Iranian Journal of Diabetes and Metabolism *Original Article*

Official Journal of the Endocrinology and Metabolism Research Institute, Tehran University of Medical Sciences. 2025; Vol. 25; No. 5

A Method for Diabetes Diagnosis Using Simulated Annealing and K-Nearest Neighbor Algorithms

Hossein Azgomi*[1], Ali Asghari[2]

1. Department of Computer Engineering, Ra.C., Islamic Azad University, Rasht, Iran
2. Department of Computer Engineering, Shafagh Institute of Higher Education, Tonekabon, Iran

Abstract

Background: Diabetes is a chronic disease where the body cannot use or store glucose properly. Diabetes occurs when the pancreas is unable to produce insulin, or the body cannot use the insulin produced. Nowadays, diabetes is a common disease worldwide, and providing automated methods for its diagnosis is critically important.
Methods: This paper introduces a novel method for diagnosing diabetes using artificial intelligence (AI) algorithms. The proposed method is based on metaheuristic and classification algorithms. The simulated annealing (SA) metaheuristic algorithm was used for feature selection. Diabetes diagnosis was performed using the improved K-nearest neighbor (KNN) classification algorithm. In addition to the proposed method, the performance of two other methods, named MVMCNN and WKNN, was studied for diabetes diagnosis.
Results: The proposed method has been compared practically with the two other methods for diagnosing diabetes. The comparisons are based on the accuracy rate of disease diagnosis. In the experiments, the proposed method (SAKNN) demonstrated 95% accuracy, the MVMCNN method showed 93% accuracy, and the WKNN method demonstrated 90% accuracy. Thus, the proposed method outperformed the others. The proposed method also had acceptable performance in terms of time and several other criteria.
Conclusion: The proposed method for diagnosing diabetes, using metaheuristic and classification algorithms, provides higher accuracy compared to other methods. These results indicate that the proper use of AI techniques can offer effective solutions for the automatic diagnosis of diabetes and can be used as an auxiliary tool for doctors and researchers.

Keywords: Diabetes, Artificial Intelligence, Feature Selection, K-Nearest Neighbor

Please cite this article as:
Azgomi H, Asghari A. A Method for Diabetes Diagnosis Using Simulated Annealing and K-Nearest Neighbor Algorithms. *ijdld*. 2025; 25(5):453-471.
DOI: 10.18502/ijdl.v25i5.20342

[DOI: 10.18502/ijdl.v25i5.20342]

*Corresponding Author: Hossein Azgomi; **Email:** Hossein.Azgomi@iau.ac.ir
Taleshan Bridge, Lakan Road, Islamic Azad University, Rasht Branch, Rasht, Gilan, Iran. Tel: +981333424080

مجله دیابت و متابولیسم ایران
مقاله پژوهشی

نشریه پژوهشگاه علوم غدد و متابولیسم دانشگاه علوم پزشکی تهران
دورهٔ ۲۵، شمارهٔ ۵، سال ۱۴۰۴، صفحات ۴۵۳ تا ۴۷۱

روشی برای تشخیص بیماری دیابت با استفاده از الگوریتم‌های تبرید شبیه‌سازی شده و کی–نزدیک‌ترین همسایه

حسین ازگومی*[۱]، علی اصغری[۲]

۱- گروه مهندسی کامپیوتر، واحد رشت، دانشگاه آزاد اسلامی، رشت، ایران
۲- گروه مهندسی کامپیوتر، مؤسسه آموزش عالی شفق، تنکابن، ایران

چکیده

مقدمه: دیابت نوعی بیماری مزمن است که در آن بدن نمی‌تواند از گلوکز استفاده و یا آن را ذخیره کند. دیابت زمانی رخ می‌دهد که لوزالمعده قادر به ساخت انسولین نباشد یا بدن نتواند از انسولین تولید شده استفاده کند. امروزه بیماری دیابت یک بیماری شایع در جهان است و ارائه روش‌هایی خودکار برای تشخیص آن بسیار حائز اهمیت است.

روش‌ها: در این مقاله، روشی نوین برای تشخیص دیابت با استفاده از الگوریتم‌های هوش مصنوعی معرفی شده است. روش پیشنهادی مبتنی بر الگوریتم‌های فرا ابتکاری و طبقه‌بندی است. برای انتخاب ویژگی‌ها از الگوریتم فرا ابتکاری تبرید شبیه‌سازی شده (SA) استفاده شد. تشخیص دیابت نیز با استفاده از الگوریتم طبقه‌بندی بهبودیافته کی–نزدیک‌ترین همسایه (KNN) انجام می‌شود. علاوه بر روش پیشنهادی، عملکرد دو روش دیگر با نام‌های MVMCNN و WKNN در تشخیص دیابت مورد مطالعه قرار گرفتند.

یافته‌ها: روش پیشنهادی با دو روش دیگر برای تشخیص دیابت به‌صورت عملی مقایسه شده است. مقایسه‌ها براساس میزان دقت حاصل از تشخیص بیماری صورت گرفت. در آزمایش‌ها روش پیشنهادی (SAKNN) دقت ۹۵٪، روش MVMCNN دقت ۹۳٪ و روش WKNN دقت ۹۰٪ را ارائه کردند. بنابراین روش پیشنهادی نسبت به سایر روش‌ها عملکرد بهتری از خود نشان داده است. از نظر زمانی و چند معیار دیگر نیز روش پیشنهادی عملکرد قابل قبولی داشت.

نتیجه‌گیری: روش پیشنهادی برای تشخیص دیابت به کمک الگوریتم‌های فرا ابتکاری و طبقه‌بندی دقت بالاتری نسبت به دیگر روش‌ها ارائه می‌دهد. این نتایج نشان می‌دهد که استفاده مناسب از تکنیک‌های هوش مصنوعی می‌تواند راه‌حل‌های مؤثری برای تشخیص خودکار بیماریی دیابت ارائه دهد و می‌تواند به‌عنوان یک ابزار کمکی برای پزشکان و محققین به‌کار گرفته شود.

واژگان کلیدی: دیابت، هوش مصنوعی، انتخاب ویژگی، کی–نزدیک‌ترین همسایه

تاریخ دریافت: ۱۴۰۳/۱۰/۲۵ **تاریخ پذیرش:** ۱۴۰۴/۰۱/۲۳

به این مقاله، به صورت زیر استناد کنید:

Azgomi H, Asghari A. A Method for Diabetes Diagnosis Using Simulated Annealing and K-Nearest Neighbor Algorithms. *ijdld*. 2025; 25(5):453-471.

* ***نویسندهٔ مسئول:*** حسین ازگومی، **آدرس:** گیلان، رشت، جادهٔ لاکان، پل طالشان، دانشگاه آزاد اسلامی واحد رشت، تلفن: ۰۱۳۳۴۲۴۰۸۰، پست الکترونیک: Hossein.Azgomi@iau.ac.ir

[DOI: 10.18502/ijdl.v25i5.20342]

مقدمه

بیماری دیابت زمانی اتفاق می‌افتد که بدن انسان قادر به جذب قند یا همان گلوکز به سلول‌های خود و استفاده از آن برای تولید انرژی نباشد. این نقص باعث ایجاد قند اضافی در جریان خون می‌شود. بیماری قند خون کنترل نشده می‌تواند منجر به عواقب جدی شود و به طیف وسیعی از اندام‌ها و بافت‌های بدن، از جمله قلب، کلیه‌ها، چشم‌ها و اعصاب آسیب برساند [۱].

فرآیند هضم، شامل تجزیهٔ غذایی که مصرف می‌شود به منابع مختلف است. وقتی کربوهیدرات، مثل نان، برنج و ماکارونی مصرف می‌شود، بدن انسان آن را به گلوکز تجزیه می‌کند. هنگامی‌که گلوکز وارد جریان خون می‌شود، به کمک نیاز دارد تا به مقصد نهایی خود، یعنی داخل سلول‌های بدن برسد و در آنجا مورد استفاده قرار گیرد. در این حالت انسولین به گلوکز برای رسیدن به مقصد نهایی کمک می‌کند. انسولین هورمونی است که توسط پانکراس یا همان لوزالمعده، اندامی که در پشت معده قرار دارد، ساخته می‌شود. پانکراس، انسولین را در جریان خون آزاد می‌کند. انسولین حکم کلید را دارد که قفل در دیوارهٔ سلولی را باز می‌کند و به گلوکز اجازه می‌دهد تا وارد سلول‌های بدن شود. گلوکز، سوخت یا انرژی مورد نیاز بافت‌ها و اندام‌ها را برای عملکرد مناسب فراهم می‌کند [۱].

در افراد دیابتی دو حالت امکان دارد رخ دهد. یکی از حالات این است که پانکراس انسولین تولید نمی‌کند یا اگر تولید کند، مقدارش کافی نیست. حالت دیگر این است که پانکراس انسولین تولید می‌کند، اما سلول‌های بدن به آن پاسخ نمی‌دهند و نمی‌توانند آن‌طور که باید از آن استفاده کنند. اگر گلوکز نتواند وارد سلول‌های بدن شود، در جریان خون باقی می‌ماند و سطح گلوکز خون افزایش می‌یابد. امروزه بیماری قند خون در اکثر کشورهای جهان یک بیماری شایع و فراگیر است. یکی از علل اصلی این امر مصرف بی رویهٔ مواد حاوی گلوکز در اکثر کشورهای پیشرفته و در حال توسعه است. بنابراین ارائه روش‌هایی که بیماری قند خون را به‌صورت غیرتهاجمی و خودکار تشخیص دهد، بسیار ارزشمند است.

