

# تشخیص بیماری های مزمن کلیه با استفاده از ترکیب الگوریتم فازی خوشن کسری بهبود یافته، k-MENS خوشه بندی و بهینه سازی ازدحام موش ها و مروری بر روش های استفاده شده در تشخیص بیماریهای مزمن کلیوی

سید حسن نجات<sup>1</sup>، سید حسین هاشمی<sup>2</sup>، مهیار ثابت<sup>3</sup> و محمد سردشتی فرد<sup>4</sup>

<sup>1</sup> دانشگاه آزاد اسلامی واحد تهران مرکز-دکتری تخصصی کامپیوتر مهندسی نرم افزار

<sup>2</sup> دانشگاه آزاد اسلامی واحد تهران مرکز-دکتری تخصصی کامپیوتر مهندسی نرم افزار

<sup>3</sup> دانشگاه آزاد اسلامی واحد تهران مرکز-دکتر تخصصی مدیریت

<sup>4</sup> دانشگاه آزاد اسلامی واحد ملارد-کارشناسی ارشد مهندسی نرم افزار

## چکیده

در این مقاله، پیش‌بینی و طبقه‌بندی بیماری مزمن کلیوی مبتنی بر داده‌های بزرگ Hadoop با استفاده از خوشه‌بندی K-means فازی خوشن کسری (IF-RFKM) و XG تقویت کننده ازدحام موش ها پیشنهاد شده است. در اینجا، روش خوشه بندی IF-RFKM در نظر گرفته شده است برای پیش بینی بیماری این بیماری با استفاده از طبقه بندی تقویت کننده XG طبقه بندی مراحل بیماری های مزمن کلیوی به عنوان طبیعی و غیر طبیعی الگوریتم بهینه سازی ازدحام موش (RSO) برای بهینه سازی پارامترهای پیشنهاد شده است طبقه بندی XGboost در ابتدا، داده ها به طور تصادفی از مجموعه داده های CKD تولید می شوند

شبیه سازی به زبان python انجام شده است. از شبیه سازی، روش پیشنهادی به دقت بالاتر 99.57٪، 98.28٪ و 97.35٪، فراخوان بالاتر 98.23٪، 88.34٪ می رسد همچنین 78.96٪ و زمان اجرای کمتر 92.15٪، 90.25٪ و 92.48٪ در مقایسه با روش های همانند تشخیص بیماری مزمن کلیوی و طبقه بندی با حذف ویژگی های بازگشتی با استفاده از درخت تصمیم موجود بدست آمده است. روش استفاده شده طبقه بندی و خوشه بندی بیماری با استفاده از رگرسیون لجستیک (CCKD-LR) و طبقه بندی بیماری مزمن کلیه با استفاده از دستگاه ناقل پشتیبانی چند هسته ای با (CCKD-MKC-SVM) بکار برده شده است.

## اطلاعات مقاله

ناریخچه مقاله:

تاریخ دریافت مقاله: 1401/10/05

تاریخ پذیرش مقاله: 1401/11/07

تاریخ انتشار مقاله: 1401/11/20

کلمات کلیدی:

تشخیص

بیماری

کلیه

الگوریتم فازی

ازدحام موش ها

خوشه بندی

کلان داده

k-MENS

## 1 - مقدمه

امروزه فناوری ها با تغییرات اساسی توسعه یافته اند که منجر به بهبود فناوری هایی با حجم عظیم داده می شود. اگر از داده ها استفاده نشود، به مقدار داده ای تبدیل می شود که قابل استفاده نیست. بنابراین، در صورتی که با استفاده از روشی به نام داده کاوی به درستی مدیریت شوند، مقداری از داده های غیرقابل استفاده می توانند به عنوان یک مبنای داده بسیار ارزشمند استفاده شوند.

داده کاوی برای کشف طرح کلی محرک از اطلاعات در یک پایگاه داده پنهان استفاده می شود. با استفاده از تکنیک های یادگیری ماشینی و محاسبات آماری، کشف الگو برای ساخت مدل هایی آماده می شود که عملکرد داده ها را پیش بینی می کنند. طبقه بندی تکنیکی برای کشف الگوها از داده ها با حجم عظیمی از داده ها به سرعت توسط تکنیک داده کاوی تجزیه و تحلیل می شوند، در واقع، در زمینه آمار، تجارت، پیش بینی و مهندسان ارتباطات، داده کاوی تکنیک پیشرفته ای نیست که از الگوها در جستجوی داده ها برای پیش بینی، تأیید و شناسایی استفاده کند. در حال حاضر پایگاه های بالینی پزشکی زیادی بر اساس پیشرفت سیستم مدیریت پایگاه داده سلامت ایجاد شده است. با تکنیک های داده کاوی، پایگاه های داده ترکیبی به راحتی حفظ و نگهداری و قابل استفاده می شوند.

داده کاوی برای شناسایی بیماری از پرونده داده های پزشکی بیمار در بخش مراقبت های بهداشتی به کار می رود. طبقه بندی یک استراتژی داده کاوی است. طبقه بندی بیزی، درخت تصمیم، شبکه های عصبی، رگرسیون لجستیک، ماشین های بردار پشتیبان نمونه هایی از تکنیک های طبقه بندی هستند. برای پیش بینی و طبقه بندی، درخت تصمیم به عنوان رایج ترین تکنیک مورد استفاده قرار گرفته می شود.

در فرآیند جستجو، طبقه بندی می تواند برای نمایش ها یا توابع معمولی استفاده شود که کلاس های داده یا طبقه بندی را متمایز می کند، با این هدف که این تکنیک بتواند برای محاسبه دقیق کلاس های هدف استفاده شود. نمونه ای از روش طبقه بندی در فناوری، طبقه بندی تهدیدات در توسعه نرم افزار است. همچنین در طرح مراقبت از سلامت انسان، فناوری ارتباط با اطلاعات دارای شهود اضافی است. طرح رعایت بهداشت و علم پزشکی می تواند درآمد و آورده های هنگفتی را روز به روز ایجاد کند. روش های تجزیه و تحلیل پزشکی و ارائه دهندگان خدمات بهداشتی، حجم عظیمی از داده ها از جمله صدا، متن، ویدئو، تصویر، که قابل

مدیریت بوده و بسیار مشکل ساز بوده و با روش های مرسوم مورد بررسی قرار می گیرند. روش های داده کاوی و روش های یادگیری ماشین، روش ها و ابزارهایی را برای اصلاح این داده های مختلف به عنوان داده های نتیجه ای برای تصمیم گیری توزیع می کنند. روش ها با یادگیری نیمه نظارتی و همچنین یادگیری بدون نظارت مورد بررسی قرار گرفتند. نیمه نظارت گروهی از استراتژی های یادگیری ماشینی است که با یادگیری یک تکنیک، از نمونه های برچسب دار و بدون برچسب استفاده می کند.

اساساً یادگیری بدون نظارت جایگزین خوشه بندی است. به طور کلی، خوشه بندی راه حلی برای کلاس های آشکار شده در داده ها است. خوشه بندی یک روش رایج برای تجزیه و تحلیل عددی داده ها و همچنین تکنیکی از داده های بدون نظارت است. به طور معمول در زمینه کاربردهای پزشکی می توان از خوشه بندی برای پیش آگهی بیماری های مختلف استفاده کرد. از بین بیماری های مختلف، بیماری مزمن کلیه یک بیماری قابل توجه برای عموم مردم است.

CKD وضعیتی است که در آن کلیه ها آسیب می بینند و نمی توانند خون را آنطور که باید فیلتر کنند. به همین دلیل، مایعات اضافی و مواد زائد خون در بدن باقی می ماند و ممکن است باعث مشکلات سلامتی دیگری مانند بیماری قلبی و سکتته شود. در مرگ در عملکرد کلیه، CKD نیز نتیجه متقابل را آغاز می کند. افراد مبتلا به CKD از این بیماری رنج می بردند که عملکرد کلیه را کاهش می دهد. علاوه بر این، باعث ایجاد اختلالات اضافی مانند تغذیه نامناسب، استخوان های ضعیف تر، لذت خون بالا، کم خونی، آسیب عصبی می شود. در این بین، بیماری کلیوی خطر ابتلای بیمار به بیماری قلبی را در کنار بیماری های خونی افزایش می دهد. ارثی، پرفشاری خون و دیابت نیز توسط CKD منشاء می گیرند. پیش آگهی زودرس CKD برای درمان مناسب برای جلوگیری از نارسایی کلیه ضروری است.

پیش بینی های اولیه از طریق رویکردهای یادگیری ماشینی، مانند Naïve Bayes، SVM، درخت تصمیم و ANN و غیره انجام می شوند. جنبه داده می تواند توسط طبقه بندی کننده برای پیش آگهی CKD ساخته شود.

