های مولد تخصصی (GAN) که قادر به تولید داده‌های مصنوعی مشابه نمونه‌های واقعی هستند، می‌تواند به بهبود صحت مدل‌ها و رفع محدودیت‌های داده‌ای کمک کند. با این حال، تاکنون مطالعات محدودی (مانند [۵۶]) به بررسی کاربرد GAN در بهبود تشخیص بیماری‌های کبدی با داده‌های عددی پرداخته‌اند و این حوزه همچنان نیازمند تحقیقات بیشتر و توسعه مدل‌های پیشرفته‌تر است.

هدف پژوهش حاضر، توسعه مدلی است که با بهره‌گیری از تکنیک‌های پیش‌پردازش پیشرفته، افزایش داده مبتنی بر شبکه‌های مولد تخصصی (GAN) و الگوریتم یادگیری تجمعی استکینگ، بتواند صحت تشخیص بیماری کبد را به طور قابل توجهی افزایش دهد. این مدل با تمرکز بر داده‌های عددی حاصل از آزمایش‌های خون، درصدد جبران کمبود داده‌های واقعی و فراهم آوردن امکان تصمیم‌گیری دقیق‌تر و بهینه‌تر برای پزشکان برآمده است.

روش کار

شکل ۱ نمای کلی چارچوب روش پیشنهادی را نمایش می‌دهد که فرآیند پژوهش را در چند مرحله مشخص و منظم سامان‌دهی می‌کند. این ساختار گام‌به‌گام، به منظور طراحی مدلی دقیق و قابل اعتماد برای تشخیص بیماری کبد تدوین شده است.

- **انتخاب مجموعه داده اولیه:** انتخاب مجموعه داده‌ای استاندارد شامل اطلاعات بیماران مبتلا به بیماری‌های کبدی و افراد سالم به عنوان پایه‌ای برای آموزش مدل‌ها
- **آماده‌سازی و پیش‌پردازش داده‌ها:** پاک‌سازی، نرمال‌سازی و مقیاس‌بندی داده‌ها و سپس تقسیم آن‌ها به سه بخش آموزش، اعتبارسنجی و آزمون جهت ارزیابی علمی
- **متادل‌سازی داده‌ها با SMOTE:** استفاده از تکنیک Synthetic Minority Over-sampling Technique برای تولید نمونه‌های مصنوعی از کلاس اقلیت و رفع عدم تعادل
- **افزایش حجم داده با شبکه‌های مولد متخاصم (GAN):** تولید نمونه‌های مصنوعی مشابه داده‌های واقعی جهت کاهش خطر بیش‌برازش و ارتقای قابلیت تعمیم مدل
- **آموزش مدل با یادگیری جمعی (Ensemble Learning):** ارائه داده‌های غنی‌شده به الگوریتم‌های ترکیبی جهت افزایش دقت پیش‌بینی و پایداری [۱۷].



شکل ۱: نمای کلی چارچوب روش پیشنهادی

۱۰ ویژگی کلیدی که در مدل پیشنهادی به عنوان ورودی اصلی مدل های یادگیری ماشین به کار گرفته شده اند عبارتند از:

- سن (Age): عددی (۴ تا ۹۰ سال)
- جنسیت (Gender): مرد یا زن
- بیلی روبین کل (Total Bilirubin): شاخص سطح بیلی روبین در خون
- بیلی روبین مستقیم (Direct Bilirubin): بخش محلول بیلی روبین
- پروتئین کل (Total Proteins): مجموع پروتئین های موجود در سرم
- آلبومین (Albumin): پروتئینی مرتبط با عملکرد کبد
- آلکالین فسفاتاز (Alkaline Phosphatase): آنزیمی مرتبط با اختلالات کبد و استخوان
- آلانین آمینوترانسفراز (ALT): شاخص آسیب سلولی کبد
- آسپاراتات آمینوترانسفراز (AST): آنزیمی مکمل در ارزیابی عملکرد کبد
- نسبت آلبومین به گلوبولین (A/G Ratio): شاخص تعادل پروتئین های خون

آماده سازی و پیش پردازش داده

اقدامات اصلی در این مرحله به منظور سازگاری و بهبود کیفیت داده ها به شرح زیر انجام شد:

- **پاک سازی داده ها:** مشابه با مطالعه [۱۵]، مقادیر گم شده عمدتاً در ویژگی (A/G Ratio) با استفاده از میانگین یا میانه ستون مربوطه پر شدند (Imputation). مقادیر پرت (Outliers) با روش آماری IQR (Interquartile Range) شناسایی و مدیریت شدند تا نویز کاهش یابد.
- **تبدیل داده ها:** متغیرهای کیفی نظیر جنسیت با روش کدگذاری دودویی به مقادیر عددی تبدیل شدند (مرد = ۱، زن = ۰).
- **انتخاب ویژگی:** انتخاب ویژگی به صورت صریح انجام نشد و تمامی ۱۰ ویژگی به منظور جلوگیری از کاهش اطلاعات مفید بالینی حفظ شدند.
- **استاندارد سازی و مقیاس بندی:** استاندارد سازی با تبدیل ویژگی ها به توزیعی با میانگین صفر و انحراف معیار یک انجام شد. مقیاس بندی نیز با تکنیک Min-Max Scaler جهت انتقال مقادیر به بازه [۰,۱] یا [-۱,۱] صورت گرفت تا تفاوت های شدید مقیاس برطرف شود.
- **تبدیل لگاریتمی (log1p):** برای مقابله با توزیع بسیار نامتقارن متغیرهایی مانند آنزیم های کبدی، از تبدیل لگاریتمی استفاده شد تا توزیع به سمت نرمال نزدیک تر شود.

محیط پیاده سازی

تمامی مراحل پیاده سازی مدل ها و تحلیل داده ها با استفاده از زبان برنامه نویسی Python (نسخه ۳.۸ و بالاتر) انجام شد. برای پردازش و مدیریت داده ها از کتابخانه های Pandas و NumPy استفاده گردید. متعادل سازی داده ها با استفاده از کتابخانه imbalanced-learn و الگوریتم SMOTE انجام شد. آموزش و ارزیابی مدل های یادگیری ماشین با بهره گیری از کتابخانه scikit-learn صورت گرفت و پیاده سازی شبکه مولد متخاصم نیز با استفاده از چارچوب PyTorch انجام شد. تمامی آزمایش ها در محیط های Jupyter Notebook و Google Colab اجرا گردیدند.

افزایش داده با روش شبکه مولد متخاصم (GAN)

به منظور مواجهه با محدودیت های ذاتی داده های پزشکی جدولی، از جمله کمبود نمونه های واقعی، عدم توازن میان کلاس ها، و افزایش احتمال بیش برآزش در مدل های یادگیری ماشین، در این پژوهش از شبکه مولد متخاصم به عنوان یکی از راهبردهای افزایش داده استفاده شده است. شبکه های مولد متخاصمی از جمله روش های یادگیری عمیق هستند که با بهره گیری از چارچوبی رقابتی میان دو



شبکه، امکان تولید نمونه‌های مصنوعی واقع‌نما و آماری‌سازگار با داده‌های واقعی را فراهم می‌سازند. این ویژگی، GAN را به ابزاری مناسب برای کاربردهای پزشکی تبدیل می‌کند؛ زیرا در چنین داده‌هایی، دسترسی به حجم کافی از نمونه‌های متنوع و متوازن معمولاً دشوار است.

ساختار GAN از دو مؤلفه اصلی تشکیل می‌شود: مولد G و متمایزگر D. مولد با دریافت یک بردار نویز تصادفی Z از یک توزیع ساده، نمونه‌ای مصنوعی $x^* = G(z)$ تولید می‌کند؛ در مقابل، متمایزگر می‌کوشد بین نمونه‌های واقعی X و نمونه‌های تولیدشده توسط مولد تمایز قائل شود. فرایند آموزش این دو شبکه به‌صورت هم‌زمان و خصمانه انجام می‌شود؛ به‌گونه‌ای که مولد تلاش می‌کند متمایزگر را فریب دهد، در حالی که متمایزگر می‌کوشد نمونه‌های جعلی را شناسایی کند. در نتیجه این رقابت، کیفیت داده‌های مصنوعی به‌تدریج افزایش می‌یابد و مولد قادر می‌شود توزیعی نزدیک‌تر به داده‌های واقعی را بازتولید کند. تابع هدف کلاسیک GAN به‌صورت رابطه ۱ بیان می‌شود:

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(x)} [\log D(x)] + E_{z \sim p_z(z)} [\log (1 - D(G(z)))] \quad \text{رابطه ۱}$$

در این معادله، $p_{data}(x)$ توزیع داده‌های واقعی، $p_z(z)$ توزیع نویز ورودی، $D(x)$ احتمال واقعی بودن نمونه ورودی از دید متمایزگر، و $G(z)$ خروجی مولد است. هدف متمایزگر بیشینه‌سازی توان تفکیک میان نمونه‌های واقعی و مصنوعی، و هدف مولد کمینه‌سازی این تابع و در نتیجه افزایش شباهت داده‌های تولیدی به داده‌های واقعی است.