یادگیری ماشین روشی برای تجزیه‌وتحلیل داده است که ساخت مدل تحلیلی را خودکار می‌کند. این شاخه‌ای از هوش مصنوعی است که مبتنی بر این ایده است که سیستم‌ها می‌توانند از داده‌ها یاد بگیرند، الگوها را شناسایی کنند و با کمترین مداخلهٔ انسانی تصمیم بگیرند [۲]. یادگیری ماشین در زمینه‌های مختلفی مانند کلان داده [۳]، تشخیص تقلب [٤]، شبکه‌های کامپیوتری [۵] و غیره کاربرد دارد. تشخیص بیماری‌های مختلف از جمله دیابت توسط یادگیری ماشین یکی از مسائل مطرح و حائز اهمیت در تحقیقات است. یکی از شاخه‌های اصلی یادگیری ماشین جهت تشخیص، طبقه‌بندی است. طبقه‌بندی عملیات جداسازی موجودیت‌های مختلف به چندین کلاس است. این کلاس‌ها را میتوان با قوانین تجاری، مرزهای کلاس یا برخی تابع‌های ریاضی تعریف کرد. عملیات طبقه‌بندی ممکن است براساس یک رابطه بین یک تخصیص کلاس شناخته شده و ویژگی‌های موجودیت طبقه‌بندی شود. این نوع طبقه‌بندی تحت نظارت نامیده می‌شود [٦].

الگوریتم کی- نزدیک‌ترین همسایه یک الگوریتم یادگیری ماشین نظارت شده متداول است که می‌تواند برای حل مسائل طبقه‌بندی و رگرسیون استفاده شود. در این الگوریتم ابتدا فاصله داده ورودی با همهٔ داده‌های موجود محاسبه می‌شود. محاسبهٔ فاصله توسط یک معیار مانند فاصلهٔ اقلیدسی یا فاصلهٔ منهتن انجام می‌شود. سپس به تعداد k عدد نزدیک‌ترین داده به داده ورودی انتخاب می‌شوند. در پایان داده ورودی به مکررترین کلاس k داده انتخاب شده تعلق می‌یابد. در این الگوریتم مقدار k از ورودی دریافت می‌شود [۷].

الگوریتم‌های فرا ابتکاری دسته‌ای از الگوریتم‌ها هستند که سعی دارند تا با جستجوی تصادفی در فضای جستجوی مسئله، مسائل زمانبر را با یافتن جواب‌های نزدیک به بهینه، در زمان قابل قبولی حل نمایند. از جمله این الگوریتم‌های الگوریتم‌های میتوان به الگوریتم روابط اجتماعی درختان [۸] و الگوریتم بهینه‌سازی آب [۹] اشاره کرد. الگوریتم تبرید شبیه‌سازی شدهٔ یک الگوریتم فراابتکاری دیگر برای حل مسائل بهینه‌سازی بدون محدودیت است. این روش فرآیند فیزیکی گرم کردن یک ماده را مدل‌سازی می‌کند و سپس به آرامی دما را کاهش می‌دهد. در هر تکرار از الگوریتم تبرید شبیه‌سازی شده، یک راه‌حل جدید به‌طور تصادفی تولید می‌شود. فاصلهٔ راه‌حل جدید از راه‌حل فعلی یا وسعت جستجو براساس توزیع احتمال با مقیاسی متناسب با دما است. الگوریتم تمام راه‌حل‌های جدیدی را که بهتر باشند، می‌پذیرد. امّا با احتمال مشخص، راه‌حل‌هایی را که ضعیف‌تر باشند، می‌پذیرد. با پذیرش راه‌حل‌های ضعیف، الگوریتم از گیر افتادن در بهینه‌های محلی جلوگیری می‌کند و می‌تواند در سطح کلی برای راه‌حل‌های ممکن کاوش کند [۱۰]. الگوریتم تبرید شبیه‌سازی شده یک برنامهٔ سیستماتیک برای کاهش دما با ادامه در هر تکرار از الگوریتم دارد. با کاهش دما، الگوریتم میزان جستجوی خود را کاهش می‌دهد تا به‌صورت متمرکز جستجو

[DOI: 10.18502/ijdl.v25i5.20342]

نماید. الگوریتم تبرید شبیه‌سازی‌شده دارای ویژگی‌های مثبت متعددی است که آن را به یک روش مؤثر در حل مسائل بهینه‌سازی پیچیده تبدیل کرده است. این الگوریتم توانایی فرار از بهینه‌های محلی را دارد، زیرا در مراحل اولیه با احتمال بیشتری به بررسی نقاط کمتر مطلوب می‌پردازد و در مراحل بعدی با کاهش تدریجی دما، به تدریج به سمت نقاط با ارزش بالاتر حرکت می‌کند. این ویژگی به این الگوریتم امکان می‌دهد تا فضای جستجو را به‌طور گسترده‌تری بررسی کند و احتمال یافتن بهینه سراسری را افزایش دهد.

در این پایان‌نامه هدف ارائه یک روش جدید برای تشخیص بیماری دیابت براساس یادگیری ماشین است. روش پیشنهادی بر پایه الگوریتم‌های فرا ابتکاری و طبقه‌بندی است. در این روش از الگوریتم فرا ابتکاری تبرید شبیه‌سازی شده و الگوریتم طبقه‌بندی کی-نزدیک‌ترین استفاده می‌شود. در روش پیشنهادی انتخاب ویژگی به کمک الگوریتم تبرید شبیه‌سازی شده صورت می‌گیرد. تشخیص بیماری نیز به کمک الگوریتم کی-نزدیک‌ترین همسایه انجام می‌شود. هدف اصلی پژوهش افزایش دقت در تشخیص بیماری دیابت است. دستاورد علمی این مقاله عبارت است از:

- ارائه روشی برای تشخیص بیماری دیابت بر پایهٔ یادگیری ماشین
- انتخاب ویژگی به‌صورت بهینه با استفاده از الگوریتم تبرید شبیه‌سازی شده
- استفاده از الگوریتم کی-نزدیک‌ترین همسایه جهت تشخیص بیماری

ادامهٔ مقاله به این صورت سازمان‌دهی شده است که در بخش بعد راه‌کارهای مرتبط با مسئله تشخیص بیماری دیابت به کمک یادگیری ماشین مرور می‌شود؛ به‌دنبال آن روش پیشنهاد شده برای تشخیص بیماری دیابت ارائه می‌گردد. و در نهایت روش پیشنهادی به‌صورت عملی مورد ارزیابی قرار گرفته و با سایر روش‌های مشابه مقایسه می‌شود. نتیجه‌گیری حاصل از این پژوهش نیز در انتها ارائه می‌شود.

روش‌ها

راه‌کارهای مرتبط

در پژوهش‌های زیادی مسئله تشخیص بیماری دیابت با استفاده از یادگیری ماشین مورد بررسی قرار گرفته است. این پژوهش‌ها از الگوریتم‌های طبقه‌بندی برای تشخیص بیماری استفاده می‌کنند. با توجه به این که روش‌های ارائه شده براساس یادگیری ماشین و طبقه‌بندی هستند، از دو فاز آموزش و آزمایش تشکیل شده‌اند. در ادامه به بررسی برخی از مهم‌ترین پژوهش‌ها در این زمینه می‌پردازیم.

در مطالعه‌ای روشی برای تشخیص بیماری دیابت براساس الگوریتم ماشین بردار پشتیبان ارائه شده است [۱۱]. روش پیشنهادی از سه مرحله استخراج ویژگی، کاهش ویژگی و تشخیص تشکیل شده است. دقت روش پیشنهاد شده در این پژوهش حدود ۸۹ درصد است. در پژوهشی دیگر روشی بر پایهٔ الگوریتم کی- نزدیکترین همسایه برای تشخیص بیماری دیابت پیشنهاد شده است [۱۲]. در این روش دو مقدار ۳ و ۵ برای پارامتر k در الگوریتم کی- نزدیک ترین همسایه در نظر گرفته شده است. در این پژوهش دقت و خطای روش پیشنهادی مورد تحلیل قرار گرفت. در مطالعه‌ای دیگر الگوریتم‌های طبقه‌بندی در داده‌کاوی برای تشخیص بیماری دیابت مورد بررسی قرار گرفتند [۱۳]. الگوریتم‌های بررسی شده در این پژوهش الگوریتم درخت تصمیم، الگوریتم جنگل تصادفی، الگوریتم رگرسیون لجستیک، الگوریتم بیزین ساده و شبکه‌های عصبی مصنوعی هستند که الگوریتم جنگل تصادفی بهترین عملکرد را داشت. در بررسی دیگری از الگوریتم درخت تصمیم برای تشخیص بیماری دیابت استفاده شد [۱٤]. البته در این تحقیق مجموعه‌ای از درخت‌های تصمیم به‌صورت کیسه‌گذاری مورد استفاده قرار گرفتند. الگوریتم تصمیم مورد استفاده شده از نوع J48 (c4.5) است که عملکرد مناسبی دارد. در تحقیقی دیگر نیز الگوریتم طبقه‌بندی برای تشخیص بیماری دیابت مورد تحلیل قرار گرفتند [۱۵]. الگوریتم‌های بررسی شده در این پژوهش الگوریتم ماشین بردار پشتیبان، الگوریتم درخت تصمیم و الگوریتم بیزین ساده است. در این تحقیق الگوریتم بیزین ساده عملکرد بهتری نسبت به دو روش دیگر داشت. در پژوهشی دیگر نیز روشی برای تشخیص بیماری دیابت با استفاده از الگوریتم جنگل تصادفی ارائه شده است [۱٦]. البته از یک نسخهٔ بهبود یافته الگوریتم جنگل تصادفی در این تحقیق استفاده شده است. برای بهبود الگوریتم جنگل تصادفی، از الگوریتم ژنتیک استفاده شده است. روش پیشنهادی عملکرد قابل قبولی از نظر دقت دارد.