در طول دهه های گذشته، مدل های داده کاوی متعددی برای تشخیص بیماری مزمن کلیوی پیشنهاد شده اند. کاربرد مدیریت دانش به دلیل کارایی آن در طول طبقه بندی در امتداد پیش بینی بیماری به سرعت در بخش های مراقبت های بهداشتی در حال رشد است و برای تجزیه و تحلیل تعداد زیادی از داده ها

- روش پیشنهادی در پایتون فعال می‌شود، عملکرد آن با معیارهای عملکرد مانند دقت، امتیاز F1، دقت، یادآوری بررسی می‌شود.
- کارایی روش پیشنهادی با روش‌های موجود مانند CCKD-RFE-DT، CCKD-LR 23 و CCKD-MKC- SVM مقایسه می‌شود.

مابقی این مقاله به این صورت طراحی شده است: بخش 2 مطالعات اخیر را مرور می‌کند، بخش 3 در مورد روش پیشنهادی توضیح می‌دهد، بخش 4 اثبات می‌کند. نتیجه با بحث، بخش 5 این مقاله را به پایان می‌رساند.

## 2- مرور مطالعات اخیر

بسیاری از مطالعات مربوط به مدل تحلیلی Hadoop-big data برای پیش‌بینی و طبقه‌بندی بیماری‌های مزمن کلیوی در ادبیات پیشنهاد شده‌اند، برخی از مطالعات اخیر در اینجا مرور می‌شوند.

Jerlin Rubini و Perumal یک طبقه‌بندی بیماری مزمن کلیوی را با استفاده از دستگاه بردار پشتیبان چند هسته با رویکرد بهینه‌سازی مگس میوه پیشنهاد کرده‌اند. از مجموعه قابل دسترس ویژگی‌ها، FFOA برای انتخاب توابع بهینه به کار گرفته شد. توابع از طریق مجموعه داده‌های پزشکی برای پیش پردازش انتخاب شدند و به ماشین بردار پشتیبان چند هسته برای طبقه‌بندی داده‌های بالینی داده شدند. پیامد تحلیلی مشخص می‌کند که روش طبقه‌بندی CKD ارائه شده به ارزش دقیق طبقه‌بندی می‌رسد.

نیتیا و همکاران CCKD-KKMC-NNC-CSOA را برای طراحی و بهبود روشی برای تشخیص CKD با استفاده از مجموعه‌ای از طبقه بندی پیشنهاد کرده‌اند. از جمله روش خوشه بندی، در عمل و تشکیل درمان برای تصاویر سونوگرافی و ام آر آی، شناسایی سنگ کلیه با قطعه بندی تکنیک ضروری بود. در تصاویر سونوگرافی، تقسیم سنگ کلیه اغلب به صورت فیزیکی در آموزش پزشکی انجام می‌شد. با استفاده از طبقه‌بندی کننده شبکه عصبی، تصاویر به‌عنوان عادی و غیرعادی طبقه‌بندی شدند. تصاویر غیرطبیعی به فاز تقسیم‌بندی برای تقسیم تومور با ناحیه سنگ به صورت جداگانه با استفاده از رویکرد خوشه‌بندی چند کرنل K-means مشخص شدند.

Pradeepa و Jeyakumar یک الگوریتم خوشه‌بندی طیفی خود تنظیم کارآمد برای پیش‌بینی بیماری مزمن کلیوی ارائه

ضروری است. برای این منظور، داده کاوی تأثیر قابل توجهی برای افزایش تحقیقات پزشکی می‌گذارد. علاوه بر این، رویکردها به متخصصان بالینی در فرآیند تصمیم گیری کمک می‌کنند. بسیاری از داده های مراقبت های بهداشتی با ابزارهای پایگاه داده بزرگ ترکیب می‌شوند. حالت های تحلیلی به داده های جمع آوری شده برای حمایت از پیش بینی دقیق بیماری ها در سیستم های مراقبت بهداشتی بستگی دارد. علاوه بر این، چالش های مختلفی در طول تجزیه و تحلیل داده های بزرگ از طریق برنامه های کاربردی مختلف مراقبت های بهداشتی به وجود می‌آید. این اشکالات باعث الهام گرفتن این کار شده است.

یک روش خوشه‌بندی IF-RFKM در این مقاله برای پیش‌بینی و طبقه‌بندی بیماری مزمن کلیوی پیشنهاد شده است. این بیماری با استفاده از طبقه‌بندی کننده XG boost طبقه‌بندی می‌شود تا مراحل بیماری‌های مزمن کلیوی را به صورت نرمال و غیرطبیعی طبقه‌بندی کند. علاوه بر این، برای بهینه‌سازی پارامترهای طبقه‌بندی کننده تقویت XG، الگوریتم بهینه‌سازی ازدحام موش‌ها پیشنهاد شده است.

مشارکت های اصلی این مقاله عبارتند از:

- پیش‌بینی و طبقه‌بندی بیماری‌های مزمن کلیوی مبتنی بر داده‌های بزرگ Hadoop با استفاده از خوشه‌بندی IF-RFKM و بهینه‌سازی ازدحام موش‌های صحرائی تقویت کننده (IF-RFKM-XG-RSO) XG پیشنهاد شده است.
- حالت تحلیلی مبتنی بر داده های بزرگ پیشنهادی، تکنیک داده کاوی پیشرفته از جمله خوشه بندی، یادگیری ماشین، رویکردهای بهینه سازی را ترکیب می‌کند.
- روش خوشه‌بندی IF-RFKM برای پیش‌بینی بیماری مزمن کلیوی در نظر گرفته می‌شود.
- این بیماری با استفاده از طبقه‌بندی XGboost 21 برای طبقه‌بندی مراحل بیماری‌های مزمن کلیوی به‌عنوان نرمال و غیر طبیعی طبقه‌بندی می‌شود.
- برای بهینه‌سازی پارامترهای طبقه‌بندی کننده تقویت XG، الگوریتم بهینه‌سازی ازدحام موش‌ها 22 پیشنهاد شده است.
- داده ها به طور تصادفی از مجموعه داده های بیماری مزمن کلیه تولید می‌شوند.

Jongbo و همکاران رشد روش گروهی را برای تجزیه و تحلیل بیماری مزمن کلیوی ارائه کرده اند. که در آن، 2 روش مشترک گنجانده شد یعنی بسته بندی و زیرفضای تصادفی بر روی 3 یادگیرنده پایه، بهبود عملکرد مجموعه فن آوری ها، از جمله K-Neighbors، Nav Bay و درخت تصمیم. برای رسیدگی به مقادیر از دست رفته، پیش پردازش داده ها و همچنین افزایش داده ها برای عادی سازی انواع الگوهای آزاد شده پذیرفته شد.

### 3- روش پیشنهادی

در این مقاله، پیش بینی و طبقه بندی بیماری های مزمن کلیوی مبتنی بر داده های بزرگ Hadoop با استفاده از خوشه بندی IF-RFKM و بهینه ساز ازدحام موش های تقویت کننده -CCKD (XG RSO) پیشنهاد شده است. بلوک دیاگرام روش پیشنهادی CCKD-IF-RFKM-XG-RSO در شکل 1 آورده شده است.

#### داده کاوی:

داده های بزرگ برای ذخیره حجم زیادی از داده ها برای هدف معین ذخیره می شود. این استراتژی های داده کاوی عمدتاً در زمینه پزشکی به کار می روند، که به پزشکان امکان می دهد در حین پردازش تصمیم گیری، اطلاعات مربوطه را به دست آورند. الگوها توسط فرآیند داده کاوی از داده ها استخراج می شوند. بر اساس ویژگی های داده کاوی، الگوها برای مجموعه داده ایجاد می شوند. توصیفی و پیشگویی دو جنبه مهم داده کاوی هستند که در آنها توصیفی به شناسایی مجموعه داده های معمولی کمک می کند. پیش بینی کننده بسیاری از پیش بینی ها را از نظر مجموعه داده های قابل دسترس انجام می دهد. داده کاوی به پارامترهای خاصی نیاز دارد، یعنی تجزیه و تحلیل الگو، طبقه بندی، خوشه بندی، توالی، ارتباط برای ارزیابی داده ها، که در آن، تداعی نشان دهنده الگوهایی است که ارتباط بین داده ها را تعیین می کند، به همین ترتیب، طبقه بندی نشان دهنده الگوهای جدید از نظر اهداف از پیش تعریف شده است. این خوشه ها شامل اجسام کوچکی است که شبیه یکدیگر هستند.

#### K-means خشن کسری بهبود یافته برای خوشه بندی:

در این بخش اطلاعات کلان داده های طولی را توصیف می کند که از روش های تصویربرداری غیرتهاجمی با استفاده از رویکرد K-means فازی کسری (IF-RFKM) بهبود یافته خوشه بندی

کرده اند. یک خوشه بندی طیفی خود تنظیم برای کاهش اندازه گیری بالا و مسائل اضافی در داده ها ارائه شد. قرار گرفتن در معرض استخراج همراه با ایجاد اطلاعات ضروری شناخته شده از داده های بیماران بالینی و همچنین آزمایشگاهی بود که حداکثر حمایت برای کمک به جراحان در افزایش دقت شناسایی CKD قبلاً بر سطح جدی تأثیر می گذاشت. نتایج خوشه بندی از روش های یادگیری ماشینی استفاده شد، یعنی ماشین بردار پشتیبان، شبکه عصبی مصنوعی، شبکه عصبی عمیق، نزدیک ترین همسایه K و جنگل تصادفی مورد استفاده قرار گرفته است.

