

## تشخیص بیماری کبد با استفاده از مدل جنگل تصادفی مبتنی بر الگوریتم بهینه‌سازی خرچنگ

نازنین صالح امین و فرهاد سلیمانیان قره چپق\* و ندا محمدی کیا

### چکیده

بیماری‌های کبدی سبب آسیب به عملکرد طبیعی کبد می شوند، تشخیص زود هنگام می‌تواند در پیشگیری، کنترل و درمان موثر باشد. در این مقاله یک مدل نوین بر مبنای بهینه‌سازی مدل جنگل تصادفی با الگوریتم خرچنگ برای تشخیص بیماری کبد پیشنهاد شده است. برای ارزیابی از مجموعه داده بیماران کبدی با ۴۱۶ مبتلا، ۱۶۷ نفر سالم و ۱۰ ویژگی برای هر بیمار استفاده شده است. نتایج نشان می‌دهد که عملکرد روش پیشنهادی در مقایسه با پنج مدل یادگیری ماشین از نظر معیارهای صحت، دقت، فراخوانی و معیار  $F_1$  ۱ دقیق‌تر بوده و به ترتیب برابر ۰/۷۳۵۶، ۰/۷۳۴۱، ۰/۸۴۶۷ و ۱/۰۰۰۰ بدست آمده که در تشخیص و طبقه‌بندی افراد سالم و مبتلا نقشی موثر داشته و برای بالا بردن میزان تشخیص، به حداقل رساندن درصد خطا و کاهش هزینه‌ها و زمان به عنوان یک ابزار کمکی قابل استفاده است. در آینده با داده‌های بیشتر و بهبود مدل‌ها می‌توان دقت و اطمینان را افزایش داد و گام مهمی در جهت پزشکی هوشمند داشت.

کلیدواژه: بیماری کبد، هوش مصنوعی، جنگل تصادفی، الگوریتم بهینه‌سازی خرچنگ

کبد از نظر اندازه و وزن بزرگترین اندام داخلی بدن انسان است، که بین ۲ تا ۳ درصد وزن بدن را تشکیل می‌دهد [۱]. مسئولیت بسیاری از عملکردها، از جمله سنتز صفرا تا تجزیه و پردازش سموم در خون برعهده کبد است. صفرا یک ترشح ضروری برای بقای انسان است، زیرا محیط اولیه برای دفع کلسترول و سایر مواد سمی است که کلیه‌ها قادر به پردازش آنها نیستند [۲]. می‌توان نتیجه گرفت که عملکرد مؤثر و جامع کبد لازمه یک زندگی سالم است. به عنوان یک عضو ضروری، تعداد زیادی از ویروس‌ها برای آلوده کردن کبد انسان تکامل یافته‌اند. شایع‌ترین آنها ویروس‌های هپاتیت نامیده می‌شوند که نام آن‌ها برگرفته از شرایطی است که منجر به ایجاد التهاب کبد می‌شوند. بیماری‌های غیر عفونی مختلف مانند کلانژیت صفراوی اولیه<sup>۱</sup> و بیماری کبد چرب غیرالکلی<sup>۲</sup> نیز بر کبد تأثیر می‌گذارد.

هوش مصنوعی حوزه وسیعی است که در زمینه‌های مختلف ریاضیات و علوم کاربرد دارد. تشخیص هر گونه بیماری و پیشگیری از شیوع آن مستلزم بررسی مداوم داده‌ها است. اقدام سریع براساس داده‌های دقیق تأثیر اجتماعی و مالی قابل توجهی بر زندگی مردم در سراسر جهان دارد [۳]. استفاده از هوش مصنوعی در مراقبت‌های بهداشتی، جمع‌آوری و پردازش داده‌های ارزشمند و در سطوح بالاتر، برنامه‌ریزی تشخیص بیماری را بهبود داده است [۴]. هوش مصنوعی در حوزه مراقبت‌های بهداشتی نحوه جمع‌آوری، تجزیه و تحلیل و توسعه اطلاعات برای مراقبت از بیماران را تغییر داده است [۵].

الگوریتم جنگل تصادفی یک روش هوش مصنوعی و از گروه یادگیری آماری است که توسط لئو بریمن در سال ۲۰۰۱ معرفی شده است [۶]. الگوریتم جنگل تصادفی به طور گسترده در بسیاری از زمینه‌های کاربردی استفاده شده است، البته نه تنها به دلیل عملکرد پیش‌بینی دقیق، بلکه به دلیل انعطاف‌پذیری این الگوریتم می‌باشد. در نتیجه، الگوریتم جنگل تصادفی یکی از روش‌های ترجیحی فعالان حوزه آمار، داده و سایر محققان داده است. مهمترین رکن برای افزایش دقت الگوریتم جنگل تصادفی تعیین مقدار دقیق پارامترهای این الگوریتم توسط الگوریتم‌های فراابتکاری می‌باشد. الگوریتم‌های فراابتکاری به توصیف اکتشافی راه حل‌های بهینه می‌پردازند که برای حل طیف وسیعی از مسائل استفاده شده‌اند [۷]. اثرگذاری استفاده از این الگوریتم‌ها برای حل مسائل پیچیده، به دست آمدن بهترین راه حل‌ها حتی برای داده‌های بسیار بزرگ در مدت زمان کمی می‌باشد. الگوریتم‌های فراابتکاری یک راه حل عملی و دقیق برای بسیاری از مسائل تک‌هدفه و چندهدفه ارائه می‌دهند و برای دستیابی به راه حل‌های تقریبی/بهینه در زمان‌های معقول برای مسائل بهینه‌سازی NP-Hard بهره گرفته شده‌اند [۸].

الگوریتم بهینه‌سازی خرچنگ<sup>۳</sup> یکی از الگوریتم‌های فراابتکاری می‌باشد که رفتار خرچنگ‌ها در پناه گرفتن در یک مکان خنک، رقابت و جستجوی غذا را شبیه‌سازی کرده است [۹]. مرحله پناه گرفتن در یک مکان خنک نشان‌دهنده اکتشاف است، درحالی‌که مرحله رقابت و جستجوی غذا نشان‌دهنده مرحله بهره‌برداری هستند. اکتشاف و بهره‌برداری در الگوریتم بهینه‌سازی خرچنگ توسط دما تنظیم می‌شوند. زمانی که دما بیش از حد بالا باشد، خرچنگ‌ها برای پناه گرفتن به غارها می‌روند یا برای تصاحب غار با یکدیگر رقابت می‌کنند. زمانی که دما مناسب باشد، خرچنگ‌ها بر مبنای اندازه غذا، رفتارهای مختلفی در جستجوی غذا نشان می‌دهند. الگوریتم پیشنهادی در این مقاله، یک روش بهینه‌سازی نوآورانه و الهام گرفته از رفتارهای طبیعی خرچنگ‌ها است که پارامترهای الگوریتم جنگل تصادفی را برای مسئله طبقه‌بندی

<sup>۱</sup> Primary Biliary Cholangitis

<sup>۲</sup> Non-alcoholic fatty liver disease (NAFLD)

<sup>۳</sup> Crayfish optimization algorithm

بیماران کبدی بهینه‌سازی می‌نماید. الگوریتم بهینه‌سازی خرچنگ بهترین پارامترها را با توجه به نتایج ارزیابی شده در طول تکرارها برای تشخیص بیماری کبدی شناسایی می‌کند.

ساختار این مقاله بشرح زیر صورت است: در بخش دوم، مطالعات پیشین در حوزه تشخیص بیماری کبد بررسی می‌شوند. در بخش سوم، روش پیشنهادی، پیش‌پردازش داده‌ها، جنگل تصادفی، الگوریتم بهینه‌سازی خرچنگ، معیارهای ارزیابی و مراحل پیاده‌سازی روش پیشنهادی بیان می‌شوند. در بخش چهارم، مقایسه و ارزیابی روش پیشنهادی انجام می‌شود. نهایتاً در بخش پنجم نتیجه‌گیری و کارهای آینده ارائه خواهد شد.