در تحقیقات چند سال اخیر نیز روی مسئله تشخیص بیماری دیابت به کمک روش‌های یادگیری تأکید شده است. برای نمونه در یک مطالعه بیماری دیابت به کمک الگوریتم کی- نزدیک‌ترین همسایه تشخیص داده شده است [۱۷]. در این تحقیق از انواع مختلف الگوریتم کی- نزدیک ترین همسایه استفاده شده است. در

[DOI: 10.18502/ijdl.v25i5.20342]

این تحقیق نوع‌های بهبودیافته الگوریتم کی- نزدیک‌ترین همسایه برای تشخیص بیماری دیابت معرفی گردید. در یک بررسی دیگرروشی برای تشخیص بیماری دیابت بر پایهٔ الگوریتم بیزین ارائه شده است [۱۸]. در این پژوهش از الگوریتم بیزین ساده استفاده شد. نتایج این تحقیق نشان می‌دهد که با استفاده از روش‌های یادگیری ماشین می‌توان با دقت بالاتری بیماری دیابت را نسبت به روش‌های سنتی تشخیص داد. در مطالعه‌ای نیز مدلی برای تشخیص بیماری دیابت بر پایهٔ یادگیری ماشین ارائه شده است [۱۹]. مدل ارائه شده با الگوریتم رگرسیون لجستیک، الگوریتم بیزین ساده و الگوریتم کی- نزدیک‌ترین همسایه مورد تجزیه‌وتحلیل قرار گرفت. نتایج نشان داد که الگوریتم رگرسیون لجستیک عملکرد مناسب‌تری دارد. در بررسی دیگری [۲۰] یک چارچوب جدید برای تشخیص بیماری دیابت براساس الگوریتم کی- نزدیک‌ترین همسایه پیشنهاد شده است [۲۰]. در این پژوهش از انواع مختلف الگوریتم کی- نزدیک‌ترین همسایه استفاده شده است. نتایج عملی در این تحقیق نشان می‌دهد که چارچوب پیشنهادی بسیار کارآمد است. محققان در بررسی دیگری روشی مبتنی بر الگوریتم کی- نزدیک‌ترین همسایه، شبکه‌های عصبی و منطق فازی برای تشخیص بیماری دیابت ارائه کردند [۲۱]. این روش adaptable neuro-fuzzy inference K Nearest Neighbourhood (AF-KNN) نامیده شد که بر پایه ویژگی‌های رفتاری بیماران است. نتایج عملی نشان داده که روش پیشنهاد شده دقت بالاتری نسبت به روش‌های مشابه ارائه می‌کند. مسئلهٔ پیش‌بینی شروع دیابت به کمک تکنیک‌های یادگیری ماشین در یک مطالعه مورد بررسی قرار گرفت [۲۲]. در این پژوهش انواع الگوریتم‌های یادگیری بر روی یک مجموعه داده جمع‌آوری شده با ۸ ویژگی آزمایش شدند. نتایج نشان داد که بهترین الگوریتم، two-class boosted decision tree است که بهترین دقت را دارد. در پژوهشی دیگر روش برای تشخیص دیابت به کمک رنگ عنبیهٔ چشم ارائه شده است [۲۳]. روش پیشنهاد شده در این تحقیق بر پایهٔ الگوریتم ماشین بردار پشتیبان است و هدف آن افزایش دقت تشخیص است. در مطالعه‌ای دیگر نیز روش دیگری برای تشخیص دیابت براساس رنگ عنبیه ارائه شد که براساس الگوریتم کی- نزدیک‌ترین همسایه است [۲۴]. علاوه بر دیابت در این تحقیق بیماری کووید-۱۹ نیز مد نظر قرار گرفت و مشخص شد که رابطهٔ مشترکی برای تشخیص این دو بیماری براساس روش پیشنهاد شده وجود دارد. در یک پژوهش جدید چارچوبی برای پیش‌بینی دیابت از طریق رویکرد یادگیری ماشین ترکیبی توسعه یافته معرفی شده است [۲۵]. این چارچوب نوع مدل ماشین بردار پشتیبان و شبکهٔ عصبی مصنوعی که دو تکنیک محبوب در یادگیری ماشین هستند. را ترکیب می‌کند. این مدل‌ها داده‌ها را برای تعیین مثبت یا منفی بودن تشخیص دیابت تجزیه‌وتحلیل می‌کنند. در آزمایش‌های علمی مشخص شد که روش پیشنهادی با دقت بالاتری به تشخیص می‌پردازد. در یک مطالعهٔ جدید دیگری نیز روشی برای تشخیص دیابت با استفاده از یادگیری عمیق مبتنی بر توجه ارائه شد [۲۶]. این روش بر پایهٔ کاوش شبکه‌های عصبی مکرر با سازکارهای مبتنی بر توجه است. روش پیشنهادی از نظر دقت و زمان یادگیری و تشخیص عملکرد مناسبی در مقایسه با روش‌های مشابه دارد. جدول ۱، راه‌کارهای مرور شدهٔ تشخیص بیماری دیابت براساس الگوریتم‌های یادگیری ماشین لیست شده است.

همان طور که قابل مشاهده است در سال‌های مختلف و به‌خصوص در سال‌های اخیر به تحقیق در مسئله تشخیص دیابت به کمک الگوریتم‌های یادگیری ماشین توجه شده است. الگوریتم کی-نزدیک‌ترین همسایه در بسیاری از تحقیق‌ها برای تشخیص این بیماری مورد استفاده قرار گرفته است. علاوه بر این، یکی از مهم‌ترین عامل‌ها در تشخیص بیماری‌ها به کمک یادگیری ماشین، انتخاب ویژگی‌های مناسب می‌شود. انتخاب ویژگی با استفاده از الگوریتم‌های فرا ابتکاری می‌تواند به خوبی انجام شود. بنابراین ارائه روشی برای تشخیص دیابت براساس انتخاب ویژگی و الگوریتم کی-نزدیک‌ترین همسایه می‌تواند حائز اهمیت باشد.

روش پیشنهادی

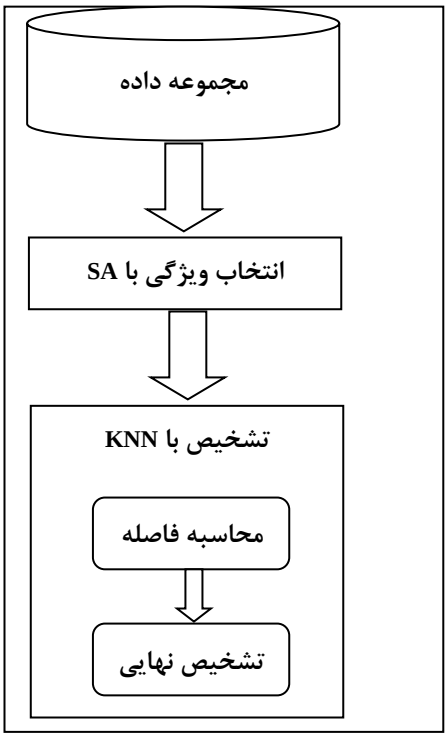
روش پیشنهادی برای تشخیص بیماری دیابت بر پایه الگوریتم تبرید شبیه‌سازی شده و الگوریتم کی- نزدیک‌ترین همسایه است. بنابراین روش پیشنهادی از دو گام کلی تشکیل شده است. گام اول مربوط به انتخاب ویژگی است. در روش پیشنهادی از الگوریتم تبرید شبیه‌سازی شده برای انتخاب ویژگی استفاد می‌شود. در این گام ویژگی‌های مناسب برای تشخیص بیماری قند خون از بین همهٔ ویژگی‌ها انتخاب می‌شود. این کار باعث می‌شود که هم سرعت و هم دقت تشخیص افزایش یابد. زیرا فقط از ویژگی‌های مناسب استفاده می‌شود.

[DOI: 10.18502/ijdl.v25i5.20342]

جدول ۱- راه‌کارهای مرورشدهٔ تشخیص دیابت براساس یادگیری ماشین

پژوهش	سال	الگوریتم‌های استفاده شده
(Calisir &el al)	۲۰۱۱	ماشین بردار پشتیبان
(Saxena &el al)	۲۰۱٤	کی- نزدیک‌ترین همسایه
(Nai-Arun &el al)	۲۰۱۵	درخت تصمیم، جنگل تصادفی، رگرسیون و بیزین
(Perveen &el al)	۲۰۱٦	درخت تصمیم
(Sisodia &el al)	۲۰۱۸	بیزین، درخت تصمیم و ماشین بردار پشتیبان
(Komal Kumar &el al)	۲۰۱۹	جنگل تصادفی
(Ali &el al)	۲۰۲۰	کی- نزدیک‌ترین همسایه
(Priya &el al)	۲۰۲۰	الگوریتم بیزین
(Khaleel &el al)	۲۰۲۱	رگرسیون، بیزین و کی- نزدیک‌ترین همسایه
(Suyanto &el al)	۲۰۲۲	کی- نزدیک‌ترین همسایه
(Prasad &el al)	۲۰۲۳	کی- نزدیک‌ترین همسایه، شبکهٔ عصبی و منطق فازی
(Chou &el al)	۲۰۲۳	رگرسیون، شبکهٔ عصبی، جنگل تصادفی، درخت تصمیم
(Shabdiz-a &el al)	۲۰۲٤	ماشین بردار پشتیبان
(Shabdiz-b &el al)	۲۰۲٤	کی- نزدیک‌ترین همسایه
(Reddy &el al)	۲۰۲۵	ماشین بردار پشتیبان و شبکهٔ عصبی
(Tiwari &el al)	۲۰۲۵	یادگیری عمیق

در گام دوم روش پیشنهادی تشخیص بیماری صورت می‌گیرد. این کار به کمک الگوریتم کی- نزدیک‌ترین همسایه انجام می‌شود. در این الگوریتم فاصله تمامی داده‌ها با دادهٔ ورودی محاسبه می‌شود. سپس تعدادی از نزدیک‌ترین داده‌ها انتخاب می‌شود و تشخیص به کمک آنها صورت می‌گیرد. برای این منظور داده ورودی به پُرتکرارترین کلاس از بین داده‌های انتخاب شده، اختصاص می‌یابد. در شکل ۱ نمای کلی روش پیشنهادی برای تشخیص بیماری دیابت نمایش داده شده است.



شکل ۱- نمای کلی روش پیشنهادی برای تشخیص دیابت

[DOI: 10.18502/ijdl.v25i5.20342]

همان‌طور که مشاهده می‌شود ابتدا یک مجموعه داده از بیماران موجود است که دارای برچسب است. سپس در گام اول ویژگی‌های مناسب از مجموعه داده به کمک الگوریتم تبرید شبیه‌سازی شده انتخاب می‌شود. در گام دوم نیز به کمک الگوریتم کی- نزدیک‌ترین همسایه تشخیص انجام می‌شود. طبق روال این الگوریتم ابتدا نیاز است که فاصلهٔ بین داده ورودی و داده‌های موجود محاسبه شود و سپس عمل تشخیص بیماری صورت بگیرد. در ادامه مراحل روش پیشنهادی مورد بررسی قرار می‌گیرد.