Akben سطح اولیه CKD را با استفاده از مدل های داده کاوی برای آزمایش ادرار، آنالیز خون، وقایع بیماری ارائه کرده است. روش ارائه شده از تجزیه و تحلیل پزشکی با استفاده از نتایج مربوط به بیماری، ادرار و آزمایش خون پشتیبانی می کند. برای طبقه بندی الگوریتم از تکنیک داده کاوی استفاده شد. مرحله اولیه روش استخراج، پیش پردازش بر روی داده های CKD انجام شد. پیش پردازش را می توان با روش خوشه بندی K-means انجام داد. برای تشخیص CKD، مدل های طبقه بندی، مانند KNN، Naïve Bayes، SVM برای داده های از پیش پردازش شده کاربردی بودند.

Aldhyani و همکاران خوشه بندی نرم را برای تقویت تحلیل بیماری های مزمن در الگوریتم های یادگیری ماشین ارائه کرده اند. هدف توسعه روش بررسی قرار گرفتن در معرض بیماری های مزمن بود که برای امنیت بقای مردم استفاده می شد. روش ارائه شده برای بهبود شناسایی بیماری مزمن با استفاده از الگوریتم یادگیری ماشینی ایجاد شده است.

Khamparia و همکاران طبقه بندی بیماری مزمن کلیوی را با یادگیری داده های چند رسانه ای با استفاده از سیستم رمزگذار خودکار پشته ای عمیق ارائه کرده اند. مدل یادگیری عمیق جدیدی برای طبقه بندی CKD با استفاده از رمزگذار خودکار پشته ای با طبقه بندی کننده Softmax پیشنهاد شد. برای استخراج توابع ارزشمند از طریق پایگاه داده، رمزگذار خودکار انباشته شده در نظر گرفته که توسط طبقه بندی کننده Softmax برای محاسبه جلسه تکمیل استفاده شد. روش The multi mode نسبت به طبقه بندی های محافظه کار تخمین زده شده با دقت طبقه بندی برای بیماری مزمن کلیوی دقیق تر تلقی می شود.

ویژگی دلالت دارد. اگر یک شی با تقریب حداقلی قرار گیرد، شی برای خوشه های دیگر رقابت نمی کند. بنابراین، این درجه عضویت فازی شی در حداقل تقریب با 1 تخصیص داده شده و عضویت فازی جدید در معادله نمایش داده میشود.

$$\mu_{mk} = \begin{cases} \frac{1}{\sum_{j=1}^l (h_{mk}/h_{jk})^{\frac{2}{n-1}}} & \text{if } A_k \in \hat{D}_m, \\ 1 & \text{if } A_k \in \bar{D}_m, \end{cases}$$

که در آن  $h_{mk}$  به معنای حداقل فاصله بین  $A_k$  در امتداد کل مراکز خوشه است. اگر فاصله  $h_{jk}$  بین  $A_k$  وجود داشته باشد، انحراف فاصله  $h_{jk}-h_{mk}$  کمتر از آستانه فعلی است و  $D_m$  نشان دهنده تعداد خوشه است، عبارت فوق دارای مرزهای فازی واضح در خوشه RFKM با تقریب کمتر است. اندازه گیری فازی محلی، شی مرزی را محاسبه می کند که متعلق به خوشه های کمتر مرتبط است، در حالی که معیار فازی سراسری، هر شی را که به همه خوشه ها تعلق دارد، محاسبه می کند. اگر شی  $A_2$  در بین خوشه های 1 و 2 در محلی واقع شده باشد، عضویت فازی متعلق به خوشه های 1، 2 است که به صورت محلی ارتقا یافته اند. این برای افزایش فازی محلی پاسخگو است. محاسبه مرکز خوشه تکراری در معادله نمایش داده شده است.

$$u_m = \begin{cases} T_g \times \frac{\sum_{A_k \in D_m} \mu_{mk}^n A_k}{\sum_{A_k \in D_m} \mu_{mk}^n} + T_f \times \frac{\sum_{A_k \in \hat{D}_m} \mu_{mk}^n A_k}{\sum_{A_k \in \hat{D}_m} \mu_{mk}^n} & \text{if } D_m \neq \phi \wedge \hat{D}_m \neq \phi, \\ \frac{\sum_{A_k \in D_m} \mu_{mk}^n A_k}{\sum_{A_k \in D_m} \mu_{mk}^n} & \text{if } D_m \neq \phi \wedge \hat{D}_m = \phi, \\ \frac{\sum_{A_k \in D_m} \mu_{mk}^n A_k}{\sum_{A_k \in D_m} \mu_{mk}^n} & \text{if } D_m = \phi \wedge \hat{D}_m \neq \phi, \end{cases}$$

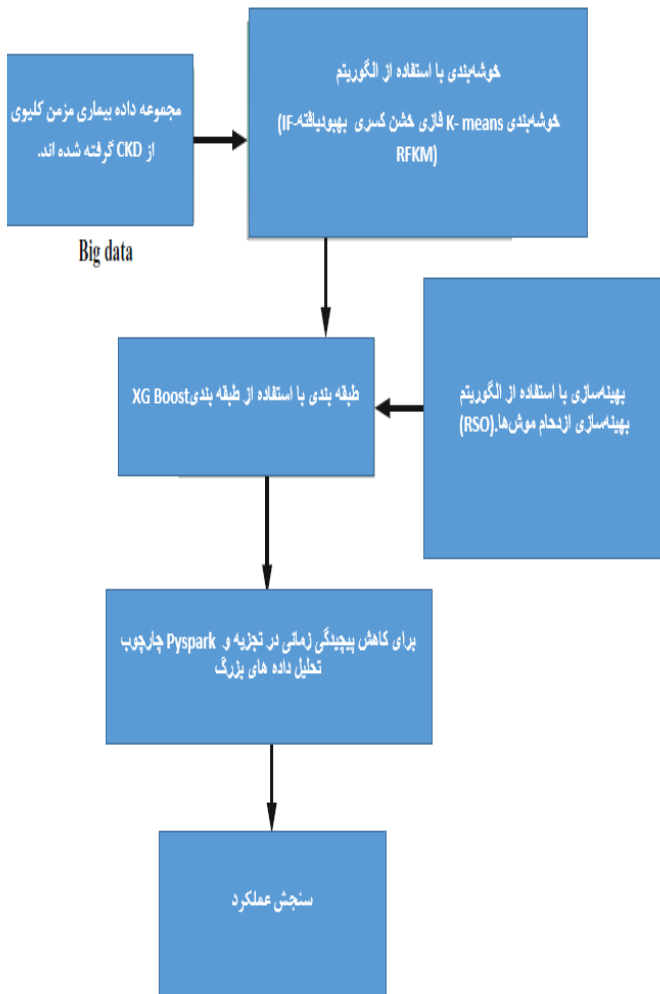
که در آن  $T_g$  و  $T_f$  نشان دهنده ضرایب وزنی اهمیت نسبی  $T_g > T_f$  و  $T_g + T_f = 1$  مناطق مرزی معمولاً پایین تر است. پس از برخی بهبودهای پیشرونده، روش کلاسیک Rough K-Means انجام شد،  $A_k$  به شی در فاصله  $h_{mk}$  از مرکز  $u_m$  خوشه  $D_m$  نیاز دارد. اشیاء در منطقه محدود کمتر مربوطه در امتداد درجات

شده اند. IF-RFKM یک روش شناخته شده است که در پیش بینی بیماری استفاده می شود.