در این مطالعه، از GAN برای افزایش حجم داده، بهبود تنوع نمونه‌ها، و تقویت قابلیت تعمیم مدل نهایی استفاده شد. پس از تکمیل چرخه آموزش، حدود ۳۰۰۰ نمونه مصنوعی توسط شبکه مولد تولید و با داده‌های واقعی ادغام شد تا مجموعه داده آموزشی غنی‌تر و متوازن‌تری حاصل شود. این راهبرد نه تنها به کاهش اثر کمبود داده کمک کرد، بلکه به‌ویژه در شناسایی کلاس اقلیت، منجر به بهبود عملکرد مدل شد. بر اساس نتایج گزارش‌شده، استفاده از این داده‌های مصنوعی همراه با سایر مراحل پیش‌پردازش و متعادل‌سازی، نقش مهمی در ارتقای دقت، حساسیت، و شاخص‌های کلی عملکرد ایفا کرد.

به‌طور کلی، به‌کارگیری GAN در این پژوهش به‌منزله یک سازوکار داده‌افزایی مبتنی بر یادگیری عمیق، امکان تولید نمونه‌های مصنوعی معتبر و مفید را فراهم ساخت و در نهایت به افزایش پایداری و کارایی سامانه طبقه‌بندی کمک کرد.

روش پشته‌سازی در الگوریتم یادگیری جمعی

در روش پشته‌سازی (Stacking Ensemble)، فرایند آموزش در دو مرحله انجام می‌شود. ابتدا چندین مدل پایه (مانند درخت تصمیم، رگرسیون لجستیک، SVM و KNN) با داده‌ها آموزش می‌بینند. سپس خروجی این مدل‌ها به یک مدل نهایی (Meta-Learner) داده می‌شود. مدل نهایی یاد می‌گیرد که در چه شرایطی کدام مدل پایه عملکرد بهتری دارد و بر اساس آن، وزن‌های متفاوتی به پیش‌بینی‌ها اختصاص می‌دهد که این امر دقت و پایداری را به میزان قابل‌توجهی افزایش می‌دهد [۱۶، ۱۷].

ارزیابی مدل

در این بخش، چارچوب و محیط آزمایشگاهی مورد استفاده برای سنجش عملکرد مدل‌های پیشنهادی شرح داده می‌شود. این موارد شامل تشریح مجموعه داده‌های به کار رفته در فاز آزمون، معرفی تکنیک اعتبارسنجی (Cross-Validation) و همچنین تعریف دقیق معیارهای ارزیابی مدل (از جمله دقت، صحت، فراخوانی و امتیاز F1) می‌باشد.

معرفی مجموعه داده

در این پژوهش، از مجموعه داده ILPD استخراج‌شده از مخزن داده‌های یادگیری ماشین UCI Machine Learning Repository استفاده شده است. این مجموعه داده شامل اطلاعات ۵۸۳ نمونه (۴۱۶ بیمار و ۱۶۷ فرد سالم؛ ۴۴۱ مرد و ۱۴۲ زن) می‌باشد. انتخاب این مجموعه داده به‌صورت آگاهانه صورت گرفته است؛ چراکه این دیتاست، پرکاربردترین مجموعه داده عمومی در مطالعات تشخیص بیماری کبدی محسوب می‌شود (بیش از ۱۵۰ مقاله معتبر) و چالش‌های واقعی داده‌های بالینی نظیر حجم کم،

عدم تعادل شدید کلاس‌ها (۴/۷۱٪ بیمار در برابر ۶/۲۸٪ سالم)، مقادیر گمشده، و توزیع نامتقارن را به خوبی بازتاب می‌دهد. آمار توصیفی کلیدی ویژگی‌های بالینی بر اساس تحلیل استاندارد مجموعه داده UCI، در جدول ۱ ارائه شده است.

جدول ۱: آمار توصیفی کلیدی ویژگی‌های بالینی

ویژگی	میانگین (Mean)	انحراف معیار (Std)	حداقل (Min)	حداکثر (Max)
سن	۴۴/۷۵	۱۶/۱۶	۴	۹۰
جنسیت	۰/۷۵۶	۰/۴۲۹	۰	۱
بیلی روبین کل	۳/۳	۶/۲	۰/۴	۷۵
بیلی روبین مستقیم	۱/۵	۲/۸	۰/۱	۱۹/۱
آلکالین فسفاتاز	۲۹۰/۶	۳۴۲/۹	۶۳	۲۱۱۰
آلاتین آمینوترانسفراز	۸۰/۸	۱۸۲/۶	۱۰	۲۰۰۰
آسپاراتات آمینوترانسفراز	۱۰۹/۹	۲۸۸/۹	۱۰	۴۹۳۹
پروتئین کل	۶/۵	۱/۱	۲/۷	۹/۶
آلبومین	۳/۱	۰/۸	۰/۹	۵/۵
نسبت آلبومین به گلوبولین (A/G)	۰/۹۵	۰/۳۵	۰/۳	۲/۸