تکنیک‌های داده‌کاوی شامل ماشین بردار پشتیبان، CHAID، Exhaustive CHAID و نسخه تقویت‌شده الگوریتم C5.0 برای تشخیص بیماری کبد پیشنهاد شده‌اند [۱۰]. این مطالعه از نوع توصیفی است و با بهره‌گیری از ۷۲۱ داده جمع‌آوری شده از بیماران مبتلا به بیماری کبد در شهر زاهدان انجام شده است. نتایج حاصل از تحلیل‌ها نشان می‌دهد که بالاترین دقت مربوط به الگوریتم C5.0 تقویت‌شده با ۹۴.۰۹ درصد است. الگوریتم Exhaustive CHAID با دقت ۸۸.۷۱ درصد، ماشین بردار پشتیبان با دقت ۸۷.۰۹ درصد و الگوریتم CHAID با دقت ۸۵.۴۷ درصد در رده‌های بعدی قرار دارند.

الگوریتم بهینه‌سازی ازدحام ذرات و سیستم استنتاج عصبی-فازی تطبیقی برای شناسایی اختلالات کبدی در افراد سالم و بیمار استفاده شده است [۱۱]. داده‌ها از سایت معتبر UCI جمع‌آوری شدند. به‌منظور ارزیابی روش پیشنهادی، معیارهایی همچون دقت، حساسیت و صحت استفاده شده‌اند. روش پیشنهادی با دقت میانگین ۹۹.۱۴ درصد، نتایج حاصل از این مدل نشان داد که سیستم استنتاج عصبی-فازی تطبیقی از قابلیت بالایی در تشخیص بیماری‌های کبدی برخوردار است.

برای تشخیص بیماری کبد از منطق فازی استفاده شده است و پارامترهای آن توسط الگوریتم ژنتیک بهینه شده‌اند [۱۲]. بیماری کبد با استفاده از سیستم‌های استنتاج فازی genfis<sub>2</sub> (برمبنای تابع سوگنو) و genfis<sub>3</sub> (برمبنای تابع ممدانی) تشخیص داده شده است. دقت تشخیص بیماری کبد با استفاده از سیستم استنتاج فازی genfis<sub>2</sub> با هشت ویژگی، برابر با ۹۱.۶۶ درصد و با استفاده از سیستم استنتاج فازی genfis<sub>3</sub> با شش ویژگی، برابر با ۸۷.۸۹ درصد است. همچنین، خطای genfis<sub>2</sub> و genfis<sub>3</sub> به‌ترتیب برابر با ۰.۰۳۴ و ۰.۰۴۷ به‌دست آمده است.

هفت الگوریتم بوستینگ معروف شامل Gradient Boosting (GB)، AdaBoost، LogitBoost، SGBost، XGBoost و LightGBM برای تشخیص بیماری کبد استفاده شده‌اند [۱۳]. دو مجموعه داده عمومی و معتبر شامل مجموعه داده افراد مبتلا به بیماری کبد (LDPD) و مجموعه داده بیماران هندی مبتلا به بیماری کبد (ILPD) استفاده شده‌اند. برای افزایش دقت پیش‌بینی، از تکنیک‌هایی مانند تنظیم هایپرپارامترها، نرمال‌سازی و افزایش نمونه‌ها استفاده شده است. در بین هفت الگوریتم بوستینگ، روش Gradient Boosting (GB) به‌عنوان بهترین مدل شناخته شد. این الگوریتم دقت‌های قابل توجهی معادل ۹۸.۸۰٪ روی مجموعه داده LDPD و ۹۸.۲۹٪ روی مجموعه داده ILPD به‌دست آورده است.

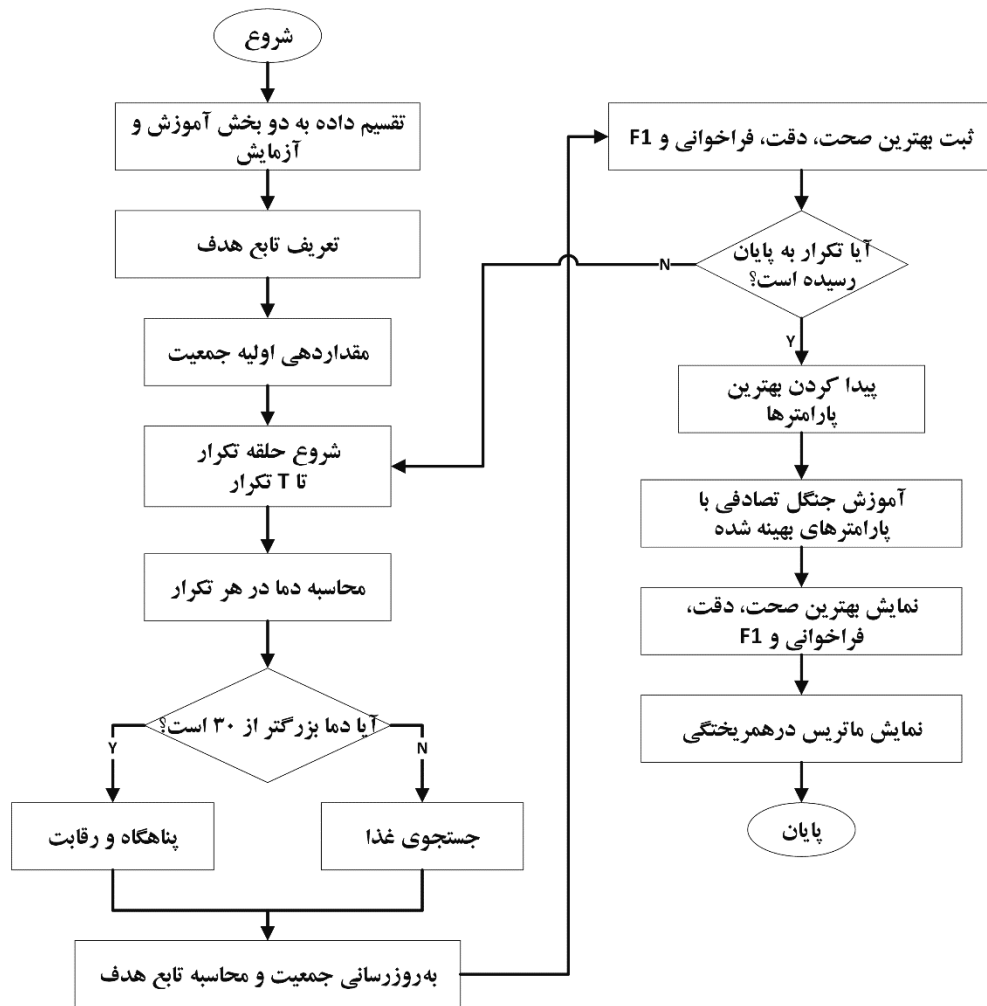
مدل ترکیبی برمبنای الگوریتم بهینه‌سازی شیرمورچه و k نزدیک‌ترین همسایه برای طبقه‌بندی افراد به دو گروه سالم و مبتلا به بیماری کبد طراحی شده است [۱۴]. الگوریتم بهینه‌سازی شیرمورچه برای انتخاب ویژگی استفاده شده است. مجموعه داده به‌صورت تصادفی به ۲۰ دسته تقسیم شده است که هر دسته شامل داده‌های آموزش و آزمون است. نتایج مدل در بهترین حالت نشان می‌دهد که صحت کلی براساس تمام ویژگی‌ها برابر با ۹۵.۲۳ درصد است. معیارهای ویژگی

و حساسیت به ترتیب برابر با ۹۳.۹۵ درصد و ۹۴.۱۱ درصد هستند. صحت مدل پیشنهادی با استفاده از ۵ ویژگی به ۹۸.۶۳ درصد رسیده است.