IF-RFKM سریعتر از سایر روشهای خوشه بندی است، به خصوص اگر خوشه ها کروی باشند. از این رو، تعداد زیادی پیکسل با کمک IF-RFKM خوشه بندی می شوند. داده ها به عنوان مجموعه ای از خوشه جدا می شوند. در اینجا، پارامترهای ورودی داده شده به واحد شفاف سازی ویژگی هایی را نشان می دهد که برای پیش بینی وجود یا عدم وجود بیماری مزمن کلیوی در نظر گرفته می شود. که در آن 25 ویژگی مانند فشار خون، جاذبه دقیق، آلومین، قند، سلول چرک، گلبول های قرمز، توده های سلول چرک، باکتری برای پیش بینی بیماری مزمن کلیوی در نظر گرفته می شود. سپس، متغیرهای خروجی از واحد فاززدایی به پیش بینی وجود یا عدم وجود بیماری مزمن کلیوی اشاره دارد. پارامترهای فازی نتیجه به طور خودکار در IF-RFKM تنظیم می شوند. پارامترهای توابع عضو را می توان با توجه به دامنه واریانس داده های ورودی همراه با نامشخص مشخص کرد. محاسبه فاصله اقلیدسی توسط درجات عضویت فازی که با معیارهای خاصی آغاز شده اند تغییر می کنند. رویکرد خوشه بندی K-means با جداسازی داده ها عمل می کند، زیرا مرکزها توسط مجموعه ای از خوشه ها تعیین می شوند. در این، خوشه های 1، 2 در امتداد اشیاء مرزی مرتبط با آن نشان داده شده اند، از این نقطه، شی مرزی خوشه 1 با موقعیت خوشه مرتبط با آن قرار دارد. اما، شی مرزی خوشه 2 در منطقه ای در میان خوشه 1، 2 قرار می گیرد. این ها به طور تصادفی از مجموعه داده ها تولید می شوند. داده های بزرگ برای پیش بینی بیماری مزمن کلیه به طور تصادفی از مجموعه داده های CKD تولید می شود. درجه عضویت فازی توسط فرمول زیر محاسبه می شود.

$$\mu_{mk} = \frac{1}{\sum_{j=1}^l (h_{mk}/h_{jk})^{\frac{2}{n-1}}},$$

که در آن پارامتر  $n$  تعداد فازی فایر (fuzzifier) را نشان می دهد و  $\exists n \in (1, \infty)$  درجات فازی مختلفی را ایجاد می کند، پارامتر  $\mu \in [0, 1]$  و  $\sum_{m=1}^l \mu_{mk} = 1$ ، پارامتر  $h_{mk}$  ماتریس فاصله را نشان می دهد. درجات عضویت فازی محدوده تغییرات ورودی را تعیین می کند که نشان دهنده ویژگی های پیش بینی بیماری مزمن کلیوی و سطح عدم قطعیت است که بر سطح تنوع مقادیر



شکل 1 نمودار بلوکی برای حالت تحلیلی مبتنی بر داده های بزرگ پیشنهادی برای پیش بینی بیماری کلیوی با طبقه بندی

اجازه دهید پارامتر  $v$  بردار ورودی را مشخص کند،  $b$  بردار خروجی را از طریق طرح فازی مشخص کند،  $bc$  بردار پارامترهای بعدی را مشخص کند.  $\vartheta$  برای کاهش پیچیدگی زمانی در تجزیه و Pyspark چارچوب تحلیل داده های بزرگ

$$C_a^{1-p}h(a) = \frac{1}{\mathfrak{S}(1-p)} \int_0^a (a-\tau)^{-p} \frac{c}{ca} h(\tau) c \tau,$$

مشتق  $h$  به عنوان  $p$  است، که در آن  $\mathfrak{S}(1-p)$  نشان دهنده تابع گاما و  $h(a)$  نشان دهنده فاصله نرمال شده بین عنصر داده است،  $h(\tau)$  نشان دهنده اشیاء در ناحیه مرزی است. پارامترهای متعاقب  $\vartheta$  تنظیم برای کاهش زمان محاسبات و همچنین به دست آوردن تصمیم بهتر در معادله نشان داده شده است.

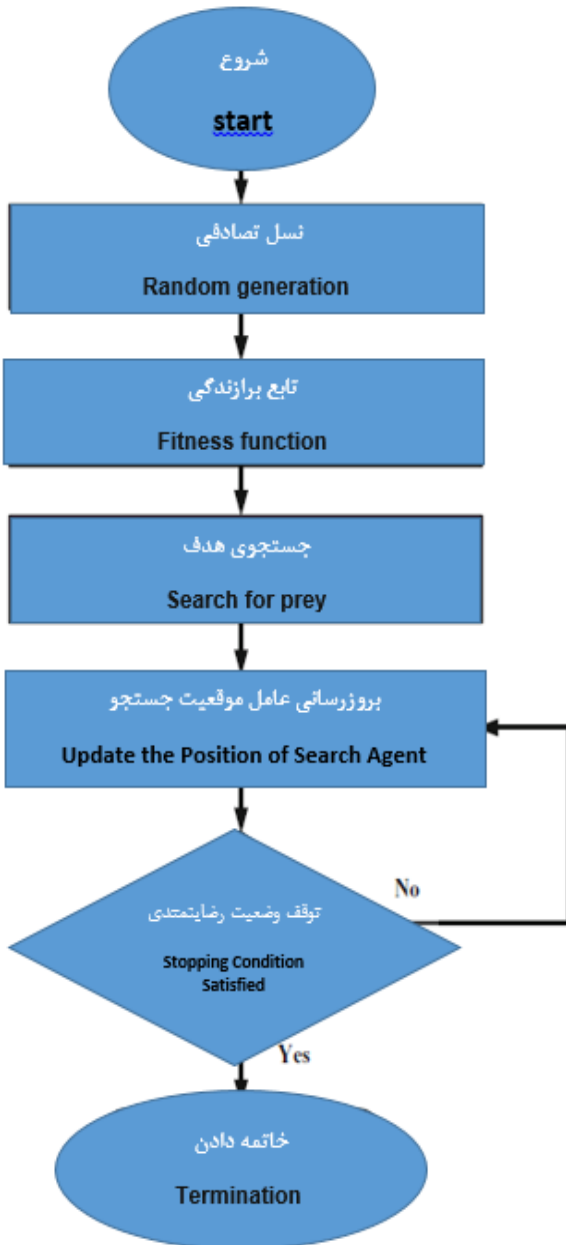
عضویت فازی متنوع قرار دارند. با استفاده از IF-RFKM، پارامترهای خروجی از طریق سیستم فازی به دست می آیند. همچنین قوانین انطباق خاصی نتیجه تجزیه و تحلیل پایداری است که سطح لغزش را به عنوان "صفر" همگرا می کند.

$$R = D_a^{1-p}F + \alpha D_x^{-p}F.$$

در فرمول بالا،  $R$  تعداد خوشه های همپوشانی را نشان می دهد و  $\alpha$  نشان دهنده پایداری مثبت، پارامتر  $F$  نشان دهنده خطای تخمین است. سپس  $D_x$  و  $D_a$  نشان دهنده مرز پایین و بالای خوشه است. سپس خطای تخمین به صورت معادله محاسبه می شود.

$$F = b(\vartheta/v) - b_c.$$

بلوک دیاگرام الگوریتم پیشنهادی:



شکل 2 فلوچارت برای الگوریتم بهینه سازی ازدحام موش برای بهینه سازی افزایش تحرک شدید سپس بهترین نقطه تقسیم تخمین زده می شود. پس از آن، مقادیر وزن خاصی را برای هر گره برگ جدید تعیین کنید. در نهایت، گره های با سود منفی با فرآیند هرس حذف می شوند. اجازه می دهیم مجموعه ویژگی  $U_i$  شامل بردارهای ویژگی  $IR_i$  باشد.  $U_1, U_2, \dots, U_m$ , here,  $i = 1, 2, \dots, m$  امتیاز اهمیت ویژگی را تخمین بزنید بر اساس اینکه هر ویژگی چند بار برای توزیع داده های آموزشی به صورت مورب در کل درختان استفاده می شود. نمره اهمیت ویژگی به صورت معادله داده می شود.

$$\vartheta = [t_1, \dots, t_N, \bar{t}_1, \dots, \bar{t}_N]^W.$$

در اینجا پارامتر تعداد قوانین را مشخص می کند. محدوده متغیر ورودی به بخش ها جدا می شود. درجه عضویت برای هر بخش در نظر گرفته می شود. پارامترهای درجه عضویت با توجه به محدوده متغیرهای ورودی در امتداد سطح نامشخص مورد بررسی قرار می گیرند. مقدار کلی قوانین معادل مقدار کلی درجات عضویت است. روش IF-RFKM مزایای مجموعه های فازی مرتبه کسری و شبکه خوشه بندی RFKM را ارائه می دهد. این امکان خوشه بندی کلان داده ماهرانه را در حین پردازش تصمیم گیری فراهم می کند.

### طبقه بندی XGboost برای طبقه بندی بیماری مزمن کلیدی:

در این مقاله طبقه بندی XGboost برای طبقه بندی بیماری مزمن کلیدی مورد بحث قرار گرفته است. XGBoost انتخاب شده است، زیرا بر تعدادی از معایب ناشی از روش های موجود غلبه می کند. به همین دلیل است که XGBoost به عنوان مدل طبقه بندی ترجیحی هنگام برخورد با مسائلی که با مشکلات طبقه بندی کلمه واقعی مواجه هستند انتخاب می شود. XGBoost تقریباً 10 برابر سریع تر از روش های موجود در یک پلتفرم است. انعطاف پذیری ارائه شده توسط XGBoost بسیار زیاد است و با یک سایت خاص ادغام نشده است، بنابراین، طبقه بندی CKD با کمک XGBoost ممکن است مستقل از پلتفرم باشد، این یک مزیت قابل توجه است. منظم سازی یک ویژگی حیاتی الگوریتم XGBoost است، زیرا به جلوگیری از مشکلات بیش از حد داده ها کمک می کند. تکنیک تقویت گرادیان شدید (XG boost) یکی از کارآمدترین روش های مقیاس پذیر مدل های یادگیری ماشینی با استراتژی تقویت درخت است.