انتخاب ویژگی با الگوریتم تبرید شبیه‌سازی شده[1] (SA)

همان‌طور که اشاره شد، در گام اول روش پیشنهاد شده، انتخاب ویژگی انجام می‌شود. انتخاب ویژگی در کاربردهای یادگیری ماشین حائز اهمیت است. به‌خصوص زمانی که تعداد ویژگی‌ها زیاد باشند. با انتخاب ویژگی مناسب دقت تشخیص بالا می‌رود. زیرا فقط ویژگی‌های مناسب در عمل تشخیص مورد استفاده قرار می‌گیرند. همچنین این کار باعث کاهش پردازش و افزایش سرعت می‌شود. زیرا تعداد ویژگی‌های کمتری مورد پردازش قرار می‌گیرند.

مسئلهٔ انتخاب ویژگی یک مسئله سخت است. زیرا تعداد حالات ممکن برای انتخاب ویژگی بسیار زیاد و به‌صورت نمایی است. بنابراین روش‌های متداول برای حل این مسئله بسیار زمان‌بر هستند و نیاز است از الگوریتم‌های تقریبی برای این منظور استفاده شود. در روش پیشنهادی از الگوریتم فراابتکاری تبرید شبیه‌سازی شده برای انتخاب ویژگی استفاده می‌شود.

الگوریتم تبرید شبیه‌سازی شده با یک جواب تصادفی کار خود را آغاز می‌کند. سپس در یک حلقه تکرار، تعدادی جواب همسایه جواب فعلی تولید می‌شود. آنگاه در صورت تولید جواب بهتر در همسایه‌ها، آن جواب جایگزین جواب فعلی می‌شود. درصورتی‌که جواب بهتر نیز تولید نشود، باز هم شانس برای قرارگیری بهترین جواب همسایه به‌جای جواب فعلی وجود دارد. در هر مرحله از الگوریتم این روال انجام می‌شود تا شرط توقف حلقه برقرار شود و در پایان جواب فعلی به‌عنوان جواب نهایی مسئله در نظر گرفته می‌شود. در ادامه جزئیات الگوریتم تبرید شبیه‌سازی شده برای انتخاب ویژگی در روش پیشنهادی مطرح می‌گردد.

تولید راه‌حل اولیه

در الگوریتم تبرید شبیه‌سازی شده باید ابتدا نحوهٔ نمایش جواب‌ها مشخص گردد. در روش پیشنهادی از روش باینری برای نمایش جواب‌ها استفاده می‌شود. در این روش برای هر ویژگی یک بیت در نظر گرفته می‌شود. آنگاه یک بیت مقدار یک می‌گیرد، اگر ویژگی مربوط به آن، در جواب انتخاب شده باشد. اگر یک ویژگی در جواب انتخاب نشده باشد، بیت مربوط به آن مقدار صفر می‌گیرد. بنابراین طول جواب‌ها برابر با تعداد ویژگی‌ها است. به‌عنوان نمونه فرض کنید شش ویژگی از A1 تا A6 موجود باشد. در این حالت طول جواب‌ها برابر با شش است و به هر ویژگی یک بیت اختصاص داده می‌شود. یک جواب فرضی در روش پیشنهادی به‌صورت شکل ۲ است.

1	1	0	1	0	0
A1	A2	A3	A4	A5	A6

شکل ۲- نمایش راه‌حل‌ها در روش پیشنهادی

همان‌طور که مشاهده می‌شود طول جواب برابر با شش است و برای هر ویژگی یک بیت اختصاص داده شده است. در این جواب فرضی، ویژگی‌های اول، دوم و چهار انتخاب شده‌اند، زیرا بیت مربوط به آنها مقدار یک دارد. سایر ویژگی‌ها انتخاب نشده‌اند.

پس از مشخص نمودن نحوهٔ نمایش جواب‌ها، نحوهٔ تولید جواب اولیه باید مشخص گرد. الگوریتم تبرید شبیه‌سازی شده بر خلاف اکثر الگوریتم‌های فراابتکاری فقط با یک جواب به جستجو می‌پردازد. البته نسخه‌هایی از این الگوریتم نیز ارائه شده است که با یک جمعیت از جواب‌ها کار می‌کند. امّا در روش پیشنهادی برای کاهش ویژگی از نسخهٔ اصلی الگوریتم تبرید شبیه‌سازی شده استفاده می‌شود. این جواب در روش پیشنهادی به‌صورت کاملاً تصادفی تولید می‌گردد. یعنی در ابتدای کار یک رشته بیت شامل صفرها و یک‌ها به‌طول تعداد ویژگی‌ها به‌صورت تصادفی تولید شده و به‌عنوان جواب اولیه الگوریتم در نظر گرفته می‌شود. همچنین این الگوریتم یک پارامتر به نام دما دارد که تولید جواب‌ها به کمک آن کنترل می‌شود. برای این پارامتر باید در ابتدای کار یک مقدار در نظر گرفت.

تابع برازندگی

در الگوریتم تبرید شبیه‌سازی شده نیاز است که همواره در مراحل مختلف میزان مناسب بودن جواب مشخص شود. این

[1] Simulated Annealing

[DOI: 10.18502/ijdl.v25i5.20342]

کار به کمک تابع برازندگی انجام می‌شود. تابع برازندگی یک تابع است که جواب را دریافت می‌کند و خروجی آن یک عدد است که میزان مناسب بودن آن جواب را نشان می‌دهد.

در روش پیشنهادی برای تابع برازندگی، ویژگی‌های انتخاب شده آن جواب به کمک طبقه‌بندی کی- نزدیک‌ترین همسایه آموزش داده می‌شود و میزان خطای حاصل شده محاسبه می‌شود. خطای در نظر گرفته شده در روش پیشنهادی، خطای میانگین مربعات است. قابل ذکر است که برای تابع برازندگی تمامی داده‌ها در نظر گرفته نمی‌شوند و فقط زیرمجموعه‌ای از آنها مورد استفاده قرار می‌گیرد تا محاسبات تابع برازندگی زمانبر نباشد. تابع برازندگی در روش پیشنهادی به‌صورت رابطهٔ ۱ است.

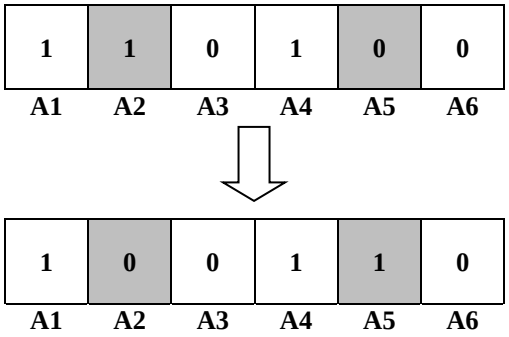
$$Fitness = \frac{1}{n} \times \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 \qquad (۱)$$

در این رابطه n برابر با تعداد داده‌ها، Y_i برابر با observed values و $\hat{Y}_i$ برابر با predicted values است. بنابراین با توجه به موارد مطرح شده هرچه قدر میزان خطا یا همان برازندگی برای جوابی کمتر باشد، آن جواب بهتر است. مقدار بالای خطا نیز نشان دهندهٔ ضعیف بودن جواب است.

تولید جواب‌های همسایه

پس از تولید جواب اولیه، الگوریتم وارد یک حلقهٔ تکرار می‌شود و در هر تکرار از حلقه تعدادی جواب همسایه تولید می‌شود. تعداد جواب‌های همسایه تولید شده در هر مرحله با پارامتر NN مشخص می‌شود که از ورودی دریافت می‌گردد.

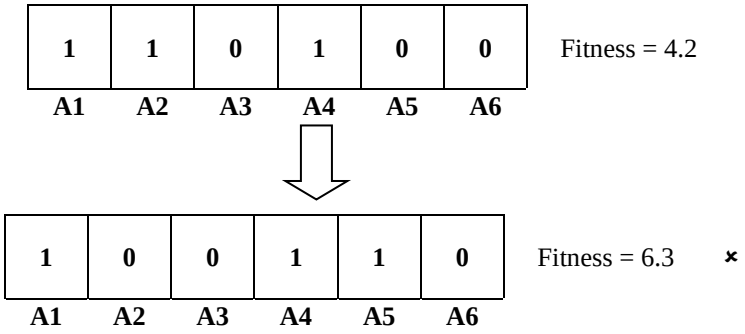
برای تولید جواب همسایه در روش پیشنهادی از روش تعویضی استفاده می‌شود. در این روش دو بیت به‌صورت تصادفی انتخاب می‌شوند و مقادیر آنها با یکدیگر تغییر می‌کند و به این صورت یک جواب جدید تولید می‌شود. نمونه‌ای از تولید جواب به روش تعویضی در شکل ۳ نمایش داده شده است.



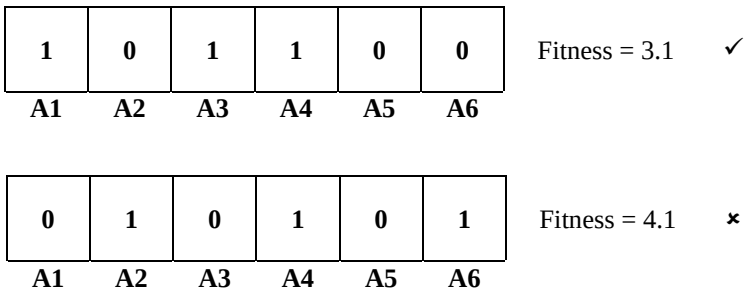
شکل ۳- تولید جواب همسایه به روش تعویضی

همان‌طور که مشاهده می‌شود در این نمونه بیت‌های دوم و پنجم به‌صورت تصادفی انتخاب شده‌اند و مقادیر آنها با یکدیگر جابه‌جا شده است و جواب جدیدی به‌وجود آمده است.

پس از تولید جواب‌های همسایه دو حالت ممکن است رخ دهد. یکی از حالات این است که بهترین جواب همسایه از جواب فعلی بهتر باشد. در این صورت بهترین جواب همسایه جایگزین جواب فعلی است. نمونه‌ای از این حالت با فرض NN=3 در شکل ٤ نشان داده شده است.