انتخاب ویژگی در یادگیری ماشین، داده کاوی و تشخیص الگو اهمیت بسزایی دارد و هدف آن بهبود الگوریتم های یادگیری ماشین از طریق تصفیه ویژگی ها است. الگوریتم بهینه سازی طوطی بهبود یافته برای فرآیند انتخاب ویژگی در داده های پزشکی استفاده شده است [۱۵]. ۷۰ درصد مجموعه داده به عنوان مجموعه آموزشی و ۳۰ درصد باقی مانده به عنوان مجموعه آزمایشی استفاده شده است. نتایج آزمایش ها نشان می دهد که درصد صحت مجموعه داده ILPD بالای ۸۰ درصد بوده است. الگوریتم بهینه سازی کپلر بهبود یافته برای انتخاب ویژگی در مجموعه داده هایی با ابعاد بالا طراحی شده است [۱۶]. این مدل با بهره گیری از یادگیری مبتنی بر مخالفت و عملگر فرار محلی مبتنی بر طبقه بندی k نزدیک ترین همسایه به عنوان یک ابزار قدرتمند برای تشخیص بیماری کبد شناخته می شود. نتایج مدل روی مجموعه داده ILPD برای معیارهای صحت، حساسیت، دقت و F۱ بترتیب ۷۶ درصد، ۵۰ درصد، ۵۰ درصد و ۵۰ درصد بدست آمده است. ارتقای الگوریتم آدابوست با ترکیب الگوریتم کرم شب تاب برای تشخیص بیماری انجام شده است [۲]. با بهره گیری از توانمندی های الگوریتم کرم شب تاب، دقت و کارایی مدل آدابوست در تشخیص اختلالات کبدی افزایش می یابد. مجموعه داده شامل ۵۸۳ رکورد با ۱۰ ویژگی مرتبط با بیماری کبد است. برای آموزش و آزمون، داده ها به صورت تصادفی تقسیم شدند؛ به طوری که ۸۰ درصد داده ها برای آموزش و ۲۰ درصد باقی مانده برای ارزیابی استفاده شده اند. نتایج نشان می دهد که استفاده از انتخاب ویژگی منجر به افزایش دقت و کارایی می شود. در بهترین حالت، مدل ترکیبی با استفاده از پنج ویژگی، دقتی معادل ۹۸.۶ درصد را به دست آورده است. با تمام ویژگی ها، دقت مدل به ۹۴.۱ درصد رسیده است.

## ۲. روش کار

در این مقاله، برای تشخیص بیماری کبد از الگوریتم بهینه سازی خرچنگ جهت تنظیم پارامترهای الگوریتم جنگل تصادفی در مدل پیشنهادی استفاده شده است. در شکل (۱) فلوچارت روش پیشنهادی ترسیم شده است.



شکل ۱: فلوجارت روش پیشنهادی برای تشخیص بیماری کبد

در این مقاله برای تشخیص بیماری کبد از مجموعه داده ILPD<sup>۴</sup> استفاده شده است. این مجموعه داده شامل سوابق ۴۱۶ نفر مبتلا به بیماری کبد و ۱۶۷ نفر بدون بیماری کبد است. مجموعه داده از نظر مقادیر گمشده بررسی و در صورت لزوم مقداردهی اولیه شده است. سپس ویژگی‌ها و برجسب‌ها از یکدیگر تفکیک شده‌اند و داده‌ها به دو بخش آموزشی و آزمون (۸۰ به ۲۰) تقسیم شده‌اند.

جدول ۱: متغیرهای مجموعه داده ILPD

شماره	نام متغیر	نوع	توضیحات	محدوده
۱	Age	عددی	سن	۴ تا ۹۰
۲	Gender	اسمی	جنسیت	مرد-زن
۳	TB (Total Bilirubin)	عددی	مجموع بیلی‌روبین	۰/۴ تا ۷۵

<sup>۴</sup> <https://archive.ics.uci.edu/dataset/۲۲۵/ilpd+indian+liver+patient+dataset>

۴	DB (Direct Bilirubin)	عددی	بیلی روبین مستقیم	۰/۱ تا ۱۹/۷
۵	Alkphos Alkaline Phosphotase	عددی	آلکالین فسفاتاز	۶۳ تا ۲۱۱۰
۶	SGPT Alamine Aminotransferase	عددی	آلانین آمینوترانسفراز	۱۰ تا ۲۰۰۰
۷	SGOT Aspartate Aminotransferase	عددی	آسپاراتات آمینوترانسفرا	۱۰ تا ۴۹۲۹
۸	TP (Total Protiens)	عددی	مجموع پروتئین	۲/۷ تا ۹/۶
۹	ALB (Albumin)	عددی	آلبومین	۵/۵ تا ۰/۹
۱۰	A/G Ratio (Albumin and Globulin Ratio)	عددی	نسبت آلبومین به گلوبولین	۲/۸ تا ۰/۳
۱۱	Selector field	اسمی	فیلد انتخاب کننده ناسالم (۱)، سالم (۲)	۱ یا ۲

## ۱.۲. الگوریتم جنگل تصادفی

الگوریتم جنگل تصادفی به دلیل توانایی بالا در تحلیل داده‌های پیچیده و مقاومت بالا در برابر بیش‌برازش، به عنوان یک مدل پیش‌بینی در این مقاله استفاده شده است. این الگوریتم متشکل از تعدادی درخت تصمیم‌گیری است که به صورت تصادفی، ویژگی‌ها و نمونه‌های گوناگون را برای تولید پیش‌بینی‌های نهایی ترکیب می‌کنند. پارامترهای کلیدی الگوریتم جنگل تصادفی که توسط الگوریتم بهینه‌سازی خرچنگ بهینه‌سازی می‌شوند شامل تعداد درخت‌های تصمیم‌گیری، عمق هر درخت تصمیم‌گیری و حداقل تعداد نمونه‌ها برای تقسیم یک گره هستند. تعداد درخت‌ها تأثیر مستقیمی بر دقت مدل دارد. افزایش بیش از حد آن ممکن است زمان اجرا را افزایش دهد ولی دقت را بهتر می‌کند. مقدار بهینه برای تعداد درخت‌ها برابر با ۱۵ درخت است. عمق درخت می‌تواند از بیش‌برازش جلوگیری کند. عمق بیش از حد بالا ممکن است به یادگیری جزئیات غیرضروری منجر شود. مقدار بهینه برای عمق درخت برابر با ۸ درخت است. حداقل تعداد نمونه‌ها برای تقسیم گره مشخص می‌کند که حداقل چه تعداد نمونه باید در هر گره وجود داشته باشند تا تقسیم جدیدی انجام شود. مقدار بهینه برای حداقل تعداد نمونه‌ها برابر با ۵ نمونه است.

## ۲.۲. الگوریتم بهینه‌سازی خرچنگ

الگوریتم بهینه‌سازی خرچنگ به عنوان یک الگوریتم فراابتکاری برای بهینه‌سازی جنگل تصادفی انتخاب شده است. مراحل الگوریتم بهینه‌سازی خرچنگ بشرح زیر است:

**مقداردهی اولیه جمعیت:** در ابتدا یک جمعیت تصادفی برای پارامترهای مدل (از جمله تعداد درخت‌ها، عمق درخت‌ها و حداقل تعداد نمونه لازم برای تقسیم داده‌ها) ایجاد می‌شود. در مسئله بهینه‌سازی چندبعدی، هر خرچنگ به صورت یک ماتریس  $1 \times dim$  نمایش داده می‌شود. هر سطر در این ماتریس، یک راه‌حل برای مسئله تلقی می‌شود. مجموعه متغیرهای  $(X_{i,1}, X_{i,2}, \dots, X_{i,dim})$  باید در محدوده بین کران بالا و کران پایین تعریف شوند. مقداردهی اولیه الگوریتم

شامل تولید تصادفی مجموعه‌ای از راه‌حل‌های کاندید ( $X$ ) در فضای جستجو است. راه‌حل‌های کاندید ( $X$ ) براساس اندازه جمعیت ( $N$ ) و ابعاد مسئله مقداردهی اولیه می‌شوند. مقداردهی اولیه الگوریتم طبق معادله (۱) تعریف می‌شود.