تقویت XG با استفاده از مجموعه ای از چندین درخت تصمیم ساخته شده است که به طور موثر مشکل اضافه برازش را در طول طبقه بندی حل می کند. در ابتدا، یک درخت تصمیم با چندین گره برگ ساخته می شود. هر گره یک ویژگی واحد را تعریف می کند که برای انجام طبقه بندی ضروری است.

پیچیدگی و جلوگیری از مشکل بیش از حد برآزش در طول طبقه بندی توسط معادله ارائه شده است.

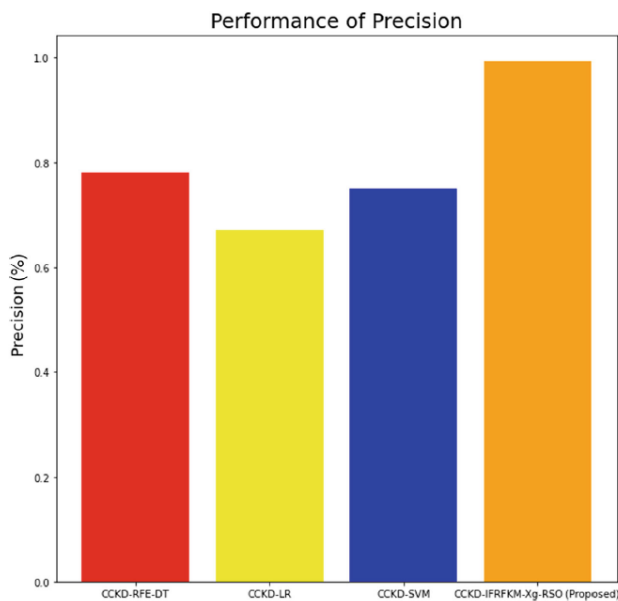
$$\Omega = \Psi^W + \frac{1}{2} \lambda \sum_{j=1}^W T_j^2,$$

در جایی که پارامتر تعداد برگها را در درخت تصمیم نشان می دهد، پارامتر  $T_j^2$  امتیاز روی برگ  $j$  را نشان می دهد، نشان دهنده تابع تلفات محدب قابل تشخیص است که تفاوت را در میان پیش بینی اندازه گیری می کند، و  $\Psi$  دلالت بر دنباله ای پیوسته از یک سری زمانی دارد. بنابراین، تابع هدف "Z" مدل تقویتی XG پیشنهادی در معادله بیان شده است.

$$Z = K + \Omega.$$

در طبقه بندی کننده تقویت XG، میانگین و انحراف استاندارد به عنوان نزول گرادیان برای بهینه سازی تابع هدف استفاده می شود. مدل تقویت XG در معادله زیر بیان شده است.

$$Z(w) = \sum_{i=1}^n \left[ h_i q_w(A_i) + \frac{1}{2} g_i q_w^2(A_i) \right] + \Omega(q_w),$$



شکل 4 عملکرد صحت

در اینجا پارامتر  $A_i$  نشان دهنده بردار ویژگی در نقطه داده است و پارامتر  $n$  تعداد تکرارهای انجام شده برای دستیابی به حداکثر عمق را نشان می دهد. تابع هدف تقویت XG به عنوان معادله تعریف می شود.

$$F_{obj}(\theta) = K(\theta) + \Omega(\theta),$$

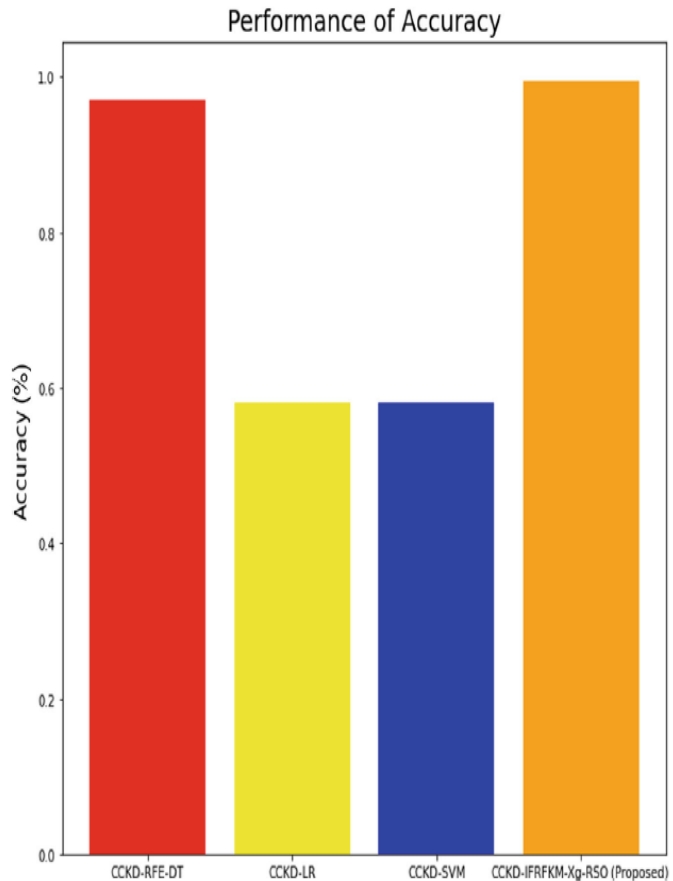
$$K(\theta) = k(\hat{b}_i, b_i), \Omega(\theta) = \mathfrak{F} + \frac{1}{2} \lambda \|T\|^2,$$

$$IR_i = \{A | A = T_i U_i\}.$$

در اینجا پارامتر  $T_i$  مقدار وزن را برای هر ویژگی مشخص می کند و پارامتر  $U_i$  مجموعه ویژگی ها را نشان می دهد، خروجی طبقه بندی در نقطه داده در معادله بیان می شود.

$$B'_i = \sum_{m=1}^m q_i,$$

در اینجا پارامتر تعداد درختان تصمیم را نشان می دهد و پارامتر  $q_i$  خروجی طبقه بندی از درخت تصمیم را نشان می دهد. سپس، توابع ضرر برای آموزش مدل تقویت XG برای دستیابی به بهترین نتایج طبقه بندی بهینه می شود و عملکرد از دست دادن برای طبقه بندی کلیه ها با کارکرد طبیعی مشخص می شود.



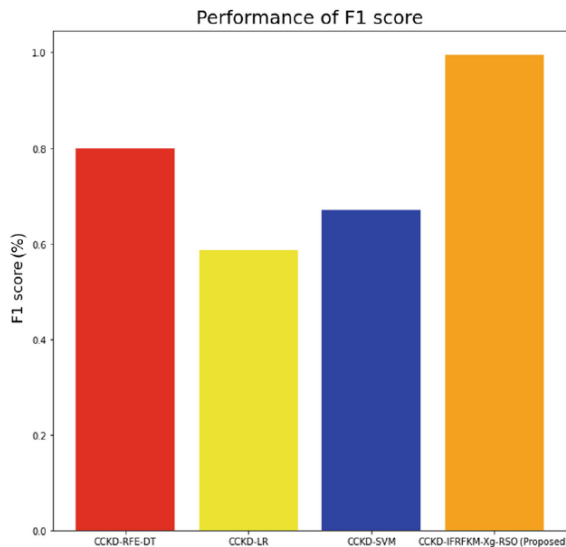
شکل 3 عملکرد دقت

بیماری در معادله زیر نشان داده شده است

$$K = -\frac{1}{N} \sum_i^N ((B_i \log e_i) + (1 - B_i) \log (1 - e_i)),$$

در جایی که پارامتر  $K$  تابع ضرر را نشان می دهد که تخمین می زند که مدل طبقه بندی پیشنهادی چقدر پیش بینی می کند،  $B_i$  نشان دهنده زمان جدی تابع ضرر است. تابع تنظیم برای کنترل





شکل 5 نتیجه میزان کارایی F1

Save translation

اولین جمعیت موش را به صورت  $q_i$  راه اندازی کنید، و برای  $i = 1, 2, \dots, m$  مقاداردهی اولیه جمعیت لانه های میزبان  $bi$  تا جایی که

$$q = X.q_i(a) + D.(q_s(a) - q_i(a)).$$

اجازه دهید  $X$  موقعیت موش ها را مشخص کند،  $q_i(a)$  و  $q_s(a)$  حد پایین و بالای منطقه جستجو را در بعد  $a$  مشخص می کند،  $D$  عدد تصادفی یکنواخت را در محدوده  $[0, 1]$  مشخص می کند.

### مرحله 2: تولید تصادفی

هنگامی که فرآیند مقاداردهی اولیه انجام می شود، پارامترهای ورودی به صورت تصادفی ایجاد می شوند. حداکثر مقادیر برازندگی بسته به شرایط دقیق فوق پارامتر تعیین می شود. سپس به صورت تصادفی جمعیت ارزش فرآیند را برای کاهش نرخ خطا ایجاد می کند.