[DOI: 10.18502/ijdl.v25i5.20342]



شکل ۴- قرارگیری بهترین جواب همسایه

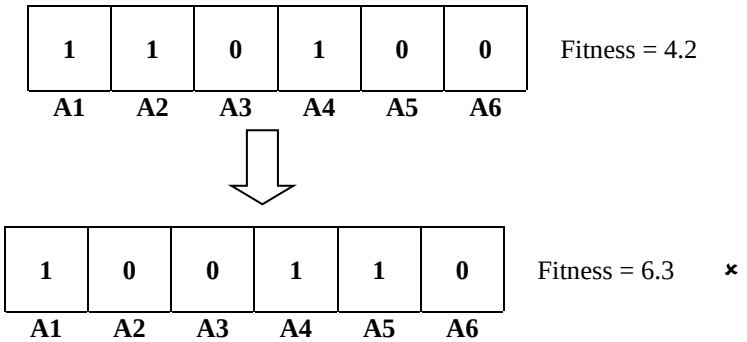
همان‌طور که قابل مشاهده است در این نمونه، سه جواب همسایه تولید شده است که یکی ضعیف‌تر از جواب فعلی است و دو جواب دیگر بهتر هستند. جواب همسایه دوم در این مثال بهترین جواب است که جایگزین جواب فعلی می‌شود.

حالت دیگری که ممکن است رخ دهد این است که هیچ‌کدام از جواب‌های همسایه بهتر از جواب فعلی نباشند. در این حالت جواب‌های همسایه کاملاً کنار گذاشته نمی‌شوند و احتمالی برای حضور آنها در نظر گرفته می‌شود. در این حالت بهترین جواب همسایه با احتمال بولتزمن جایگزین جواب فعلی می‌شود. تابع احتمال بولتزمن در رابطهٔ ۲ نشان داده شده است.

$$Pre = e^{(-\Delta E/T)} \qquad (۲)$$

در این رابطه T مقدار دما در مرحلهٔ جاری و ΔE مقدار نرمال شده تفاوت برازندگی بین جواب فعلی و جواب همسایه است. پس از محاسبه این احتمال یک عدد تصادفی بین صفر و یک تولید می‌شود. اگر این عدد کوچک‌تر یا مساوی احتمال بولتزمن باشد، آنگاه جواب همسایه جایگزین جواب فعلی می‌شود. در غیر این صورت جواب فعلی تغییری نمی‌کند. نمونه‌ای از این حالت با فرض NN=3 در شکل ۵ نشان داده شده است.

در مثال سه جواب تولید شده که هر سه ضعیف‌تر از جواب فعلی هستند. جواب سوم همسایه بهترین جواب است و برای آن احتمال بولتزمن محاسبه می‌شود. در اینجا عدد تصادفی کوچک‌تر از احتمال بولتزمن است و جواب سوم جایگزین جواب فعلی می‌شود.



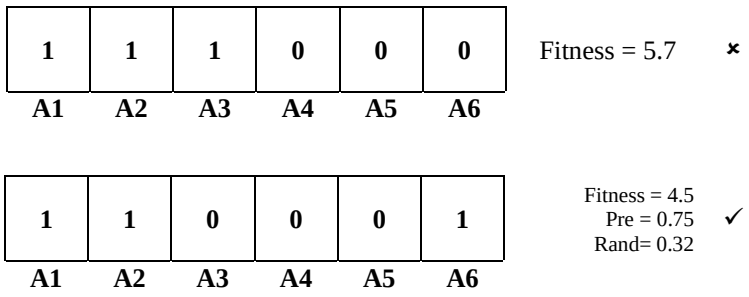
کاهش دما

در هر تکرار از الگوریتم، پس از تولید جواب‌های همسایه و جایگزینی آنها در صورت برقراری شرایط، مقدار دما باید کاهش یابد. معیارهای مختلفی برای کاهش دما در الگوریتم تبرید شبیه‌سازی شده وجود دارد که در روش پیشنهادی از قاعده خطی استفاده می‌شود. رابطهٔ کاهش دما در روش پیشنهادی در رابطهٔ ۳ نشان داده شده است.

$$T = T - \left((T_0 - T_f)/Maxiter\right) \qquad (۳)$$

در این رابطه T دمای جاری، T0 دمای اولیه، Tf دمای نهایی و Maxitrt تعداد تکرار الگوریتم است. توجه شود که در ابتدای الگوریتم دما بالا است و الگوریتم به‌صورت متنوع فضای مسئله را جستجو می‌کند. در ادامه دما کاهش می‌یابد و براساس احتمال بولتزمان فضای مسئله به‌صورت متمرکز جستجو می‌شود. بنابراین تنوع و نمرکز به خوبی اعمال می‌شود.

[DOI: 10.18502/ijdl.v25i5.20342]



شکل ۵- قرارگیری جواب همسایه ضعیف

شرط توقف

در روش پیشنهادی در هر تکرار از الگوریتم جواب‌های همسایه تولید می‌شوند و دما نیز کاهش پیدا می‌کند تا یک شرط توقف برقرار شود. شرط‌های توقف مختلفی را می‌توان برای الگوریتم‌های فرا ابتکاری در نظر گرفت. در روش پیشنهادی تعداد تکرار ثابت حلقه تکرار به‌عنوان شرط توقف در نظر گرفته شده است. این تعداد تکرار با پارامتر Maxiter مشخص می‌شود که از ورودی دریافت می‌شود.

تشخیص بیماری با کی-نزدیک‌ترین همسایه[1] (KNN)

در روش پیشنهادی پس از انتخاب ویژگی در مرحلهٔ اول به کمک الگوریتم تبرید شبیه‌سازی شده، در مرحلهٔ دوم تشخیص بیماری صورت می‌گیرد. این کار به کمک یادگیری ماشین و الگوریتم کی-نزدیک‌ترین همسایه انجام می‌شود.

در الگوریتم کی- نزدیک‌ترین همسایه ابتدا فاصلهٔ همهٔ داده‌ها با داده ورودی محاسبه می‌شود. سپس تعدادی دادهٔ نزدیک به دادهٔ ورودی انتخاب می‌شود و براساس آنها تشخیص صورت می‌گیرد. در ادامه روال تشخیص به کمک الگوریتم کی- نزدیک‌ترین همسایه در روش پیشنهادی شرح داده می‌شود.

محاسبهٔ فاصله

همان‌طور که اشاره شد، در الگوریتم کی- نزدیک‌ترین همسایه نیاز است که فاصلهٔ بین دادهٔ ورودی با داده‌های موجود محاسبه گردد. معیارهای مختلفی برای محاسبهٔ فاصلهٔ بین داده‌ها وجود دارد. در روش پیشنهادی از فاصلهٔ اقلیدسی برای محاسبهٔ فاصله استفاده می‌شود. اگر x و y دو نقطه با p مؤلفه باشند فاصلهٔ اقلیدسی طبق رابطهٔ ٤ محاسبه می‌گردد.

$$D = \sqrt{\sum_{i=1}^{p}(x_i - y_i)^2} \qquad (٤)$$

در ابتدای الگوریتم، فاصلهٔ اقلیدسی بین دادهٔ وروی با همهٔ داده‌های موجود محاسبه می‌شود. منظور از دادهٔ ورودی داده‌ای است که باید وضعیت بیماری آن تشخیص داده شود. این داده از ورودی دریافت می‌شود و ویژگی‌های آن دقیقاً مانند ویژگی‌های داده‌های موجود است.

فاصلهٔ اقلیدسی به‌دلیل سادگی رابطهٔ آن، به راحتی قابل محاسبه است و در بسیاری از الگوریتم‌های مختلف کاربرد دارد. این فاصله به‌دلیل پایه‌گذاری آن بر روی هندسهٔ اقلیدسی، به‌خوبی قابل تجسم و نمایش بصری است. در بسیاری از الگوریتم‌های یادگیری ماشین و تحلیل داده‌ها استفاده می‌شود. فاصلهٔ اقلیدسی می‌تواند در فضای چندبعدی (حتی بیش از سه بعد) نیز محاسبه شود، که این ویژگی آن را برای تحلیل داده‌های پیچیده مناسب می‌کند. همچنین این فاصله پایدار است و در برابر تغییرات مقیاس داده‌ها مقاوم است و می‌توان از آن در تحلیل داده‌ها بدون نگرانی از تغییرات مقیاس استفاده کرد. به همین علت از فاصلهٔ اقلیدسی در روش پیشنهادی استفاده می‌شود.

تشخیص بیماری

پس از محاسبهٔ فاصلهٔ اقلیدسی تمامی داده‌ها با دادهٔ ورودی، برای تشخیص بیماری داده‌ها براساس فاصله و به‌صورت نزولی مرتب می‌شوند. سپس تعدادی نزدیک‌ترین داده به دادهٔ ورودی مشخص می‌شود. این تعداد با پارامتر k مشخص می‌شود که از ورودی دریافت می‌شود.

پس از انتخاب k دادهٔ نزدیک به دادهٔ ورودی، کلاس‌هایی که آنها به آن تعلق دارند مورد بررسی قرار می‌گیرند. دادهٔ وروی به مکررترین کلاس در بین کلاس‌های داده‌های انتخاب شده تعلق می‌یابد و به این

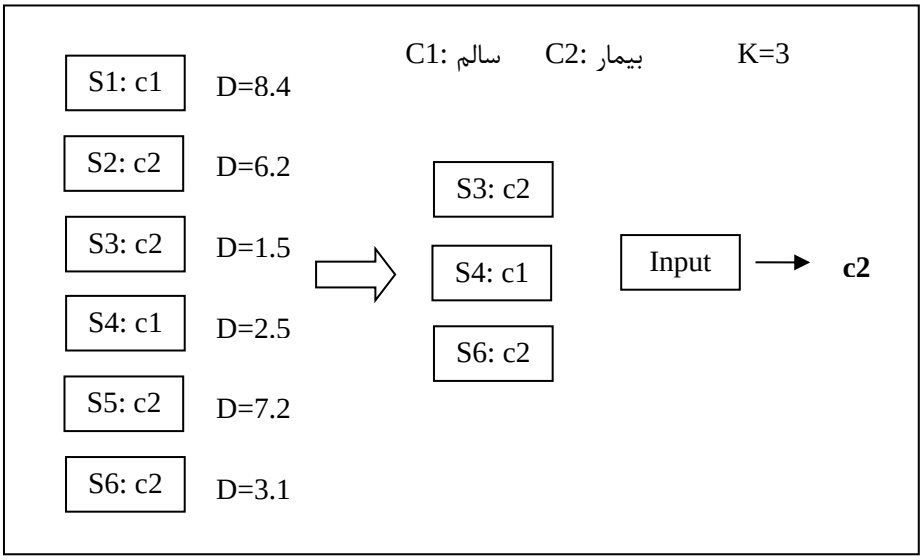
[1] K-Nearest Neighbors

[DOI: 10.18502/ijdl.v25i5.20342]

صورت تشخیص بیماری انجام می‌شود. نمونه‌ای از تشخیص در روش پیشنهادی در صورت شکل ٦ قابل مشاهده است.