ارزیابی عملکرد مدل با استفاده از روش‌های تحلیلی

پس از اتمام فرایند آموزش مدل، ارزیابی دقیق عملکرد آن برای اطمینان از کارایی و قابلیت تعمیم‌پذیری، امری ضروری است. در این راستا، دو ابزار رایج و کاربردی مورد استفاده قرار می‌گیرند: ماتریس درهم‌ریختگی و اعتبارسنجی متقابل، ماتریس درهم‌ریختگی یکی از ابزارهای تحلیلی قدرتمند در حوزه یادگیری نظارت‌شده است که عملکرد مدل را با دقت بالایی مورد سنجش قرار می‌دهد. این ماتریس با نمایش تفکیکی تعداد پیش‌بینی‌های صحیح و نادرست، امکان بررسی دقیق نتایج مدل را فراهم می‌سازد. در ساختار این ماتریس، ردیف‌ها نشان‌دهنده برچسب‌های واقعی و ستون‌ها بیانگر خروجی‌های پیش‌بینی‌شده توسط مدل هستند. از دل این ماتریس، می‌توان معیارهای کلیدی ارزیابی مدل از جمله موارد زیر را استخراج نمود:

- ❖ **صحت (Accuracy):** همان طور که در رابطه ۲ مشاهده می‌نمایید صحت نسبت پیش‌بینی‌های صحیح به کل نمونه‌ها،
- ❖ **بازخوانی (Recall):** توانایی مدل در شناسایی درست نمونه‌های مثبت، یعنی چه کسری از کل موارد مثبت واقعی، توسط مدل به درستی شناسایی شده‌اند. (مطابق رابطه ۳)،
- ❖ **دقت (Precision):** توانایی مدل در پیش‌بینی صحیح موارد مثبت، یعنی چه کسری از مواردی که مدل به عنوان مثبت پیش‌بینی کرده، واقعاً مثبت بوده‌اند. (مطابق رابطه ۴)،
- ❖ **معیار F1 (F1-Score):** میانگین همساز دقت و بازخوانی (مطابق رابطه ۵)، را استخراج و تحلیل کرد. این شاخص‌ها تصویر روشنی از عملکرد مدل در شرایط مختلف ارائه می‌دهند و پایه‌ای برای بهینه‌سازی مدل در مراحل بعدی فراهم می‌سازند.

$$\text{Accuracy} = \frac{|TN| + |TP|}{|TN| + |TP| + |FN| + |FP|} \quad \text{رابطه ۲}$$

$$\text{Recall} = \frac{|TN|}{|TN| + |FP|} \quad \text{رابطه ۳}$$

$$\text{Precision} = \frac{|TP|}{|TP| + |FP|} \quad \text{رابطه ۴}$$

$$F1\text{-measure} = \frac{2 \times |TP|}{2 \times |TP| + |FN| + |FP|} \quad \text{رابطه ۵}$$

اعتبارسنجی متقابل؛ سنجشی دقیق در شرایط داده‌های محدود

در فرآیند ارزیابی مدل‌های یادگیری ماشین، اعتبارسنجی متقابل (Cross-Validation) به‌عنوان یکی از روش‌های مؤثر بازنمونه‌گیری، نقشی کلیدی ایفا می‌کند، به‌ویژه در شرایطی که حجم داده‌ها محدود یا تنوع آن‌ها کم است. در رویکرد سنتی، داده‌ها به دو بخش آموزش و آزمون تقسیم می‌شوند؛ مدلی بر پایه داده‌های آموزشی ساخته شده و سپس عملکرد آن با داده‌های آزمون ارزیابی می‌شود. با این حال، این روش ممکن است منجر به نتایجی سوگیرانه و غیرقابل تعمیم شود، چراکه ارزیابی تنها بر اساس یک بخش خاص از داده‌ها انجام می‌گیرد که ممکن است به‌طور کامل نماینده کل توزیع داده‌ها نباشد.