Save translation

### مرحله 3: عملکرد برازندگی

در این قسمت برای دستیابی به تابع هدف، مانند کاهش نسبت خطا، پیچیدگی محاسباتی برای افزایش دقت مزمن ارزیابی می شود همچنین در اینجا، پارامترهای  $K(\theta)$  و  $\Omega(\theta)$  با استفاده از RSO بهینه می شوند.

### مرحله 4: جستجوی طعمه (هدف)

در فرمول بالا  $K(\theta)$  و  $\Omega(\theta)$  پارامترهای تنظیم شده در طبقه بندی کننده تقویت XG هستند، سپس تفاوت بین نرخ پیش بینی  $\hat{bi}$ ،  $bi$  است، همچنین  $\hat{K}$  نشان دهنده افزایش اطلاعات و XGboost این ترم 12 افزایش می یابد  $2 || T || \lambda$  مرتبط با الگوریتم متعارف GBD،  $\lambda$  پارامتر منظم شده را نشان می دهد،  $T$  دلالت بر وزن برگها دارد. به حداکثر رساندن آیتم می تواند از برازش جلوگیری کند و توانایی تعمیم ماژول را افزایش دهد. خروجی های طبقه بندی از طبقه بندی کننده تقویت XG پیشنهادی تحت مقادیر آستانه مختلف برای به دست آوردن نتایج طبقه بندی بهینه آزمایش می شوند.

### بهینه سازی پارامترها با استفاده از الگوریتم بهینه سازی ازدحام موشها

بهینه سازی ازدحام موشها (RSO) برای تنظیم پارامترهای طبقه بندی کننده تقویت XG برای دستیابی به پارامترهای بهینه استفاده می شود. این پارامترها برای دستیابی به دقت بیشتر با کاهش نرخ خطا تنظیم می شوند. الگوریتم RSO یک الگوریتم فراابتکاری است که از تعادل بهینه در میان مراحل اکتشاف و بهره برداری استفاده می کند.

این یک الگوریتم فراابتکاری مبتنی بر جمعیت است. در این مقاله، به روزرسانی با راه اندازی عوامل جستجو (موش) ایجاد می شود و توسط مجموعه ای از قوانین در تکرارهای تعیین شده ایجاد، حذف، یا اصلاح می شود. از آنجایی که به خودی خود پیشرفت کرده است، نسبت به روش تنظیم اضافی، یعنی فاز با جستجوی تصادفی، زمان تکرار کمتری را می طلبد، سپس پارامترهای هایپر بهینه را تنظیم می کند.

شکل 2 فلوچارت الگوریتم RSO را برای بهینه سازی تقویت XG نشان می دهد. فرآیند گام به گام RSO به شرح زیر است: مرحله 1: مقاداردهی اولیه

#### شکل 6 عملکرد و کارایی فراخوان

##### مرحله 6: خاتمه

فراپارامتر بهینه  $K(\vartheta)$  و  $\Omega(\vartheta)$  در تقویت XG با استفاده از RSO انتخاب می شوند که می تواند مرحله 3 را به طور مکرر تکرار کند تا زمانی که معیارهای توقف برآورده شود. در نهایت، طبقه بندی کننده تقویت XG بهترین طبقه بندی بیماری مزمن کلیوی بهینه پویا را پیدا می کند تا با کاهش پیچیدگی محاسباتی با استفاده از رویکرد RSO، دقت را افزایش دهد.

#### 4- نتیجه و بحث

در این مقاله، پیش بینی و طبقه بندی بیماری های مزمن کلیوی مبتنی بر داده های بزرگ Hadoop با استفاده از بهینه ساز ازدحام موش های صحرایی IF-RFKM و XGboost ارائه شده است. روش پیشنهادی در پایتون با استفاده از Keras، Scikit-learn، OpencvlibraryonIntelXeon با استفاده از 128 گیگابایت حافظه RAM فعال شد. یک کارت گرافیک NVIDIA GeForce GTX 1080 Ti همچنین عملکرد روش پیشنهادی با روش های موجود CCKD-RFE-DT، CCKD-LR و CCKD-MKC-31.23 و SVM مقایسه شده است. همچنین داده ها از مجموعه داده CKD جمع آوری شده است. داده ها به صورت تصادفی برای آموزش و همچنین آزمایش تولید می شوند. که در آن 124800 داده گرفته شده است. از این تعداد 80 درصد برای فرآیند آموزش و 20 درصد برای فرآیند تست استفاده می شود.

#### عملکرد اندازه گیری

معیارهای کاربردی و عملکرد، دقت، اندازه، صحت، امتیاز F1، میزان خطا تحلیل به شکل زیر تحلیل می شوند.

رفتار حرکت موش ها در طول جست و جو به صورت ریاضی با استفاده از استراتژی به روز رسانی موقعیت در معادله نشان داده شده است.

$$X = S - a \times \left( \frac{S}{Max_{iter}} \right),$$

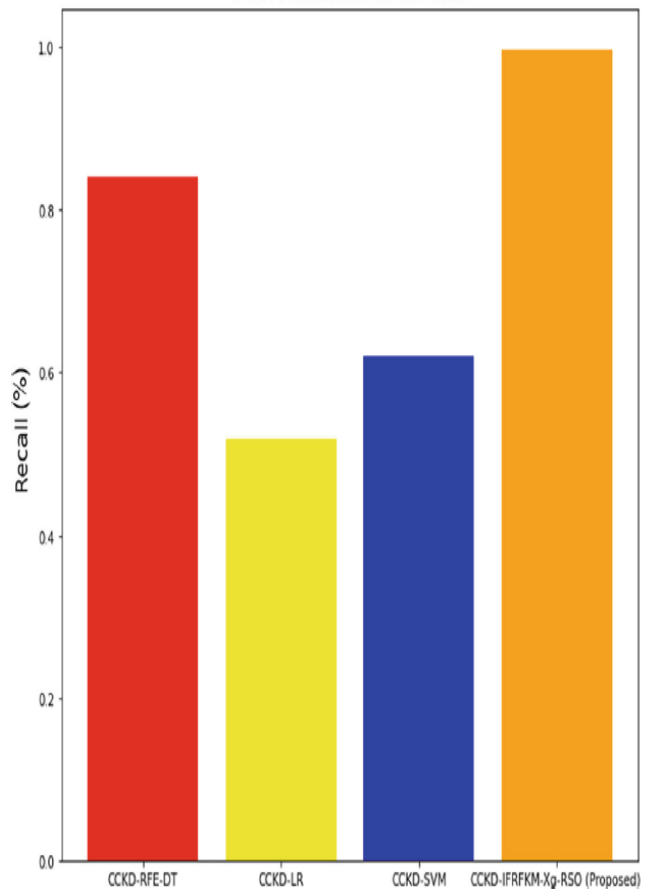
که در آن  $Max_{iter}$ ،  $a = 2, 1, 0, \dots$  و سپس S عدد تصادفی را نشان می دهد. سپس پارامتر X نشان دهنده اکتشاف و بهره برداری بهتر در طول تکرار است.

#### مرحله 5: بروز رسانی موقعیت عامل جستجو

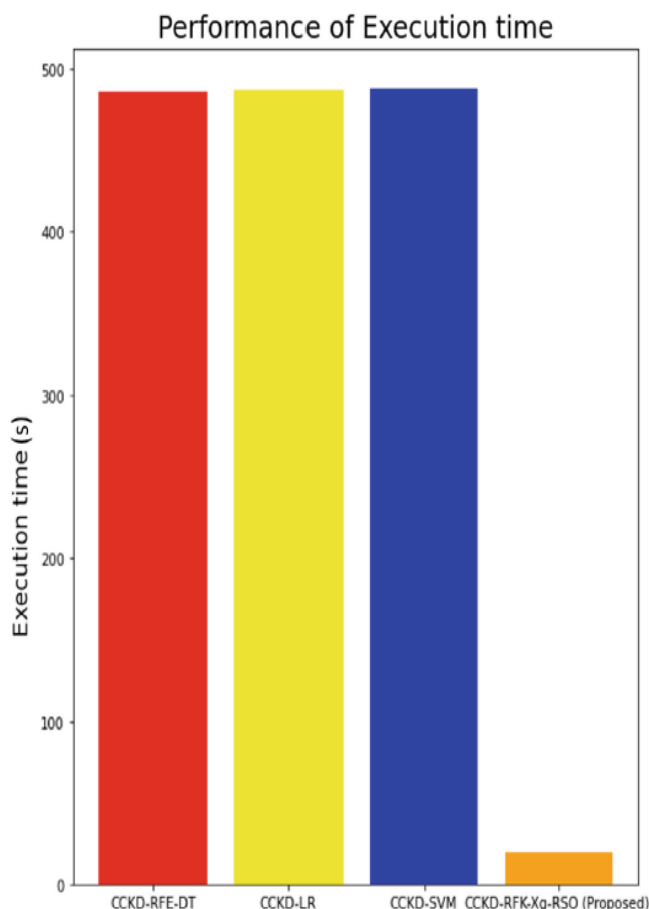
برای ارتقای مکان عوامل جستجو با بکارگیری معادله زیر از آن استفاده می کنیم. در فرمول زیر  $q_i(a + 1)$  نشان دهنده به روز رسانی موقعیت بعدی است. راه حل بهینه را با ارتقای موقعیت های دیگر عامل جستجو در مورد عامل جستجوی بعدی ذخیره می کند. سپس اکتشاف و بهره برداری با تنظیم پارامترهای X و D تضمین می شود.