همان‌طور که قابل مشاهده است، شش نمونه داده وجود دارد که کلاس هر کدام مشخص است. هر نمونه داده مربوط به یک فرد است. کلاس c1 به معنی فرد سالم و کلاس c2 به معنی فرد بیمار است. فاصلهٔ دادهٔ ورودی با هر یک از داده‌ها کنار آن درج شده است. پارامتر k نیز برابر با ۳ است.

در این مثال سه دادهٔ S3، S4 و S6 از همه نزدیک‌تر به دادهٔ ورودی هستند. این سه داده انتخاب شده و کلاس‌های آنها بررسی می‌شوند. چون دو عدد از این داده‌ها به کلاس c2 تعلق دارند و یک عدد به کلاس c1 تعلق دارد، دادهٔ ورودی به کلاس c2 تعلق پیدا می‌کند. یعنی نمونه ورودی بیمار است. به این صورت تشخیص بیماری دیابت در روش پیشنهادی انجام می‌شود.



شکل ۶- تشخیص بیماری در روش پیشنهادی

نتایج تجربی

در این بخش روش پیشنهاد شده برای تشخیص بیماری دیابت با دیگر روش‌های مشابه در این زمینه مقایسه می‌شود. در ادامه ابتدا محیط و شرایط ارزیابی‌ها معرفی می‌شود. سپس نتایج ارزیابی‌های عملی ذکر و تحلیل می‌گردد. سعی شده است که شرایط برای تمامی روش‌ها یکسان و عادلانه باشد.

● **محیط ارزیابی:** روش پیشنهادی برای تشخیص بیماری دیابت با استفاده از نرم‌افزار Matlab پیاده سازی شد. آزمایش‌ها در یک سیستم با پردازندهٔ Intel(R) Core(TM) i3-4030U 1.90GHz، حافظه‌ی رم 4GB، هارد دیسک 500GB و ویندوز ۱۰ صورت گرفت.

● **مجموعه داده:** برای آزمایش‌های عملی از مجموعه دادهٔ Diabetes 130-US hospitals[1] استفاده شد. این مجموعه داده مرتبط با بیماران دیابتی است و در آن اطلاعات حدود صد هزار بیمار جمع‌آوری شده است. در این مجموعه داده برای هر نمونه داده ۵۰ ویژگی مختلف ثبت شده است. برچسب‌های این مجموعه داده دو حالت Yes و No دارند که به معنی بیمار بودن و بیمار نبودن است. قابل ذکر است که در ارزیابی‌های انجام شده ۷۰ در صد داده‌ها برای داده‌های آموزشی و ۳۰ درصد باقی مانده برای داده‌های آزمایشی در نظر گرفته شدند. در جدول ۲ یک نمونه از ویژگی‌های انتخاب شده توسط روش پیشنهادی از بین ۵۰ ویژگی این مجموعه داده ذکر شده است.

جدول ۲- ارزیابی دقت روش‌ها در اجراهای مختلف

ویژگی	شرح
race	نژاد بیمار
gender	جنسیت بیمار

[1] https://archive.ics.uci.edu

[DOI: 10.18502/ijdl.v25i5.20342]

age	سن بیمار
admission_type_id	نوع پذیرش
time_in_hospital	مدت زمان بستری در بیمارستان
num_lab_procedures	تعداد آزمایش‌های انجام شده
number_outpatient	تعداد بازدیدهای سرپایی
diag_1	تشخیص اول
diag_2	تشخیص دوم
number_emergency	تعداد بازدیدهای اورژانس
readmitted	رخ بازگشت بیمار
num_medications	تعداد داروها
payer_code	وضعیت اقتصادی-اجتماعی
number_inpatient	تعداد بستری‌ها
A1Cresult	نتیجه آزمایش HbA1c

همان‌طور که قابل مشاهده است در این نمونه تعداد ۱۵ ویژگی انتخاب شده است. این ویژگی‌های انتخاب شده تأثیر زیادی در تشخیص بیماری دیابت دارند.

روش‌ها: در آزمایش‌های عملی از سه روش استفاده شد:

- **روش مبتنی بر وزن[1] (WKNN)** : این روش بیماری دیابت را براساس الگوریتم کی- نزدیک‌ترین همسایه تشخص می‌دهد. در این روش از یک نوع خاص الگوریتم کی- نزدیکترین همسایه که مبتنی بر وزن است استفاده شده است. جزئیات این روش در مطالعهٔ Ali و همکاران ذکر شده است [۱۷].
- **روش چند رای دهنده کمیسیونی[2] (MVMCNN)**: این روش نیز بر پایه الگوریتم کی-نزدیک‌ترین همسایه است. در این روش از یک نوع خاص الگوریتم کی- نزدیک‌ترین همسایه به نام نزدیک‌ترین همسایه چند کمیسیون چند رای دهنده استفاده می‌شود. جزئیات این روش در مطالعهٔ Khaleel و همکاران ذکر شده است [۲۰].
- **روش پیشنهادی[3] (SAKNN)**: این روش، همان روش پیشنهاد شده در این مقاله است. در این روش انتخاب ویژگی به کمک الگوریتم تبرید شبیه‌سازی شده انجام می‌شود. تشخیص بیماری قند خون نیز به کمک الگوریتم کی- نزدیک‌ترین همسایه صورت می‌گیرد. جزئیات این روش در بخش ۳ ارائه شده است. در جدول ۳ یک نمونهٔ کلی از مقادیر پارامترهای روش پیشنهادی در ارزیابی‌های عملی ذکر شده است. قابل ذکر است که در آزمایش‌های مختلف مقادیر این پارامترها بسته به شرایط تغییر می‌کند که در ادامه شرح داده شده است.

جدول ۳- مقادیر پارامترهای روش پیشنهادی

پارامتر	شرح	مقدار
N	تعداد اعضای جمعیت	1
NN	تعداد همسایه‌ها در SA	5
T0	دمای اولیه	100
Tf	دمای نهایی	0
Maxiter	تعداد تکرار الگوریتم	500
k	تعداد همسایه‌ها در KNN	10

معیار و متریک‌ها: در ارزیابی‌ها از معیار دقت (Accuracy) برای مقایسه استفاده می‌شود. معیار دقت متداول‌ترین معیار برای

[1] Weighted K-Nearest Neighbors
[2] Multi-Voter Multi-Commission Nearest Neighbor
[3] Simulated Anealing K-Nearest Neighbor

[DOI: 10.18502/ijdl.v25i5.20342]

روش‌های مبتنی بر طبقه‌بندی است. این معیار مشخص می‌کند که روش مورد بررسی تا چه میزان به درستی بیماری را تشخیص می‌دهد. هر چه مقدار دقت برای روشی بیشتر باشد، عملکرد آن روش بهتر است. از نظر مفهومی این معیار برابر با تعداد پیش‌بینی‌های صحیح صورت گرفته به تعداد کل پیش‌بینی‌های انجام شده است. معیار دقت از طریق رابطهٔ ۵ محاسبه می‌شود.

$$Acc = \frac{TP + TN}{TP + FP + TN + FN} \qquad (۵)$$

پارامترهای استفاده شده در این رابطه به شرح زیر هستند:

- مثبت واقعی[1] (TP): تعداد مواردی که به درستی بیمار تشخیص داده شده‌اند.
- مثبت کاذب[2] (FP) : تعداد مواردی که به اشتباه بیمار تشخیص داده شده‌اند.
- منفی واقعی[3] (TN): تعداد مواردی که به درستی رد شده‌اند و به درستی سالم تشخیص داده شده‌اند.
- منفی کاذب[4] (FN): تعداد مواردی که به اشتباه رد شده‌اند و به اشتباه سالم تشخیص داده شده‌اند.

برای مقایسهٔ روش‌ها براساس معیار دقت، ابتدا از اجرای مختلف استفاده شد. سپس در ادامه برای مقایسه از چهار متریک استفاده شد. یکی از متریک‌ها افزایش پارامتر k در الگوریتم کی- نزدیک‌ترین همسایه است. این پارامتر مشخص کننده تعداد همسایه‌های نزدیک به دادهٔ وروی است. متریک دیگر افزایش تعداد نمونه داده‌ها است. یعنی روند افزایش تعداد داده‌ها در روش‌ها، روی معیار دقت مورد بررسی قرار می‌گیرد. متریک دیگر افزایش تکرار الگوریتم فراابتکاری در روش‌ها است. متریک آخر افزایش تعداد اعضای جمعیت جواب‌ها است.

علاوه بر دقت، روش‌ها براساس معیارهای صحت (Precision)، فراخوانی (Recall) و امتیاز F1 (F1 Score) مورد مقایسه قرار گرفتند. این معیارها برای مقایسه الگوریتم‌های طبقه مورد استفاده قرار می‌گیرند. در انتها نیز روش‌ها از نظر زمان اجرا مقایسه شدند. هر چقدر میزان زمان اجرای روشی کمتر باشد، یادگیری و تشخیص سریع‌تر انجام می‌شود.

ارزیابی دقت در اجراهای مختلف

در این بخش نتایج ارزیابی‌های انجام شده برای مقایسهٔ روش‌ها ذکر می‌شود. برای مقایسه ابتدا روش‌ها در اجراهای مختلف و براساس دقت مورد بررسی قرار می‌گیرند. برای این منظور هر یک از روش‌ها ۱۵ بار به‌صورت مجزا اجرا شدند و در هر اجرا میزان دقت روش محاسبه شد. سپس بهترین دقت، میانگین دقت و انحراف از معیار برای هر روش محاسبه شد. نتایج حاصل از این ارزیابی‌ها در جدول ٤ نشان داده شده است.