$$X = [X_1, X_2, \dots, X_N] = \begin{bmatrix} X_{1,1} & \dots & X_{1,j} & \dots & X_{1,dim} \\ \vdots & \dots & \vdots & \dots & \vdots \\ X_{i,1} & \dots & X_{i,j} & \dots & X_{i,dim} \\ \vdots & \dots & \vdots & \dots & \vdots \\ X_{N,1} & \dots & X_{N,j} & \dots & X_{N,dim} \end{bmatrix} \quad (1)$$

که ( $X$ ) موقعیت جمعیت اولیه است، ( $N$ ) تعداد جمعیت،  $dim$  ابعاد جمعیت،  $X_{i,j}$  موقعیت فرد  $i$ ام در بعد  $j$ ام است. مقدار  $X_{i,j}$  طبق معادله (۲) تعریف می‌شود. که  $lb_j$  نمایانگر کران پایینی بعد  $j$ ام است،  $ub_j$  نمایانگر کران بالایی بعد  $j$ ام است و  $rand$  یک عدد تصادفی است.

$$X_{i,j} = lb_j + (ub_j - lb_j) \times rand \quad (2)$$

**تعریف دما و تغذیه خرچنگ:** تغییر دما بر رفتار خرچنگ تاثیر می‌گذارد و آن را وارد مراحل مختلف می‌کند. در دمای بالای ۳۰ خرچنگ مکان خنکی برای پناه گرفتن انتخاب می‌کند. در دمای مناسب، خرچنگ‌ها رفتار جستجوی غذا را انجام می‌دهند (مقدار تغذیه خرچنگ تحت تاثیر دما می‌باشد). محدوده دمای بین ۲۰ تا ۳۰ درجه سانتی‌گراد برای تغذیه در بهترین حالت (تغذیه خرچنگ با توزیع نرمال) تقریب زده می‌شود. در این الگوریتم، دامنه دما بین ۲۰ تا ۳۵ درجه سانتی‌گراد تعریف می‌شود. دما طبق معادله (۳) تعریف می‌شود.

$$temp = rand \times 15 + 20 \quad (3)$$

مدل ریاضی برای تغذیه خرچنگ توسط معادله (۴) تعریف می‌شود. پارامتر  $\mu$  مناسب‌ترین دما برای خرچنگ (میانگین دما) است که مقدارش ۲۵ تعیین شده است. یعنی مناسب‌ترین دما برای زنده ماندن خرچنگ ۲۵ درجه سانتی‌گراد است. پارامترهای  $\sigma$  و  $C_1$  برای کنترل تغذیه خرچنگ در دماهای مختلف استفاده می‌شوند. در حالت طبیعی، این پارامترها روی مقادیر ۳ و ۰/۲ تنظیم شده‌اند.

$$p = C_1 \times \left( \frac{1}{\sqrt{2} \times \pi \times \sigma} \times \exp \left( -\frac{(temp - \mu)^2}{2\sigma^2} \right) \right) \quad (4)$$

**مراحل پناهگاه (اکتشاف):** اگر  $temp > 30$  و  $rand < 0,5$  باشد، آنگاه خرچنگ‌ها مستقیماً وارد پناهگاه می‌شوند و برای بدست آوردن پناهگاه رقابت نمی‌کنند. پناه گرفتن خرچنگ طبق معادله (۵) تعریف می‌شود.  $X_{shade}$  نشان‌دهنده پناهگاه،  $X_G$  بهترین موقعیت بدست آمده در تعداد تکرارها،  $X_L$  بیانگر موقعیت بهینه در جمعیت جاری است.

$$X_{shade} = \frac{(X_G + X_L)}{2} \quad (5)$$

مبارزه خرچنگ‌ها برای پناهگاه یک رویداد تصادفی است. زمانیکه  $\text{rand} < 0,5$  باشد، به این معنی است که خرچنگ‌ها برای تصاحب غار رقابت نمی‌کنند و خرچنگ‌ها به‌طور مستقیم وارد پناهگاه می‌شوند. ورود مستقیم خرچنگ به پناهگاه توسط معادله (۶) تعریف می‌شود. که  $t$  عدد تکرار فعلی،  $t+1$  تکرار بعدی و  $C_7$  منحنی کاهش است.

$$X_{i,j}^{t+1} = X_{i,j}^t + C_7 \times \text{rand} \times (X_{\text{shade}} - X_{i,j}^t) \quad (6)$$

$$C_7 = 2 - \left(\frac{t}{T}\right) \quad (7)$$

معادله (۷) برای منحنی کاهش  $(C_7)$  تعریف شده است.  $t$  تعداد جاری و  $T$  حداکثر تعداد تکرار است. در مرحله پناهگاه، هدف خرچنگ‌ها نزدیک شدن به غار است که نمایانگر راه‌حل بهینه می‌باشد. در این مرحله، خرچنگ‌ها به سمت غار حرکت می‌کنند. این کار باعث نزدیک‌تر شدن افراد به راه‌حل بهینه و تقویت بهره‌برداری می‌شود. **مرحله رقابت (بهره‌برداری):** هنگامی که دما بیشتر از  $30$  درجه سانتی‌گراد باشد و  $\text{rand} \geq 0,5$  باشد، یعنی سایر خرچنگ‌ها نیز به پناهگاه علاقه‌مند هستند. در این حالت، خرچنگ‌ها برای تصاحب پناهگاه با یکدیگر مبارزه می‌کنند. رقابت بین خرچنگ‌ها برای تصاحب پناهگاه توسط معادله (۸) تعریف می‌شود. خرچنگ  $X_i$  موقعیت خود را بر اساس موقعیت  $X_z$  (یک خرچنگ دیگر) تنظیم می‌کند.

$$X_{i,j}^{t+1} = X_{i,j}^t - X_{z,j}^t + X_{\text{shade}} \quad (8)$$

که  $Z$  نشان‌دهنده یک فرد تصادفی از خرچنگ‌ها است، که طبق معادله (۹) تعریف می‌شود.

$$z = \text{round}(\text{rand} \times (N - 1)) + 1 \quad (9)$$

**مرحله جستجوی غذا (بهره‌برداری):** زمانی که  $\text{temp} \leq 30$  باشد، دما برای تغذیه خرچنگ‌ها مناسب است. در این حالت، خرچنگ‌ها به سمت غذا حرکت می‌کنند. پس از یافتن غذا، خرچنگ‌ها اندازه غذا را ارزیابی می‌کنند. اگر غذا بیش از حد بزرگ باشد، خرچنگ‌ها غذا را با چنگال‌های خود تکه‌تکه می‌کنند و با استفاده از دومین و سومین پای حرکتی خود غذا را می‌خورند. موقعیت غذا  $X_{\text{food}}$  طبق معادله (۱۰) تعریف می‌شود.

$$X_{\text{food}} = X_G \quad (10)$$

اندازه غذا ( $Q$ ) طبق معادله (۱۱) تعریف می‌شود.

$$Q = C_7 \times \text{rand} \times \left(\frac{\text{fitness}_i}{\text{fitness}_{\text{food}}}\right) \quad (11)$$

که  $Q$  اندازه غذا،  $C_7$  ضریب غذا را نشان می‌دهد که برابر با مقدار ثابت  $3$  است.  $\text{fitness}_i$  بیانگر مقدار برازش خرچنگ  $i$ ام و  $\text{fitness}_{\text{food}}$  بیانگر مقدار برازش موقعیت غذا است. اگر  $Q > (C_7 + 1)/2$  یعنی غذا خیلی بزرگ است،



آنگاه خرچنگ با استفاده از اولین پای چنگالی، غذا را تکه تکه می کند. معادله (۱۰) برای خرد کردن غذا تعریف شده است، که  $X_{food}$  موقعیت غذا و  $Q$  اندازه غذا می باشد.

$$X_{food} = \exp\left(-\frac{1}{Q}\right) \times X_{food} \quad (12)$$

وقتی که غذا تکه تکه شد خرچنگها توسط دومین و سومین پا، غذا را برمی دارند و به دهان منتقل می کنند. برای شبیه سازی این فرآیند متناوب، ترکیبی از توابع سینوسی و کسینوسی استفاده می شوند. لذا، معادله جستجوی غذا طبق معادله (۱۳) تعریف می شود.