$$q_i(a + 1) = |q_s(a) - q|,$$

Performance of Recall



$$\text{Precision} = \frac{TP}{TP + FP}$$



شکل 8 عملکرد زمان اجرا

### فراخوانی

فراخوانی و یادآوری بر نسبت طبقه بندی که دقیقاً شناسایی شده است دلالت دارد که بر اساس فرمول زیر محاسبه می گردد.

$$\text{Recall} = \frac{TPR}{TPR + FNR}$$

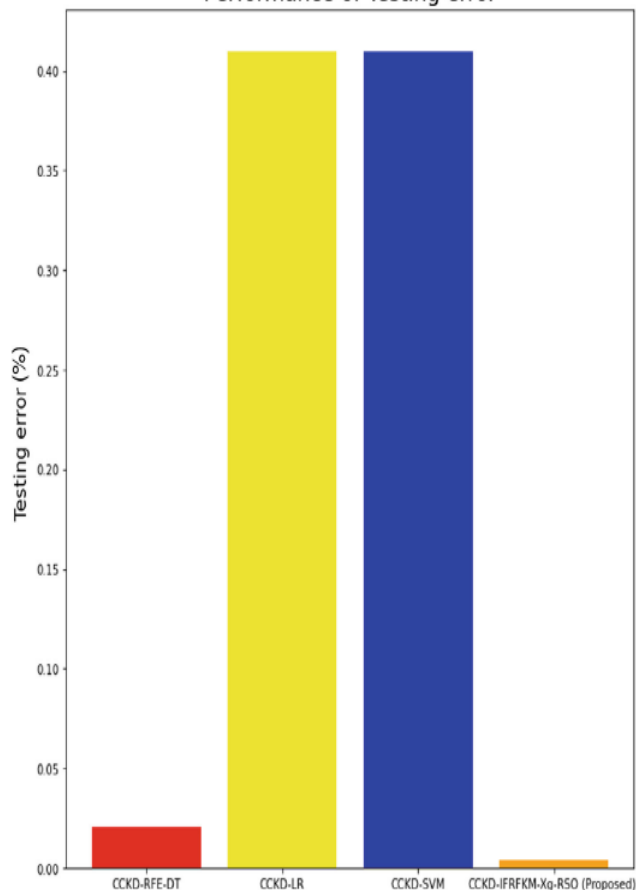
### بدست آوردن امتیاز F1

این مقیاس دقت با فراخوانی و میانگین هارمونیک دقت را با فراخوانی نشان می دهد که در معادله زیر بیان می شود.

$$\text{F-measure} = \frac{2TP}{2TP + FP + FN}$$

تحلیل و مقایسه تحلیل عملکرد کلیه مزمن

Performance of Testing error



شکل 7 عملکرد میزان خطا

- مثبت واقعی (TP): نرمال به درستی به نرمال تشخیص داده می شود.
- منفی واقعی (TN): غیر طبیعی به درستی به غیر طبیعی تشخیص داده می شود.
- مثبت کاذب (FP): غیر طبیعی به طور نامناسب به نرمال تشخیص داده می شود.
- منفی کاذب (FN): طبیعی به طور نامناسب به غیر طبیعی تشخیص داده می شود.

### محاسبه دقت

این نسبت پیش‌بینی‌های واقعی را در مقابل تعداد کلی موارد نشان می‌دهد. این با استفاده از معادله زیر به دست می آید

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

### محاسبه صحت

بهترین خطا اندازه گیری صحت و درستی است که از طریق فرمول زیر محاسبه می شود.

موجود، مانند CCKD-RFE-DT، CCKD-LR، و CCKD-MKC- SVM مقایسه می شود. در این بیماری مزمن کلیوی، فراخوانی روش پیشنهادی CCKD-IF-RFKM-Xg-RSO 17.45٪، CCKD-IF-RFKM-Xg-RSO 5.54٪، و 13.41٪ بیشتر از روش های موجود مانند CCKD-RFE-DT، CCKD-LR، و CCKD را نشان می دهد.

شکل 7 معیارهای عملکرد نرخ خطا را با روش های مختلف نشان می دهد. در اینجا، میزان خطای پیشنهادی CCKD-IF-RFKM-Xg-RSO اندازه گیری می شود و عملکرد به ترتیب با روش های موجود مانند CCKD-RFE-DT، CCKD-LR، و CCKD-MKC- SVM مقایسه می شود. در این بیماری مزمن کلیوی، میزان خطای روش پیشنهادی CCKD-IF-RFKM-Xg-RSO 75.03٪، CCKD-IF-RFKM-Xg-RSO 55.04٪ و 36.11٪ کمتر از روش های موجود، مانند CCKD-RFE-DT، CCKD-LR، و CCKD-MKC- است.

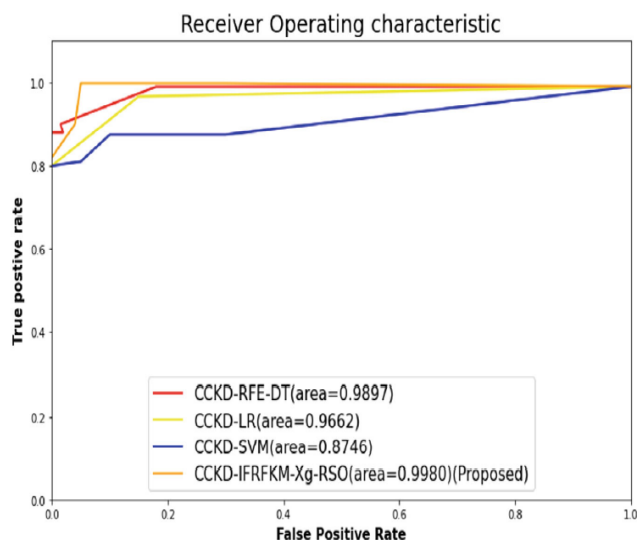
شکل 8 عملکرد زمان اجرا را با روش های مختلف نشان می دهد. در اینجا، زمان اجرای CCKD-IF-RFKM-Xg-RSO پیشنهادی اندازه گیری می شود و عملکرد با روش های موجود، مانند CCKD-RFE-DT، CCKD-LR، و CCKD-MKC-SVM مقایسه می شود. در این بیماری مزمن کلیوی، زمان اجرای روش پیشنهادی CCKD-IF-RFKM-Xg-RSO 95.03٪، CCKD-IF-RFKM-Xg-RSO 94.56٪ و 91.22٪ کمتر از روش های موجود، مانند CCKD-RFE-DT، CCKD-LR، و CCKD-MKC-SVM مقایسه می شود. در شکل 9، ROC پیشنهادی CCKD-IF-RFKM-Xg-RSO 4.6032٪، 2.0265٪ و 6.8033٪ بیشتر از روش موجود مانند CCKD-RFE-DT، CCKD-LR، و CCKD-MKC-SVM است.

## 5- نتیجه گیری

در این مقاله، پیش بینی بیماری CKD را پیشنهاد کرد.

این بیماری با استفاده از طبقه بندی XGboost برای طبقه بندی مراحل بیماری های مزمن کلیوی به عنوان نرمال و غیر طبیعی طبقه بندی می شود. همچنین RSO برای بهینه سازی پارامترهای طبقه بندی کننده XGboost پیشنهاد شده است. در اینجا، داده ها به طور تصادفی از مجموعه داده CKD تولید می شوند. فرآیند شبیه سازی به زبان پایتون اجرا می شود. روش پیشنهادی به دقت بالاتر 95.68٪، 99.32٪ و 91.13٪، امتیاز F بالا 91.57٪، 88.38٪ و 89.02٪، و نرخ خطای بالاتر 94.64٪، 86.32٪، و 85.30٪ در مقایسه با CCKD-RFE موجود می رسد.

در شکل 3 معیارهای عملکرد دقت را با روش های مختلف نشان می دهد. در اینجا، دقت CCKD-IF-RFKM-Xg-RSO پیشنهادی در نظر گرفته شده و عملکردها به ترتیب با چندین روش قبلی مانند CCKD-RFE-DT، CCKD-LR، و CCKD-MKC-SVM مقایسه می شوند.