جدول ۴- ارزیابی دقت روش‌ها در اجراهای مختلف

اجرا	WKNN	MVMCNN	SAKNN
۱	0.93	0.90	0.94
۲	0.90	0.93	0.93
۳	0.88	0.96	0.96
٤	0.87	0.92	0.98
۵	0.93	0.91	0.97
٦	0.91	0.96	0.97
۷	0.92	0.91	0.93
۸	0.87	0.96	0.95
۹	0.89	0.95	0.94
۱۰	0.90	0.90	0.97
۱۱	0.90	0.95	0.94
۱۲	0.88	0.96	0.98
۱۳	0.93	0.95	0.96
۱٤	0.91	0.90	0.95

[1] True positive
[2] False positive
[3] True negative
[4] False negative

[DOI: 10.18502/ijdl.v25i5.20342]

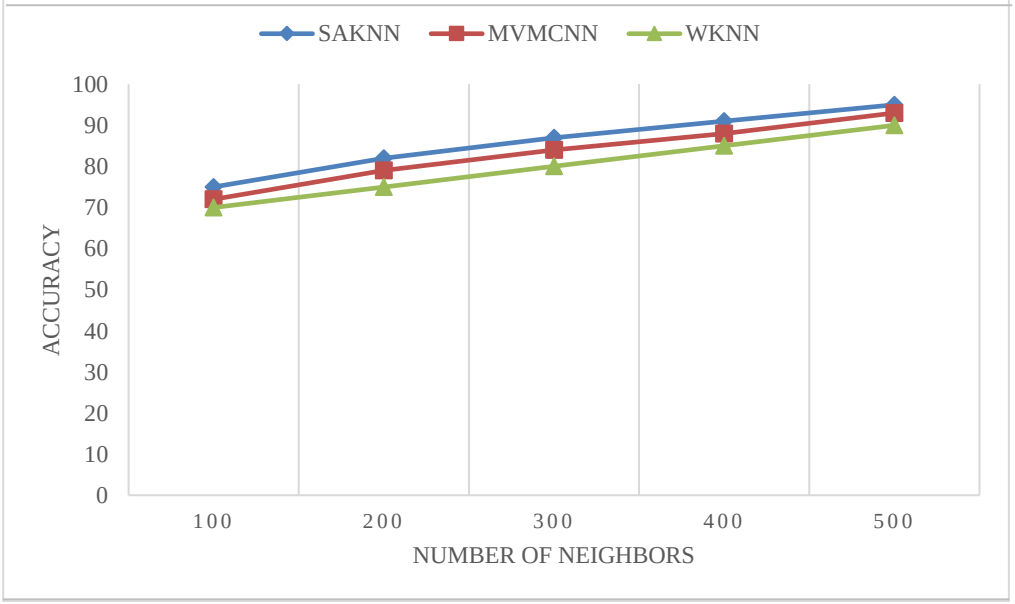
۱۵	0.92	0.94	0.98
بهترین	**0.93**	**0.96**	**0.98**
میانگین	**0.90**	**0.93**	**0.95**
انحراف از معیار	**0.0004**	**0.0005**	**0.0003**

همان طور که مشاهده می‌شود میزان دقت برای روش‌ها در ۱۵ اجرای مختلف محاسبه شده است. بهترین اجرای روش پیشنهادی یعنی SAKNN برابر با ۹۸ درصد است. بهترین اجرای روش MVMCNN برابر با ۹۶ درصد و بهترین اجرای روش WKNN برابر با ۹۳ درصد است. بنابراین بالاترین دقت را روش پیشنهادی ارائه کرده است.

از نظر میانگین دقت نیز روش پیشنهادی بهترین عملکرد را دارد. میانگین روش SAKNN برابر با ۹۵ درصد، روش MVMCNN برابر ۹۳ درصد و روش WKNN برابر با ۹۰ درصد است. از نظر انحراف از معیار نیز عملکرد روش پیشنهادی یعنی SAKNN از همه بهتر است. زیرا مقدار انحراف از معیار آن نسبت به دیگر روش‌ها نزدیک‌تر به صفر است. بنابراین در این ارزیابی‌ها بهترین عملکرد را روش پیشنهادی دارد که علت اصلی آن مدلسازی مناسب است.

ارزیابی دقت با متریک‌ها

در این بخش روش‌ها از نظر دقت و به کمک متریک‌ها مورد ارزیابی قرار می‌گیرند. ابتدا متریک افزایش تعداد همسایه‌ها در نظر گرفته می‌شود. برای این منظور روش‌ها با تعداد مختلفی از همسایه‌ها در الگوریتم KNN اجرا می‌شوند و دقت هر یک از روش‌ها محاسبه می‌گردد. تعداد همسایه‌ها همان پارامتر k است که در عملکرد روش‌ها تأثیرگذار است. در این ارزیابی‌ها برای پارامتر k مقادیر ۱۰۰، ۲۰۰، ۳۰۰، ۴۰۰ و ۵۰۰ در نظر گرفته شد. نتایج حاصل از این ارزیابی‌ها در شکل ۷ نشان داده شده است.

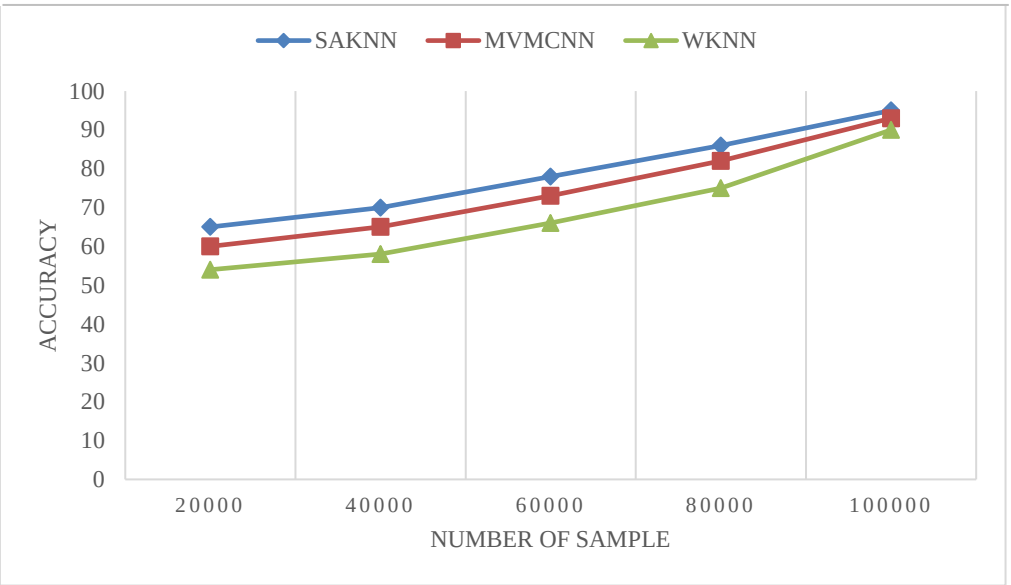


شکل ۷- ارزیابی دقت روش‌ها با افزایش تعداد همسایه‌ها

همان‌طور که مشاهده می‌شود با افزایش تعداد همسایه‌ها، دقت روش‌ها افزایش می‌یابد. زیرا با وجود تعداد همسایه‌های بیشتر، میتوان دقیق‌تر، تشخیص را انجام داد. در تمامی آزمایش‌ها، نمودار روش پیشنهادی یعنی SAKNN بالاتر از نمودار دیگر روش‌ها است. بنابراین در تمامی ارزیابی‌ها روش پیشنهادی دقت بالاتری ارائه می‌کند. در بین دو روش دیگر، روش MVMCNN دقت بالاتری نسبت به روش WKNN ارائه می‌کند. علت برتری روش پیشنهادی انتخاب ویژگی مناسب است که به کمک الگوریتم تبرید شبیه‌سازی شده انجام می‌شود. در ادامه روش‌ها از نظر دقت و با متریک افزایش تعداد نمونه داده‌ها مورد اررزیابی قرار می‌گیرند. منظور از نمونه داده‌ها همان اطلاعات افراد بیمار است. برای این منظور روش‌ها با تعداد مختلفی از داده‌ها اجرا می‌شوند و دقت هر یک از روش‌ها محاسبه می‌گردد. در این ارزیابی‌ها برای تعداد نمونه داده‌ها مقادیر ۲۰۰۰۰، ۴۰۰۰۰، ۶۰۰۰۰، ۸۰۰۰۰ و ۱۰۰۰۰۰ در نظر

[DOI: 10.18502/ijdl.v25i5.20342]

گرفته شد. نتایج حاصل از این ارزیابی‌ها در شکل ۸ نشان داده شده است.



شکل ۸- ارزیابی دقت روش‌ها با افزایش تعداد نمونه داده‌ها

همان‌طور که مشاهده می‌شود با افزایش تعداد نمونه داده‌ها، دقت ارائه شده توسط روش‌ها افزایش می‌یابد. زیرا در یادگیری ماشین و الگوریتم‌های طبقه‌بندی، هر چقدر مقدار داده‌ها بیشتر باشد، یادگیری بهتر صورت می‌گیرد و دقت تشخیص بالاتر می‌رود. در این ارزیابی‌ها نیز روش پیشنهادی یعنی SAKNN بهترین عملکرد را دارد و نمودار آن بالاتر از دیگر روش‌ها است. پس از آن، روش MVMCNN عملکرد بهتری از نظر دقت دارد. ضعیف‌ترین عملکرد را نیز روش WKNN دارد و نمودار آن در تمامی ارزیابی‌ها پایین‌تر از دو روش دیگر است.

در ادامه روش‌ها از نظر دقت و با متریک افزایش تعداد تکرار الگوریتم مورد اررزیابی قرار می‌گیرند. برای این منظور روش‌ها با تعداد تکرار مختلف الگوریتم اجرا می‌شوند و دقت هر یک از روش‌ها محاسبه می‌گردد. در این ارزیابی‌ها برای تعداد تکرار الگوریتم‌ها مقادیر ۱۰۰، ۲۰۰، ۳۰۰، ۴۰۰ و ۵۰۰ در نظر گرفته شد. نتایج حاصل از این ارزیابی‌ها در شکل ۹ نشان داده شده است.