$$X_{i,j}^{t+1} = X_{i,j}^t + X_{food} \times p(\cos(\gamma \times \pi \times rand) - \sin(\gamma \times \pi \times rand)) \quad (13)$$

اگر  $Q \leq (C_3 + 1)/2$  باشد، آنگاه خرچنگ به سمت غذا حرکت می کند و مستقیماً غذا را می خورد. این کار طبق معادله (۱۴) تعریف شده است.

$$X_{i,j}^{t+1} = (X_{i,j}^t - X_{food}) \times p + p \times rand \times X_{i,j}^t \quad (14)$$

در مرحله جستجوی غذا، خرچنگها بر مبنای اندازه غذا ( $Q$ ) از روش های تغذیه ای متفاوتی استفاده می کنند و  $X_{food}$  به عنوان بهترین راه حل در نظر گرفته می شود. هنگامی که اندازه غذا ( $Q$ ) برای تغذیه خرچنگ مناسب باشد، خرچنگ به سمت غذا حرکت می کند. اما اگر ( $Q$ ) بیش از حد بزرگ باشد، این نشان می دهد که تفاوت قابل توجهی بین موقعیت خرچنگ و بهترین راه حل وجود دارد. لذا، مقدار  $X_{food}$  باید کاهش یابد و خرچنگ به غذا نزدیک تر شود.

### ۳.۲. معیار ارزیابی

کارایی روش پیشنهادی با بهره گیری از صحت، دقت، حساسیت و معیار  $F1$  و ماتریس درهم ریختگی سنجیده می شود. معیارهای

صحت، دقت، حساسیت و معیار  $F1$  برای پیش بینی مدل به کار می روند و ماتریس درهم ریختگی کمک می کند تا اطلاعات بیشتری

از بررسی اشتباهات مدل و کارایی آن در شناسایی کلاس های گوناگون بدست آید.

شماره	معادل	معیارها
۱۵	صحت	$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$
۱۶	دقت	$Precision = \frac{TP}{TP + FP}$
۱۷	فراخوانی	$Recall = \frac{TP}{TP + FN}$

$$F1 = \frac{2 \times P \times R}{R + P} \quad \text{معیار F1} \quad 18$$

### ۳. یافته ها

برای اجرای روش پیشنهادی (جنگل تصادفی مبتنی بر الگوریتم بهینه‌سازی خرچنگ) باید در ابتدا مقداردهی اولیه انجام گیرد. تعداد تکرار و جمعیت اولیه در مدل پیشنهادی به ترتیب برابر با ۱۰۰ و ۵۰ هستند. در جدول (۲)، نتایج مدل پیشنهادی بر مبنای تکرارهای مختلف بیان شده است. مدل پیشنهادی روی مجموعه داده‌های تست در ۱۰۰ تکرار دارای بیشترین مقدار صحت ۰/۷۳۵۶، در ۴۰ تکرار دارای بالاترین مقدار دقت ۰/۷۶۸۷، در ۸۰ و ۱۰۰ تکرار بیشترین مقدار فراخوانی ۱۰۰۰۰ و در ۱۰۰ تکرار دارای بیشترین معیار F1 یعنی برابر ۰/۸۴۶۷ می‌باشد. نتایج مجموعه داده‌های تست بیانگر این است که روش پیشنهادی توانسته است، مجموعه داده‌های آموزشی را با دقت بالایی طبقه‌بندی نماید. لذا در طبقه‌بندی مجموعه داده‌های تست مشکل تشخیص نوع کلاس برای نمونه‌های جدید، پایین بوده است. اگر تعداد تکرار افزایش یابد روش پیشنهادی می‌تواند نقاط بهینه را کشف کند و همچنین به‌روزرسانی موقعیت در روش پیشنهادی برای پیدا کردن فاصله نزدیک در بهترین حالت انجام می‌گیرد. به دلیل اینکه حجم مجموعه داده‌ها بزرگ نیست، مشکل آموزش بیش‌برازش رخ نداده است.

#### ۱.۳. ارزیابی بر مبنای تعداد تکرارها

در جدول (۲) نتایج مدل پیشنهادی بر مبنای تعداد تکرار بیان شده است. با افزایش تعداد تکرارها به ۴۰، بهبود قابل توجهی در معیارها مشاهده می‌شود، صحت به ۰/۷۲۴۱، دقت به ۰/۷۶۸۷، فراخوانی به ۰/۸۸۹۷ و معیار F1 به ۰/۸۲۴۸ افزایش می‌یابند. با ۸۰ تکرار، نتایج بسیار چشمگیرتر می‌شوند، مقدار صحت و دقت هر دو به ۰/۷۲۹۸ دست یافته‌اند و فراخوانی به مقدار کامل ۱۰۰۰۰ افزایش می‌یابد، که به این معناست که روش پیشنهادی تمامی نمونه‌های مثبت واقعی را به درستی شناسایی کرده است. با ۱۰۰ تکرار، روش پیشنهادی به بالاترین سطح پایداری و کارایی دست یافته است. صحت به ۰/۷۳۵۶ و دقت به ۰/۷۳۴۱ افزایش پیدا می‌کنند، در حالی که فراخوانی در سطح ۱۰۰۰۰ باقی می‌ماند. این امر نشان‌دهنده تعادل بهینه بین شناسایی نمونه‌های مثبت و دقت در پیش‌بینی‌ها است. معیار F1 نیز به مقدار ۰/۸۴۶۷ رسیده است، که بهترین عملکرد را در میان تمام تکرارها کسب کرده است. افزایش تعداد تکرارها به‌وضوح تأثیر مثبتی بر بهبود تمامی معیارها داشته است و موجب افزایش کارایی به ویژه در شناسایی نمونه‌های مثبت واقعی شده است.

جدول ۲: نتایج مدل پیشنهادی بر مبنای تعداد تکرار

ردیف	تعداد تکرار	صحت	دقت	فراخوانی	معیار F1
۱	۲۰	۰/۷۰۶۸	۰/۷۵۶۷	۰/۸۸۱۸	۰/۸۱۴۵
۲	۴۰	۰/۷۲۴۱	۰/۷۶۸۷	۰/۸۸۹۷	۰/۸۲۴۸
۳	۶۰	۰/۷۱۲۶	۰/۷۶۵۵	۰/۸۷۴۰	۰/۸۱۶۱
۴	۸۰	۰/۷۲۹۸	۰/۷۲۹۸	۱۰۰۰۰	۰/۸۴۳۸
۵	۱۰۰	۰/۷۳۵۶	۰/۷۳۴۱	۱۰۰۰۰	۰/۸۴۶۷

### ۲.۳. مقایسه نتایج

در این بخش، مقایسه روش پیشنهادی با پنج مدل یادگیری ماشین انجام شده است. بهترین عملکرد در معیار صحت متعلق به روش پیشنهادی است. در جدول (۳) مقایسه روش پیشنهادی با مدل های دیگر بیان شده است. معیار دقت متعلق به جنگل تصادفی، معیار حساسیت متعلق به ماشین بردار پشتیبان و روش پیشنهادی می باشد. البته روش پیشنهادی در هر چهار معیار به نتایج مطلوب و عالی دست یافته است. روش پیشنهادی در معیار صحت برابر  $0.7356$ ، دقت  $0.7341$ ، فراخوانی  $1/0.000$  و معیار  $F1$  به  $0.8647$  دست یافته است. در روش ماشین بردار پشتیبان، صحت و دقت هر دو برابر  $0.7298$  هستند و فراخوانی به مقدار کامل  $1.0000$  رسیده است، که نشان می دهد تمامی نمونه های مثبت به درستی شناسایی شده اند. معیار  $F1$  نیز برابر  $0.8438$  است، که تعادل خوبی میان دقت و فراخوانی را نشان می دهد. در روش رگرسیون لجستیک، صحت  $0.7126$ ، دقت  $0.7619$ ، فراخوانی  $0.8818$  و معیار  $F1$  برابر با  $0.8175$  است. این روش عملکرد متعادلی دارد و به دلیل فراخوانی بالا می تواند تعداد زیادی از نمونه های مثبت واقعی را شناسایی کند، هر چند صحت آن کمی پایین تر است.