شکل 9 منحنی ROC برای بیماری مزمن کلیوی

بیماری کلیوی، دقت روش پیشنهادی CCKD-IF-RFKM-Xg-RSO 1.12٪، 6.77٪ و 10.8٪ بالاتر از روش های موجود مانند CCKD-RFE-DT، CCKD-LR، و CCKD-MKC- است.

شکل 4 معیارهای عملکرد دقت را با روش های مختلف نشان می دهد. در اینجا، دقت CCKD-IF-RFKM-Xg-RSO پیشنهادی در نظر گرفته شده و عملکردها به ترتیب با چندین روش قبلی مانند CCKD-RFE-DT، CCKD-LR، و CCKD-MKC-SVM مقایسه می شوند. در این بیماری مزمن کلیوی، دقت روش پیشنهادی CCKD-IF-RFKM-Xg-RSO 11.02٪، 6.37٪ و 1.18٪ بهتر از CCKD-RFE-DT، CCKD-LR، و CCKD-MKC- موجود است.

شکل 5 معیارهای عملکرد امتیاز F1 را با روش های مختلف نشان می دهد. در اینجا، امتیاز F1 CCKD-IF-RFKM-Xg-RSO پیشنهادی است اندازه گیری و عملکرد به ترتیب با چندین روش موجود مقایسه می شود، مانند CCKD-RFE-DT، CCKD-LR، و CCKD-MKC-SVM. در این بیماری مزمن کلیوی، امتیاز F1 روش پیشنهادی CCKD-IF-RFKM-Xg-RSO 6.25٪، CCKD-IF-RFKM-Xg-RSO 7.55٪ و 12.42٪ بهتر از CCKD-RFE-DT، CCKD-LR، و CCKD- موجود است.

شکل 6 معیارهای عملکرد فراخوان را با روش های مختلف نشان می دهد. در اینجا، فراخوان پیشنهادی CCKD-IF-RFKM-Xg-RSO اندازه گیری می شود و عملکرد به ترتیب با روش های

OFDM system by chaotic key generation-based Paillier

cryptography along with LDPC and its side channel estimation using machine learning technique. *J Circuits Syst Comput.* 2022;31(5):2250093.

12. Rajesh P, Shajin FH, Vijaya AN. An efficient estimation model for induction motor using BMO-RBFNN technique. *Process Integr Optim Sustain.* 2021;5(4):777-792.

13. Shajin FH, Rajesh P, Raja MR. An efficient VLSI architecture for fast motion estimation exploiting zero motion prejudgment technique and a new quadrant-based search algorithm in HEVC. *Circuits Syst Signal Process.* 2022;41(3):1751-1774.

14. Rajesh P, Muthubalaji S, Srinivasan S, Shajin FH. Leveraging a dynamic differential annealed optimization and recalling enhanced recurrent neural network for maximum power point tracking in wind energy conversion system. *Technol Econ Smart Grids Sustain Energy.* 2022;7(1):1-5.

15. Pacheco F, Exposito E, Gineste M, Baudoin C, Aguilar J. Towards the deployment of machine learning solutions in network traffic classification: a systematic survey. *IEEE Commun Surv Tutor.* 2018;21(2):1988-2014.

16. Rhee KE, Kessl S, Lindback S, Littman M, El-Kareh RE. Provider views on childhood obesity management in primary care settings: a mixed methods analysis. *BMC Health Serv Res* 2018;18(1):55.

17. Bhosale AS, Jadhav SS, Ahire HS, Jaybhay AY, Rajeswari K. Study of medicine dispensing machine and health monitoring devices. In: Gunjan VK, Suganthan PN, Haase J, Kumar A, eds. *Cybernetics, Cognition and Machine Learning Applications.* Springer; 2021:297-305.

18. Din IU, Guizani M, Rodrigues JJ, Hassan S, Korotaev VV. Machine learning in the internet of things: designed techniques for smart cities. *Future Gener Comput Syst.* 2019;100:826-843.

19. Sharma S, Singh G, Sharma M. A comprehensive review and analysis

1. Kumar S, Singh M. Big data analytics for healthcare industry: impact, applications, and tools. *Proceedings of the Big Data Mining and Analytics;* 2018:48-57; IEEE.

2. Azeroual O, Saake G, Schallehn E. Analyzing data quality issues in research information systems via data profiling. *Int J Inf Manag.* 2018;41:50-56.

3. Ahir RK, Chakraborty B. A meta-analytic approach for determining the success factors for energy conservation. *Energy.* 2021;230:120821.

4. Nemati M, Ansary J, Nemati N. Machine-learning approaches in COVID-19 survival analysis and discharge-time likelihood prediction using clinical data.

*Patterns.* 2020;1(5):100074.

5. Khare N, Devan P, Chowdhary CL, et al. SMO-DNN: spider monkey optimization and deep neural network hybrid classifier model for intrusion detection.

*Electronics.* 2020;9(4):692.

6. Zhao Y, Zhang C, Zhang Y, Wang Z, Li J. A review of data mining technologies in building energy systems: load prediction, pattern identification, fault detection and diagnosis. *Energy Built Environ.* 2020;1(2):149-164.

7. Javaid M, Khan IH. Internet of things (IoT) enabled healthcare helps to take the challenges of COVID-19 pandemic. *J Oral Biol Craniofac Res.* 2021;11(2):209-214.

8. Nadim MI, Alam J, Khanam H. Heart disease prediction using data mining approach; 2019.

9. Sornalakshmi M, Balamurali S, Venkatesulu M, et al. Hybrid method for mining rules based on enhanced Apriori algorithm with sequential minimal optimization in healthcare industry. *Neural Comput Appl.* 2020;31:1-4.

10. Wang J, Zhan Q. Visualization analysis of artificial intelligence Technology in Higher Education Based on SSCI and SCI journals from 2009 to 2019. *Int J Emerg Technol Learn.* 2021;16(8):20-23.

11. Shajin FH, Rajesh P. FPGA realization of a reversible data hiding scheme for 5G MIMO-

Downloaded from <https://iranpaper.ir>  
<https://www.tarjomano.com>  
<https://www.tarjomano.com>  
VIJAYALAKSHMI AND RAMARAJ **15 of 15**  
29. Khamparia A, Saini G, Pandey B, Tiwari S, Gupta D, Khanna A. KDSAE: chronic kidney disease classification with multimedia data learning using deep stacked autoencoder network. *Multimed Tools Appl.* 2020;79(47):35425-35440.  
30. JongboOA, AdetunmbiAO, OgunrindeRB, Badeji-Ajisafe B. Development of an ensemble approach to chronic kidney disease diagnosis. *Scientific African.* 2020;8:e00456.  
31. Nusinovici S, Tham YC, Yan MY, et al. Logistic regression was as good as machine learning for predicting major chronic diseases. *J Clin Epidemiol.*  
32. Vijayalakshmi J, Ramaraj E. A Hadoop-big data analytic model to predict and classify chronic kidney diseases using improved fractional rough fuzzy K-means clustering and extreme gradient boost rat swarm optimizer. *Concurrency Computat Pract Exper.* 2022;34(28):e7354. doi: 10.1002/cpe.7354

of supervised-learning and soft computing techniques for stress diagnosis in humans. *Comput Biol Med.* 2021;134:104450.  
20. Zhang T, Ma F, Yue D, Peng C, O'Hare GM. Interval type-2 fuzzy local enhancement based rough k-means clustering considering imbalanced clusters. *IEEE Trans Fuzzy Syst.* 2019;28(9):1925-1939.  
21. Alsharman N, Alzyoud F, Saaidah A, Almomani O. Machine learning and answer set program rules towards traffic light management. *IJATCSE.* 2020;9(3):2535-2541.  
22. Dhiman G, Garg M, Nagar A, Kumar V, Dehghani M. A novel algorithm for global optimization: rat swarm optimizer. *J Ambient Intell Humaniz Comput.* 2021;12(8):8457-8482.  
23. Chaudhuri AK, Sinha D, Banerjee DK, Das A. A novel enhanced decision tree model for detecting chronic kidney disease. *Netw Model Anal Health Inform Bioinform.* 2021;10(1):1-22.  
24. JerlinRubini L, Perumal E. Efficient classification of chronic kidney disease by using multi-kernel support vector machine and fruit fly optimization algorithm. *Int J Imaging Syst Technol.* 2020;30(3):660-673.  
25. Nithya A, Appathurai A, Venkatadri N, RamjiDR, PalaganCA. Kidney disease detection and segmentation using artificial neural network and multi-kernel k-means clustering for ultrasound images. *Measurement.* 2020;149:106952.  
26. Pradeepa P, Jeyakumar MK. An efficient self-tuning spectral clustering algorithm for chronic kidney disease prediction. *Mater Today.* 2020.  
27. Akben SB. Early stage chronic kidney disease diagnosis by applying data mining methods to urinalysis, blood analysis and disease history. *IRBM.* 2018;39(5):353-358.  
28. Aldhyani TH, Alshebami AS, Alzahrani MY. Soft clustering for enhancing the diagnosis of chronic diseases over machine learning algorithms. *J Healthc Eng.* 2020;2020:1-16.