در مقابل، اعتبارسنجی متقابل با تقسیم مجموعه داده به چند زیرمجموعه (Fold) و اجرای چندباره فرآیند آموزش و ارزیابی روی بخش‌های مختلف داده، امکان برآورد دقیق‌تر و پایدارتر عملکرد مدل را فراهم می‌سازد. در هر تکرار، یکی از زیرمجموعه‌ها به‌عنوان داده آزمون در نظر گرفته می‌شود و باقی داده‌ها برای آموزش مورد استفاده قرار می‌گیرند. در پایان، میانگین نتایج حاصل از تکرارهای مختلف، شاخصی معتبر و قابل اتکا از عملکرد کلی مدل ارائه می‌دهد. این روش نه تنها به کاهش عدم قطعیت ناشی از انتخاب تصادفی داده‌های آزمون کمک می‌کند، بلکه در طراحی مدل‌هایی با قابلیت تعمیم بالا نیز نقش تعیین‌کننده‌ای دارد.

نتایج

نتایج حاصل از این ارزیابی بر روی داده‌های واقعی، در جدول‌های ۲ و ۳ ارائه شده است. بر اساس اطلاعات جدول ۲، مدل پیشنهادی در تشخیص بیماران کبدی عملکرد بسیار موفق داشته است (Precision = ۰/۷۸)، (Recall = ۰/۹۵) با این حال، در شناسایی افراد سالم با ضعف مواجه است (Recall = ۰/۲۳) که نشان‌دهنده نیاز به بهبود عملکرد در این کلاس می‌باشد. برای تحلیل دقیق‌تر، داده‌های جدول ۳ مورد بررسی قرار می‌گیرد. ماتریس درهم‌ریختگی عملکرد مدل را در تشخیص کلاس‌های ۰ (افراد سالم) و ۱ (بیماران کبدی) نمایش می‌دهد. این ماتریس اطلاعات دقیقی درباره تعداد پیش‌بینی‌های صحیح و اشتباه مدل ارائه می‌کند. بر اساس جدول مربوط به مدل پیشنهادی، نتایج به‌شرح زیر است:

هفت فرد سالم را به‌درستی سالم تشخیص داده است. بیست و سه فرد سالم را به‌اشتباه بیمار تشخیص داده است. چهار بیمار را به‌اشتباه سالم تشخیص داده است. ۸۳ مورد، یعنی مدل ۸۳ بیمار را به‌درستی شناسایی کرده است.

جدول ۲: عملکرد مدل با استفاده از صرفاً داده واقعی

معیار	کلاس صفر (منفی)	کلاس ۱ (مثبت)	کلی
دقت	۰/۶۴	۰/۷۸	-
بازخوانی	۰/۲۳	۰/۹۵	-
معیار F1	۰/۳۴	۰/۸۶	-
تعداد نمونه‌ها	۳۰	۸۷	۱۱۷
صحت	-	-	۰/۷۷
میانگین کلان	۰/۷۱	۰/۵۹	۰/۶

جدول ۳: ماتریس در هم ریختگی داده‌های واقعی

واقعی / پیش بینی شده	کلاس +	کلاس ۱
کلاس ۰ (واقعی)	۷	۲۳
کلاس ۱ (واقعی)	۴	۸۳

سپس به منظور بهبود مدل با اعمال تکنیک SMOTE برای متعادل سازی کلاس‌ها، صحت کلی به حدود ۸۴ درصد افزایش یافت و بازخوانی کلاس اقلیت (افراد سالم) بهبود قابل توجهی پیدا کرد، اما همچنان عملکرد مدل در داده‌های واقعی محدود بود. در مرحله نهایی، با تولید حدود ۳۰۰۰ نمونه مصنوعی توسط شبکه مولد تخصصی (GAN) و افزودن آن‌ها به مجموعه آموزشی، همراه با یادگیری جمعی استکینگ، بهبود قابل توجهی در نتایج بدست آمده مطابق جدول ۴ و جدول ۵ ایجاد شد. همان طور که در جدول ۴ مشاهده می گردد در داده‌های مصنوعی تولیدشده توسط GAN به ۹۹ درصد افزایش یافت؛ در این مرحله، بازخوانی و ویژگی برای هر دو کلاس (بیمار و سالم) به طور چشمگیری بهبود یافت (به ترتیب نزدیک به ۹۸-۱۰۰ درصد) و AUC-ROC به حدود ۰/۹۹۹ رسید. این نتایج نشان می‌دهند که هرچند مدل پیشنهادی در شناسایی بیماران صحت بالایی دارد، اما برای افزایش توان تفکیک افراد سالم، به بهبود داده‌های آموزشی، تکنیک‌های نمونه‌سازی یا تنظیم بهتر آستانه تصمیم نیاز است.

این یافته‌ها حاکی از آن است که شبکه مولد متخصص توانسته است داده‌هایی با کیفیت بالا و ساختاری مشابه با توزیع داده‌های واقعی تولید کند.