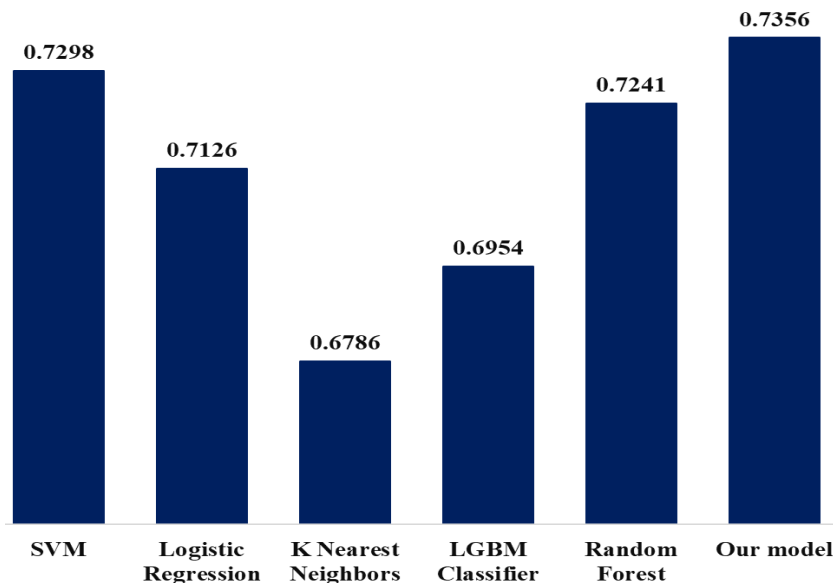
در روش  $K$  نزدیک ترین همسایه، صحت  $0.6786$ ، دقت  $0.7629$ ، فراخوانی  $0.5600$  و معیار  $F1$  برابر  $0.6700$  است. این مقادیر نشان می دهند که دقت نسبتاً بالاست، اما فراخوانی پایین به معنای ناتوانی این روش در شناسایی بسیاری از نمونه های مثبت واقعی است، که موجب کاهش معیار  $F1$  و صحت شده است. در روش ماشین گرادیان بوستینگ لایت، صحت  $0.6954$ ، دقت  $0.7720$ ، فراخوانی  $0.8267$  و معیار  $F1$  برابر  $0.7984$  است، که نشان دهنده عملکرد مناسب و تعادل خوب بین دقت و فراخوانی است. روش جنگل تصادفی، با صحت  $0.7241$ ، دقت  $0.7687$ ، فراخوانی  $0.8897$  و معیار  $F1$  برابر  $0.8247$ ، عملکرد بسیار خوبی کسب کرده است و توانسته است تعادل قابل توجهی میان تشخیص نمونه های مثبت و پیش بینی دقیق ایجاد کند. در نهایت، روش پیشنهادی بهترین نتایج را کسب کرده است؛ صحت روش پیشنهادی برابر با  $0.7356$ ، دقت  $0.7341$  و فراخوانی به مقدار کامل  $1.0000$  رسیده است، که نشان می دهد تمامی نمونه های مثبت واقعی شناسایی شده اند. معیار  $F1$  نیز با مقدار  $0.8647$  بهترین عملکرد کلی را نشان می دهد و بیانگر تعادل ایده آل میان دقت و فراخوانی است.

جدول ۳: مقایسه روش پیشنهادی با مدل های دیگر

روش ها	صحت	دقت	فراخوانی	معیار $F1$
ماشین بردار پشتیبان	$0.7298$	$0.7298$	$1.0000$	$0.8438$
رگرسیون لجستیک	$0.7126$	$0.7619$	$0.8818$	$0.8175$
$K$ نزدیکترین همسایه	$0.6786$	$0.7629$	$0.5600$	$0.6700$
ماشین گرادیان بوستینگ لایت	$0.6954$	$0.7720$	$0.8267$	$0.7984$
جنگل تصادفی	$0.7241$	$0.7687$	$0.8897$	$0.8247$
روش پیشنهادی	$0.7356$	$0.7341$	$1.0000$	$0.8647$

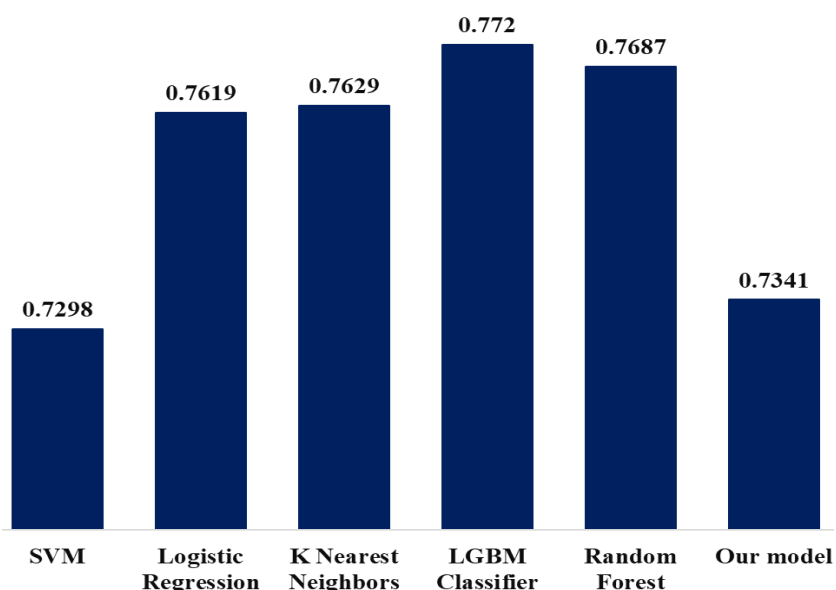
در شکل (۲) نمودار مقایسه مدل ها بر مبنای معیار صحت ترسیم شده است. ماشین بردار پشتیبان با صحت  $0.7298$  عملکرد نسبتاً بالایی دارد و یکی از مدل های موفق در شناسایی نمونه های صحیح است. رگرسیون لجستیک به صحت  $0.7126$  دست یافته است، که کمی پایین تر از ماشین بردار پشتیبان است، اما همچنان عملکرد قابل قبولی دارد. معیار

صحت در  $k$  نزدیکترین همسایه برابر با ۰.۶۷۸۶ است که کمترین مقدار را در بین تمامی روش‌ها کسب کرده است. جنگل تصادفی با صحت ۰.۷۲۴۱ به ماشین بردار پشتیبان نزدیک است و توانایی بالایی در شناسایی نمونه‌های صحیح دارد.



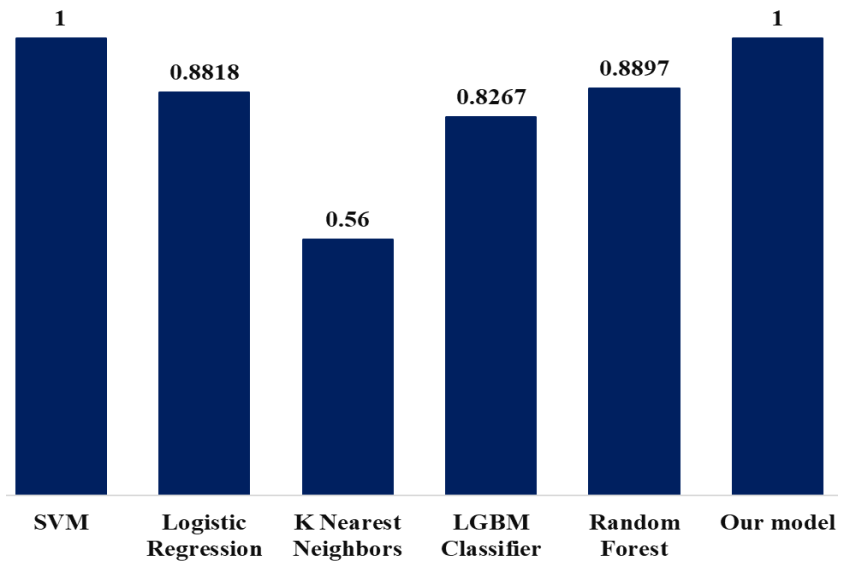
شکل ۲: نمودار مقایسه مدل‌ها بر مبنای معیار صحت

در شکل (۳) نمودار مقایسه مدل‌ها بر مبنای معیار دقت ترسیم شده است. دقت ماشین بردار پشتیبان برابر با ۰/۷۲۹۸ است. این مقدار نشان می‌دهد که ماشین بردار پشتیبان در پیش‌بینی نمونه‌های مثبت عملکرد نسبتاً خوبی کسب کرده است، اما در مقایسه با سایر روش‌ها دقت پایین‌تری دارد. دقت رگرسیون لجستیک برابر با ۰/۷۶۱۹ است که بالاتر از ماشین بردار پشتیبان است. دقت  $k$  نزدیک‌ترین همسایه برابر با ۰/۷۶۲۹ است. این مقدار نزدیک به دقت رگرسیون لجستیک است و بیانگر این است که این روش در پیش‌بینی نمونه‌های مثبت عملکرد نسبتاً مناسبی داشته است. ماشین‌گرادیان بوستینگ لایت دارای بیشترین دقت در بین تمامی مدل‌ها است و دقت آن ۰/۷۷۲۰ می‌باشد. این مقدار نشان می‌دهد که مدل‌گرادیان بوستینگ در پیش‌بینی نمونه‌های مثبت بسیار دقیق عمل کرده است و احتمال خطاهای مثبت کاذب در آن کمتر است. جنگل تصادفی با دقت ۰.۷۶۸۷ عملکرد خوبی داشته است و نزدیک به ماشین‌گرادیان بوستینگ است.



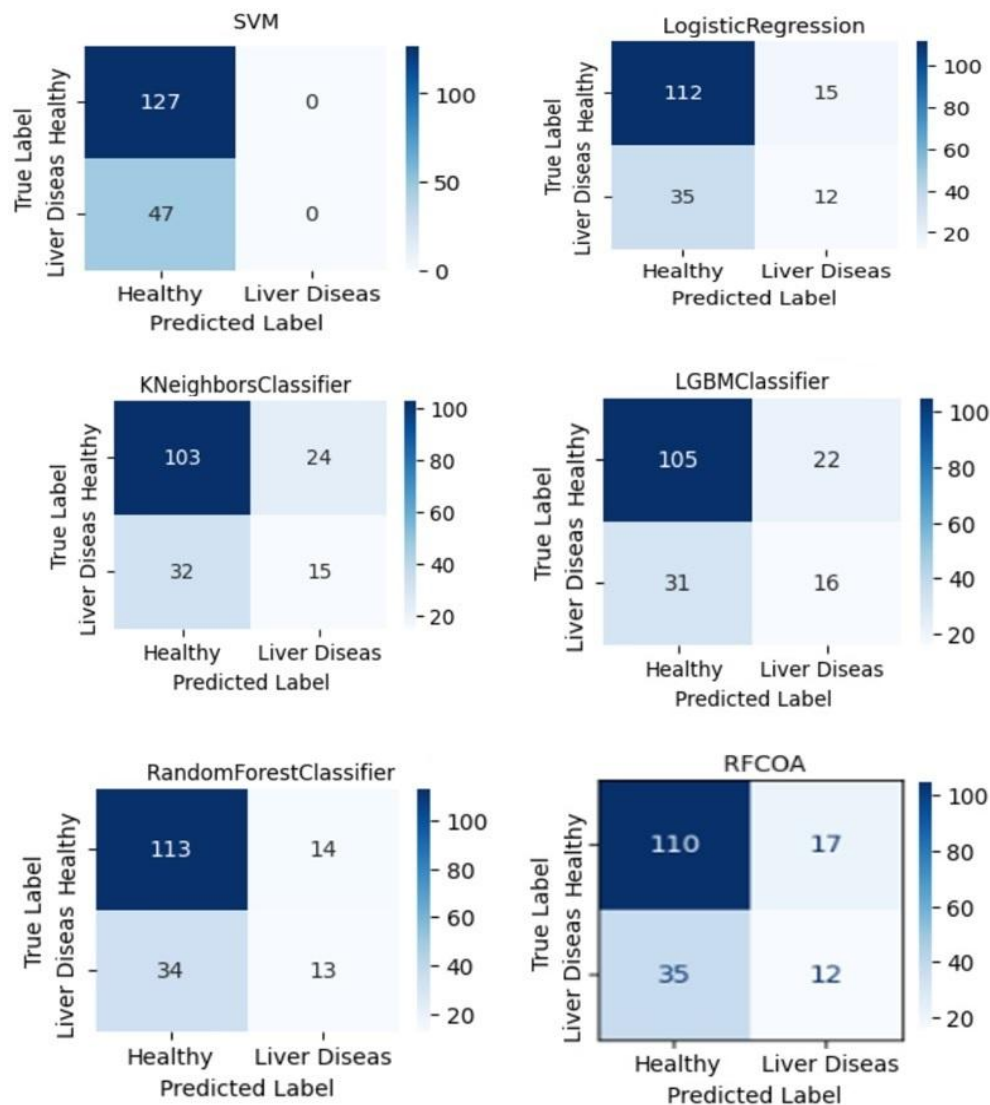
شکل ۳: نمودار مقایسه مدل‌ها بر مبنای معیار دقت

در شکل (۴) نمودار مقایسه مدل‌ها بر مبنای معیار فراخوانی ترسیم شده است. ماشین بردار پشتیبان بالاترین فراخوانی را با مقدار ۱ دارد. یعنی ماشین بردار پشتیبان توانسته که تمامی نمونه‌های مثبت واقعی را شناسایی کند و هیچ نمونه مثبت واقعی را از دست نداده است. مقدار فراخوانی برای رگرسیون لجستیک برابر با  $0/8818$  است. لذا بخش زیادی از نمونه‌های مثبت واقعی توسط این مدل شناسایی شدند، اما همچنان تعدادی از نمونه‌های مثبت واقعی از دست رفته‌اند. مقدار فراخوانی  $k$  نزدیک‌ترین همسایه برابر  $0.56$  است، که نشان می‌دهد این مدل عملکرد ضعیفی در شناسایی نمونه‌های مثبت واقعی داشته است و تعداد زیادی از نمونه‌های مثبت را شناسایی نکرده است. فراخوانی در جنگل تصادفی برابر با  $0/8897$  است، که نشان‌دهنده عملکرد بالای این مدل در شناسایی نمونه‌های مثبت واقعی است و نزدیک به رگرسیون لجستیک است. روش پیشنهادی نیز بالاترین فراخوانی ممکن را با مقدار ۱ بدست آورده است. این نتیجه نشان می‌دهد که روش پیشنهادی تمامی نمونه‌های مثبت واقعی را شناسایی کرده است.



شکل ۴: نمودار مقایسه مدل‌ها بر مبنای معیار فراخوانی

ماتریس درهم‌ریختگی برای سنجش کارایی مدل‌های یادگیری ماشین استفاده می‌شود. این ماتریس نمایانگر تعداد نمونه‌های صحیح و نادرست است. ماتریس درهم‌ریختگی داده‌ها را در چهار گروه متفاوت دسته‌بندی می‌کند. (۱) TP: تعداد نمونه‌هایی که مدل به درستی پیش‌بینی کرده است. (۲) TN: تعداد نمونه‌هایی که مدل به درستی آنها را سالم تشخیص داده است. (۳) FP: تعداد نمونه‌هایی که مدل به اشتباه پیش‌بینی کرده است که فرد بیمار است. (۴) FN: تعداد نمونه‌هایی که مدل به نادرستی پیش‌بینی کرده که سالم هستند. طبق شکل (۵) ماتریس درهم‌ریختگی برای مدل‌های ماشین بردار پشتیبان، رگرسیون لجستیک، k نزدیک‌ترین همسایه، ماشین گرادیان بوستینگ لایت، جنگل تصادفی و روش پیشنهادی بیان شده است.