- **Precision:** مقدار ۰/۹۷ برای کلاس «افراد سالم» و ۱/۰۰ برای کلاس «بیماران کبدی» بیانگر آن است که مدل تقریباً تمامی پیش‌بینی‌های مثبت خود را به درستی انجام داده و میزان خطای نوع اول بسیار پایین بوده است.
- **Recall:** مقدار ۱/۰۰ برای افراد سالم نشان‌دهنده شناسایی کامل این گروه توسط مدل است، به طوری که هیچ نمونه‌ای به اشتباه در گروه بیماران قرار نگرفته است. مقدار ۰/۹۸ برای بیماران کبدی نیز بیانگر عملکرد دقیق مدل در شناسایی این گروه بوده و تنها ۲٪ خطای نوع دوم گزارش شده است.
- **F1-Score:** مقدار ۰/۹۹ برای هر دو کلاس، تعادل مطلوب بین دقت (Precision) و بازخوانی (Recall) را تأیید می‌کند و نشان‌دهنده عملکرد بسیار پایدار مدل در هر دو کلاس است.
- **Accuracy:** نرخ صحت کلی مدل برابر با ۰/۹۹ (۹۹٪) است که عملکرد بسیار بالایی مدل در داده‌های تولیدشده را نشان می‌دهد.

جدول ۴: عملکرد مدل با استفاده بعد از تولید داده مصنوعی

معیار	کلاس صفر (منفی)	کلاس ۱ (مثبت)	کلی
دقت	۰/۹۷	۱	۰/۹۹
بازیابی	۱	۰/۹۸	۰/۹۹
معیار F1	۰/۹۹	۰/۹۹	۰/۹۹
تعداد نمونه‌ها	۲۷۰	۳۳۰	۶۰۰
صحت	-	-	۰/۹۹

جدول ۵: ماتریس در هم ریختگی بعد از تولید داده مصنوعی

واقعی / پیش بینی شده	کلاس ۰	کلاس ۱
کلاس ۰ (واقعی)	۲۶۹	۱
کلاس ۱ (واقعی)	۷	۳۲۳

ماتریس درهم ریختگی روی داده های واقعی آزمون نشان داد که از ۸۷ نمونه بیمار، ۸۳ مورد به درستی شناسایی شدند و تنها ۴ مورد به اشتباه سالم تشخیص داده شدند، در حالی که در میان ۳۰ نمونه سالم، ۲۳ مورد به اشتباه بیمار تشخیص داده شدند که این مشکل مطابق جدول ۶ با داده های مصنوعی تقریباً به طور کامل برطرف شد. همان طور که در جدول ۶ مشاهده می شود، مدل پیشنهادی ۲۶۹ فرد سالم را به درستی سالم تشخیص داده است که نشان دهنده صحت بالای مدل در شناسایی افراد سالم است. ۳۲۳ بیمار کبدی را به درستی شناسایی کرده است که نشان دهنده عملکرد بسیار خوب در تشخیص بیماران است و تنها یک فرد بیمار را به اشتباه سالم تشخیص داده است که مقدار بسیار کمی محسوب می شود و عملکرد عالی مدل را در تشخیص بیماران نشان می دهد. در نهایت مدل ۷ فرد سالم را به اشتباه بیمار تشخیص داده است که نسبت به تعداد کل نمونه ها مقدار کمی است و تأثیر زیادی بر عملکرد کلی ندارد.

بحث و نتیجه گیری

هدف این پژوهش، ارائه چارچوبی کارآمد برای تشخیص زود هنگام بیماری های کبدی بر پایه داده های آزمایشگاهی بود. با توجه به محدودیت حجم داده ها در مجموعه داده ILPD، عدم تعادل شدید کلاس ها، و نامتوازن بودن توزیع ویژگی ها، به کارگیری پیش پردازش مناسب، متعادل سازی داده ها با SMOTE، و به ویژه افزایش داده ها با شبکه مولد تخصصی (GAN) توانست عملکرد مدل را به طور چشمگیری بهبود بخشد. در این مطالعه، مدل استکینگ متشکل از رگرسیون لجستیک، ماشین بردار پشتیبان (SVM) و AdaBoost، پس از غنی سازی داده ها، به دقتی معادل ۹۹٪ دست یافت، در حالی که این مقدار بر روی داده های واقعی حدود ۷۷٪ بود. همچنین، بهبود قابل توجهی در شناسایی کلاس اقلیت مشاهده شد؛ به طوری که نرخ بازخوانی (Recall) از ۰/۲۳ به ۱/۰۰ و امتیاز F1-Score به ۰/۹۹ افزایش یافت. این نتایج نشان می دهد که تولید نمونه های مصنوعی با کیفیت می تواند نقش مؤثری در ارتقای عملکرد مدل های تشخیصی، به ویژه در مجموعه داده های پزشکی نامتعادل، ایفا کند.

یافته های این پژوهش با مطالعات پیشین همسو می باشد. برای نمونه، در مطالعه [۱۸] نیز استفاده از GAN برای افزایش داده های حاصل از تصاویر CT ضایعات کبدی، موجب بهبود نرخ بازخوانی و ویژگی طبقه بندی شد. به طور کلی، نتایج پژوهش های پیشین نشان می دهند که افزایش داده، به ویژه در داده های پزشکی محدود و نامتعادل، می تواند تعمیم پذیری مدل را افزایش داده و خطر بیش برآزش را کاهش دهد [۱۹-۲۱]. با این حال، استفاده از GAN در داده های جدولی همچنان با چالش هایی نظیر وابستگی به کیفیت داده های ورودی، نیاز به تنظیم دقیق فرآیندها و احتمال محدود بودن تنوع داده های تولید شده همراه است. این محدودیت ها نشان می دهد که توسعه روش های تخصصی تر برای تولید داده های ساختاریافته، به ویژه در کاربردهای بالینی، همچنان ضروری می باشد.

از منظر کاربردی، نتایج این مطالعه ظرفیت بالای ترکیب یادگیری جمعی و افزایش داده مبتنی بر GAN را برای استفاده در سامانه های پشتیبان تصمیم گیری بالینی نشان می دهد. با این حال، به منظور افزایش اعتبار و تعمیم پذیری نتایج، پیشنهاد می شود در پژوهش های آتی از مجموعه داده های بزرگ تر و چندمرکزی استفاده شود، الگوریتم های GAN پیشرفته تر نظیر Conditional GAN مورد ارزیابی قرار گیرند، و رویکردهای چندوجهی مبتنی بر داده های آزمایشگاهی و تصویری مانند CT و MRI نیز بررسی شوند. همچنین، گسترش حجم داده های مصنوعی و مقایسه ترکیب های مختلف الگوریتم های یادگیری ماشین می تواند به شناسایی پیکربندی های بهینه تر برای تشخیص بیماری های کبدی کمک کند.

در مجموع، این پژوهش نشان داد که بهره گیری از داده های مصنوعی تولید شده با GAN در کنار یادگیری جمعی می تواند عملکرد تشخیص زود هنگام بیماری کبد را به طور معناداری بهبود دهد و ظرفیت مناسبی برای توسعه ابزارهای هوشمند در انفورماتیک پزشکی فراهم آورد.

تعارض منافع

نویسندگان مقاله تأیید می‌نمایند که در این مقاله تعارض منافع وجود ندارد.

حمایت مالی

این مقاله در قالب مقاله مستخرج از پایان‌نامه کارشناسی ارشد در دانشکده علوم و فناوری‌های نوین تهیه شده است و فاقد حمایت مالی می‌باشد.

اعلام استفاده از هوش مصنوعی (AI)

در فرآیند آماده‌سازی مقاله از ابزار هوش مصنوعی Gemini Pro صرفاً به منظور بهبود نگارش، روان‌سازی بخش‌هایی از مقاله استفاده شده است. خروجی این ابزار توسط گروه نویسندگان به دقت بررسی شده است. طراحی پژوهشی، کد نویسی، آموزش و مدل پیشنهادی و مراحل مربوط به نگارش و تفسیر نتایج توسط گروه نویسندگان صورت گرفته است.

کد اخلاق

این مطالعه دارای مجوز کد اخلاق دانشگاه علوم پزشکی کرمان با کد IR.KMU.REC.1405.139 می‌باشد.

سهم مشارکت نویسندگان

تمامی نویسندگان در بخش‌های مختلف مقاله مشارکت داشته‌اند و از سهم مساوری برخوردار می‌باشند.