شکل ۵: نتایج ماتریس درهم‌ریختگی برای مدل‌های مختلف

#### ۴. بحث و نتیجه گیری

در این مطالعه، برای تشخیص بیماری کبد، یک مدل جدید بر مبنای ترکیب الگوریتم بهینه‌سازی خرچنگ و جنگل تصادفی پیشنهاد شد. پارامترهای جنگ تصادفی با استفاده از الگوریتم بهینه‌سازی خرچنگ بهینه شدند. نتایج روش پیشنهادی نشان داد که با استفاده از الگوریتم بهینه‌سازی خرچنگ می‌توان به‌طور موثری بهترین مقادیر را برای پارامترهای جنگ تصادفی پیدا کرد. برای روش پیشنهادی مقادیر معیارها با ۱۰۰ تکرار برای چهار معیار صحت برابر با ۰/۷۳۵۶، دقت ۰/۷۳۴۱، فراخوانی ۱/۰۰۰۰ و معیار F۱ برابر با ۰/۸۰۹۷ بدست آمد. روش‌های جنگ تصادفی و ماشین بردار پشتیبان نیز عملکرد خوبی کسب کردند. از سوی دیگر، روش K نزدیک‌ترین همسایه ضعیف‌ترین نتایج را نشان داد و تعادل مناسبی بین دقت و فراخوانی ندارد. این تحلیل نشان می‌دهد که روش پیشنهادی با دقت و کارایی بالا یک گزینه برتر برای حل تشخیص بیماری کبد می‌باشد. براساس معیار فراخوانی، روش پیشنهادی و ماشین بردار پشتیبان

بهترین عملکرد را کسب کردند و هیچ نمونه مثبتی را از دست ندادند. مدل های جنگل تصادفی و رگرسیون لجستیک نیز عملکرد مناسبی داشته اند و به فراخوانی بهینه دست یافته اند. در مقابل،  $k$  نزدیک ترین همسایه با مقدار فراخوانی بسیار پایین (۰/۵۶) ضعیف ترین عملکرد را بدست آورده است، که نشان دهنده ناتوانی این مدل در شناسایی تعداد قابل توجهی از نمونه های مثبت واقعی است. همچنین با توجه به ماتریس درهم ریختگی ۱۱۰ نمونه به عنوان بیماران واقعی، ۱۲ نمونه به عنوان افراد سالم، ۳۵ نمونه تشخیص اشتباه به عنوان بیمار و ۱۷ نمونه تشخیص اشتباه به عنوان سالم توسط روش پیشنهادی شناسایی شدند که سبب به حداقل رساندن تشخیص های نادرست و شناسایی صحیح کلاس های مختلف شده است. تحقیقات و کارهای آینده در راستای این مقاله شامل جهت گیری های زیر است:

استفاده از روش پیشنهادی روی مجموعه داده های دیگر: به عنوان مثال، استفاده از روش پیشنهادی روی حوزه های مختلف مانند پردازش تصاویر، تشخیص تقلب، پیش بینی بیماری ها یا پردازش متن، منجر به درک نقاط قوت و ضعف الگوریتم خواهد شد.

ترکیب روش پیشنهادی با سایر مدل های یادگیری ماشین: ترکیب روش پیشنهادی با سایر مدل ها می تواند منجر به طراحی مدل های ترکیبی شود. استفاده از مدل های ترکیبی باعث افزایش دقت، فراخوانی یا دقت پیش بینی شود. ترکیب چندین الگوریتم مختلف منجر به کاهش خطاهای احتمالی و افزایش توانایی مدل در شناسایی الگوهای پیچیده در داده ها می شود.

استفاده از روش های انتخاب و استخراج ویژگی: استفاده از این تکنیک ها می تواند باعث کاهش ابعاد داده، حذف ویژگی های زائد و بهبود عملکرد مدل شود. روش های استخراج ویژگی (مانند تحلیل مؤلفه های اصلی یا تجزیه مقدار منفرد) می توانند داده های ورودی را بهینه تر، سرعت و دقت روش پیشنهادی را بهبود دهند.

#### References

- [۱] Abdel-Misih SR, Bloomston M, Liver anatomy, Surg Clin North Am. ۲۰۱۰; ۹۰(۴):۲۰۱۰.
- [۲] Ardam S, Gharehchopogh F, Diagnosing liver disease using Firefly Algorithm based on Adaboost, Journal of Health Administration (JHA). ۲۰۱۹; ۲۲(۱):۲۰۱۹.
- [۳] Minaee S, Kafieh R, Sonka M, Yazdani S, and Jamalipour Soufi G, Deep-COVID: Predicting COVID-۱۹ from chest X-ray images using deep transfer learning, Medical Image Analysis. ۲۰۲۰; ۶۵(۱):۲۰۲۰.
- [۴] Kumar Y, Singla R (۲۰۲۱) Federated Learning Systems for Healthcare: Perspective and Recent Progress, in *Federated Learning Systems: Towards Next-Generation AI*, Springer International Publishing: Cham. p. ۱۴۱-۱۵۶.
- [۵] Tengnah MAJ, Sooklall R, and Nagowah SD (۲۰۱۹) Chapter ۹ - A Predictive Model for Hypertension Diagnosis Using Machine Learning Techniques, in *Telemedicine Technologies*, Academic Press. p. ۱۳۹-۱۵۲.
- [۶] Liu Y, Wang Y, and Zhang J (۲۰۱۲) New Machine Learning Algorithm: Random Forest. in *Information Computing and Applications*. Berlin, Heidelberg. ۲۴۶-۲۵۲.
- [۷] Gharehchopogh FS, Gholizadeh H, A comprehensive survey: Whale Optimization Algorithm and its applications, Swarm and Evolutionary Computation. ۲۰۱۹; ۴۸(۱):۲۰۱۹.



- [۸] Dokeroglu T, Sevinc E, Kucukyilmaz T, and Cosar A, A survey on new generation metaheuristic algorithms, *Computers & Industrial Engineering*. ۲۰۱۹; ۱۳۷(۱):۲۰۱۹.
- [۹] Jia H, Rao H, Wen C, and Mirjalili S, Crayfish optimization algorithm, *Artificial Intelligence Review*. ۲۰۲۳; ۵۶(۲):۲۰۲۳.
- [۱۰] Shahraki MR, Mesgar M, Evaluation of Data Mining Algorithms for Detection of Liver Disease, *Payavard-Salamt*. ۲۰۱۹; ۱۳(۱):۲۰۱۹.
- [۱۱] Ahmadi M, Ramezanpour M, and Khorsand R, Diagnosis of Liver Disorders Using a Combination of Adaptive Neuron-Fuzzy Inference System and Particle Swarm Optimization Algorithm, *Health Information Management*. ۲۰۱۹; ۱۶(۳):۲۰۱۹.
- [۱۲] Rezaii Farokh zad M, Soleimanian Gharehchopogh F, Determining Fuzzy Logic Parameters by using Genetic Algorithm for the Diagnosis of Liver Disease, *jhbmi*. ۲۰۱۸; ۵(۳):۲۰۱۸.
- [۱۳] Ganie SM, Dutta Pramanik PK, A comparative analysis of boosting algorithms for chronic liver disease prediction, *Healthcare Analytics*. ۲۰۲۴; ۵(۱):۲۰۲۴.
- [۱۴] Javadzadeh S, Shayanfar H, and Soleimanian Gharehchopogh F, 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*. ۲۰۲۰; ۲۸(۵):۲۰۲۰.
- [۱۵] Huang G, Wei J, Yuan Y, Huang H, and Chen H, Parrot optimization algorithm for improved multi-strategy fusion for feature optimization of data in medical and industrial field, *Swarm and Evolutionary Computation*. ۲۰۲۵; ۹۵(۱):۲۰۲۵.
- [۱۶] Houssein EH, Abdalkarim N, Samee NA, Alabdulhafith M, and Mohamed E, Improved Kepler Optimization Algorithm for enhanced feature selection in liver disease classification, *Knowledge-Based Systems*. ۲۰۲۴; ۲۹۷(۱):۲۰۲۴.