References

- [1]. Md AQ, Kulkarni S, Joshua CJ, Vaichole T, Mohan S, Iwendi C. Enhanced preprocessing approach using ensemble machine learning algorithms for detecting liver disease. *Biomedicines* 2023;11(2):581. <https://doi.org/10.3390/biomedicines11020581>
- [2]. Anuradha C, Swapna D, Thati B, Sree VN, Praveen SP. Diagnosing for liver disease prediction in patients using combined machine learning models. In 2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT); 2022 Jan 20; IEEE; 2022. p. 889-96.
- [3]. Dhiraj D, Patle G, Meshram E. Designing Disease Prediction Model Using Machine Learning Approach. *International Conference on Computing Methodologies and Communication*; 2022 Jan 20-22; Tirunelveli, India: IEEE; 2019. doi: [10.1109/ICSSIT53264.2022.9716312](https://doi.org/10.1109/ICSSIT53264.2022.9716312)
- [4]. AlAmir M, AlGhamdi M. The role of generative adversarial network in medical image analysis: An in-depth survey. *ACM Computing Surveys* 2022;55(5):1-36.
- [5]. Shorten C, Khoshgoftaar TM. A survey on image data augmentation for deep learning. *Journal of big data*. 2019;6(1):1-48.
- [6]. Chlap P, Min H, Vandenberg N, Dowling J, Holloway L, Haworth A. A review of medical image data augmentation techniques for deep learning applications. *Journal of Medical Imaging and Radiation Oncology* 2021;65(5):545-63. <https://doi.org/10.1111/1754-9485.13261>
- [7]. Abdollahi J, Nouri-Moghaddam B, Ghazanfari M. Deep Neural Network Based Ensemble learning Algorithms for the healthcare system (diagnosis of chronic diseases). *arXiv preprint arXiv:2103.08182*.
- [8]. Ghorbani R, Ghousi R. Predictive data mining approaches in medical diagnosis: A review of some diseases prediction. *International Journal of Data and Network Science* 2019;3(2):47-70.
- [9]. Alauthman M, Aldweesh A, Al-qerem A, Aburub F, Al-Smadi Y, Abaker AM, et al. Tabular data generation to improve classification of liver disease diagnosis. *Appl Sci* 2023;13(4):2678. <https://doi.org/10.3390/app13042678>
- [10]. Mutlu EN, Devim A, Hameed AA, Jamil A. Deep learning for liver disease prediction. In *Mediterranean conference on pattern recognition and artificial intelligence*. Cham: Springer International Publishing; 2021. p. 95-107.
- [11]. Nazari Nezhad S, Zahedi MH, Farahani E. Detecting diseases in medical prescriptions using data mining methods. *BioData Mining* 2022;15(1):29.



- [12]. Mahajan P, Uddin S, Hajati F, Moni MA. Ensemble learning for disease prediction: a review. Healthcare (Basel) 2023;11(12):1808. doi: [10.3390/healthcare11121808](https://doi.org/10.3390/healthcare11121808)
- [13]. Choudhary R, Gopalakrishnan T, Ruby D, Gayathri A, Murthy VS, Shekhar R. An efficient model for predicting liver disease using machine learning. Data analytics in bioinformatics: A Machine Learning Perspective 2021:443-57. <https://doi.org/10.1002/9781119785620.ch18>
- [14]. Bhusnurmath RA, Betageri S. Deep Ensemble Stacked Technique for the Classification of Liver Disease Using Artificial Neural Networks at the Base Level and Random Forest at the Meta Level. International Journal of Computer Applications 2023;975:8887.
- [15]. Rahman AS, Shamrat FJ, Tasnim Z, Roy J, Hossain SA. A comparative study on liver disease prediction using supervised machine learning algorithms. International Journal of Scientific & Technology Research 2019;8(11):419-22.
- [16]. Assegie TA, Subhashni R, Kumar NK, Manivannan JP, Duraisamy P, Engidaye MF. Random forest and support vector machine based hybrid liver disease detection. Bulletin of Electrical Engineering and Informatics 2022;11(3):1650-6. doi: <https://doi.org/10.11591/eei.v11i3.3787>
- [17]. Yousefpour H, Ghasemi J. Ensemble-based detection and Classification of liver diseases Caused by hepatitis C. Contributions of Science and Technology for Engineering 2024;1(1):32-42. doi: [10.22080/cste.2024.5012](https://doi.org/10.22080/cste.2024.5012)
- [18]. Frid-Adar M, Klang E, Amitai M, Goldberger J, Greenspan H. Synthetic data augmentation using GAN for improved liver lesion classification. 15th International Symposium on Biomedical Imaging; 2018 Apr 4; IEEE; 2018. p. 289-93
- [19]. Althnian A, AlSaeed D, Al-Baity H, Samha A, Dris AB, Alzakari N, et al. Impact of dataset size on classification performance: an empirical evaluation in the medical domain. Applied Sciences 2021;11(2):796.
- [20]. Tahmasbi H, Besharati R, Alishahi M. A Model For Diagnosing Liver Diseases Using “Machine Learning” Techniques. Studies in Medical Sciences 2022;33(11):814-22. [In Persian] [doi:10.52547/umj.33.11.814](https://doi.org/10.52547/umj.33.11.814)
- [21]. Javadzadeh S, Shayanfar H, Soleimanian QF. A hybrid model based on ant lion optimization algorithm and K-Nearest neighbors Algorithm to Diagnose Liver Disease. Journal of Ilam University of Medical Sciences 2021; 28(5): 76-89. [In Persian]