شکل ۹- ارزیابی دقت روش‌ها با افزایش تعداد تکرار الگوریتم

[DOI: 10.18502/ijdl.v25i5.20342]

همان طور که مشاهده می‌شود با افزایش تعداد تکرار الگوریتم، دقت ارائه شده توسط روش‌ها افزایش می‌یابد. زیرا هرچه تکرار الگوریتم بالاتر باشد ویژگی‌های بهتری انتخاب می‌شوند و دقت تشخیص بالاتر می‌رود. در این ارزیابی‌ها نیز نمودار روش پیشنهادی یعنی SAKNN بالاتر از دیگر روش‌ها است.

در ادامه روش‌ها از نظر دقت و با متریک افزایش تعداد اعضای جمعیت جواب‌ها مورد اررزیابی قرار می‌گیرند. برای این منظور روش‌ها با تعداد اعضای مختلف جمعیت جواب‌ها اجرا می‌شوند و دقت هر یک از روش‌ها محاسبه می‌گردد. در این ارزیابی‌ها برای تعداد اعضای جمعیت جواب‌ها مقادیر ۶۰، ۷۰، ۸۰، ۹۰ و ۱۰۰ در نظر گرفته شد. نتایج حاصل از این ارزیابی‌ها در شکل ۱۰ نشان داده شده است.

شکل ۱۰- ارزیابی دقت روش‌ها با افزایش تعداد اعضای جمعیت جواب‌ها

همان‌طور که مشاهده می‌شود با افزایش تعداد اعضای جمعیت جواب‌ها، دقت ارائه شده توسط روش‌ها افزایش می‌یابد. زیرا تعداد جواب‌ها بیشتر است و از طرق بیشتری می‌توان در فضای مسئله جستجو کرد و در نتیجه دقت تشخیص بالاتر می‌رود. در این ارزیابی‌ها نیز روش پیشنهادی یعنی SAKNN بهترین عملکرد را دارد. پس از آن، روش MVMCNN عملکرد بهتری داشته و ضعیف‌ترین عملکرد برای روش WKNN است.

در ارزیابی‌های عملی، روش‌ها از نظر دقت با پنج حالت مختلف بررسی شدند. یکی از این حالات مقایسه در اجراهای مختلف و محاسبه بهترین، میانگین و انحراف معیار بود. حالت دیگر مقایسه با افزایش تعداد همسایه‌ها بود. حالت بعدی مقایسه با افزایش تعداد نمونه داده‌ها بود. حالت بعدی مقایسه با افزایش تعداد تکرار الگوریتم بود. حالت آخر مقایسه با افزایش تعداد اعضای جمعیت بود. در تمامی ارزیابی‌ها روش پیشنهادی دقت بالاتری را از دیگر روش‌ها ارائه داد. علت این برتری مدلسازی مناسب روش پیشنهادی و استفاده از الگوریتم تبرید شبیه‌سازی شده در انتخاب ویژگی است. بنابراین در حالت کلی می‌توان نتیجه گرفت که روش پیشنهادی عملکرد بهتری از سایر روش‌های مشابه در این زمینه دارد.

ارزیابی سایر پارامترها

در این بخش روش‌ها براساس معیارهای صحت، فراخوانی و امتیاز F1 مورد ارزیابی قرار می‌گیرند. این معیارها برای ارزیابی عملکرد الگوریتم‌های طبقه‌بندی بسیار مهم هستند. معیار صحت نسبت موارد مثبت درست پیش‌بینی شده به کل موارد پیش‌بینی شده مثبت است و نشان می‌دهد که چه درصدی از پیش‌بینی‌های مثبت درست هستند. معیار فراخوانی نسبت موارد مثبت درست پیش‌بینی شده به کل موارد مثبت واقعی است و نشان می‌دهد که چه درصدی از موارد مثبت واقعی درست تشخیص داده شده‌اند. معیار امتیاز F1 میانگین هارمونیک صحت و فراخوانی است که ترکیبی از هر دو معیار را ارائه می‌دهد و برای زمانی که تعادل بین صحت و فراخوانی مورد نیاز است، مفید است. در جدول ۵ یک نمونه از مقادیر این معیارها برای سه روش اشاره شده در یک آزمایش نمونه ذکر شده است.

[DOI: 10.18502/ijdl.v25i5.20342]

جدول ۵= ارزیابی روش‌ها بر اساس پارامترهای صحت، فراخوانی و امتیاز F1

معیار	WKNN	MVMCNN	SAKNN
صحت	0.88	0.91	0.94
فراخوانی	0.87	0.92	0.95
امتیاز F1	0.87	0.91	0.94

با توجه به مقادیر برای معیارهای صحت، فراخوانی و امتیاز F1، روش SAKNN با صحت برابر ۰/۹٤، فراخوانی برابر ۰/۹۵ و امتیاز F1 برابر ۰/۹٤، عملکرد بهتری نسبت به دو روش دیگر دارد. این نتایج نشان می‌دهد که روش پیشنهادی در تشخیص موارد مثبت و درست پیش‌بینی کردن آنها بسیار قوی عمل کرده است. روش MVMCNN نیز با صحت برابر ۰/۹۱، فراخوانی برابر ۰/۹۲ و امتیاز F1 برابر ۰/۹۱، عملکرد بهتری نسبت به WKNN دارد و نشان می‌دهد که در تشخیص دیابت مؤثرتر است. روش WKNN با صحت برابر ۰/۸۸، فراخوانی برابر ۰/۸۷ و امتیاز F1 برابر ۰/۸۷، اگرچه عملکرد قابل قبولی دارد، اما نسبت به دو روش دیگر ضعیف‌تر است. بنابراین، روش پیشنهادی یعنی SAKNN بهترین عملکرد را در بین این سه روش در تشخیص دیابت ارائه می‌دهد.

ارزیابی زمانی

در این بخش روش‌ها از نظر زمان اجرا مورد ارزیابی قرار می‌گیرند. این زمان شامل زمان انتخاب ویژگی و همچنین زمان یادگیری و تشخیص بیماری است. هرچقدر زمان اجرای روشی کمتر باشد، عملکرد آن بهتر است و تشخیص بیماری با سرعت بیشتری انجام شده است. برای این ارزیابی‌ها روش‌ها هر کدام پنج مرتبه اجرا شدند و زمان اجرای آنها براساس ثانیه به‌صورت جداگانه محاسبه شد. نتایج حاصل از این ارزیابی‌ها در جدول ٦ ذکر شده است.

جدول ۶- ارزیابی زمانی روش‌ها

اجرا	WKNN	MVMCNN	SAKNN
۱	٦٥	۷۲	٦۷
۲	٦۳	۷۰	٦٥
۳	٦٥	۷۳	٦٥
٤	٦۲	۷۰	٦٦
٥	٦٤	۷۳	٦٥

همان‌طور که قابل مشاهده است زمان اجرای روش WKNN از همهٔ روش‌ها کمتر است و بهترین عملکرد را دارد. بعد از آن زمان اجرای روش پیشنهادی یعنی SAKNN بهتر از روش دیگر است. بیشترین زمان اجرا را روش MVMCNN دارد. با توجه به مقادیر این جدول می‌توان متوجه شد که میزان فاصله اجرای بین روش‌ها خیلی زیاد نیست. به‌خصوص زمان اجرای روش SAKNN به روش WKNN نزدیک است. بنابراین می‌توان گفت که عملکرد روش پیشنهادی از نظر زمانی قابل قبول است.

بحث و نتیجه‌گیری

دیابت نوعی بیماری متابولیکی و یکی از شایع‌ترین بیماری‌های غدد است که با سطح قند خون بالا یا همان گلوکز همراه است و به‌دنبال نقص در ترشح انسولین یا عملکرد آن یا هر دو ایجاد می‌شود. بیماری دیابت در سال‌های اخیر با سرعت بالایی در بین افراد گسترش یافته است. این گسترش در کشورهای پیشرفته و همچنین در حال توسعه بیشتر است. یکی از مهم‌ترین دلایل آن مصرف بی‌رویه مواد حاوی قند است. با توجه به این مورد تشخیص زودهنگام آن دارای اهمیت است. در سال‌های اخیر محققان به روش‌هایی روی آورده‌اند که این بیماری را به‌صورت غیرتهاجمی و خودکار تشخیص دهند. بیشتر بین روش‌های خودکاری که برای تشخیص بیماری دیابت ارائه شده‌اند بر پایه یادگیری ماشین است.

در این مقاله یک روش جدید برای تشخیص بیماری دیابت بر پایه الگوریتم‌های فراابتکاری و یادگیری ماشین ارائه شد. در روش پیشنهادی از الگوریتم فرا ابتکاری برای انتخاب ویژگی

[DOI: 10.18502/ijdl.v25i5.20342]

استفاده شد. الگوریتم فرا ابتکاری استفاده شده، الگوریتم تبدرید شبیه‌سازی شده است. برای تشخیص بیماری قند خون نیز از یادگیری ماشین و الگوریتم‌های طبقه‌بندی استفاده شد. الگوریتم طبقه‌بندی استفاده شده در روش پیشنهادی الگوریتم کی-نزدیک‌ترین همسایه است. در مرحلهٔ اول روش پیشنهادی انتخاب ویژگی صورت می‌گیرد و از بین ویژگی‌های موجود بهترین‌ها انتخاب می‌شوند. در مرحلهٔ دوم تشخیص بیماری صورت می‌گیرد. در مقایسه‌های عملی مشخص شد که روش پیشنهاد شده برای تشخیص بیماری دیابت عملکرد بهتری از نظر دقت نسبت به دیگر روش‌های مشابه در این زمینه دارد و تشخیص را با دقت بالاتر و خطای کمتری انجام می‌دهد.

یکی از محدودیت‌های مهم این تحقیق، تأثیر عوامل محیطی و شرایط زندگی افراد مانند تغذیه، فعالیت بدنی، سطح استرس، و غیره بر دقت تشخیص دیابت است. درصورتی‌که این عوامل به‌صورت کامل در داده‌های آموزشی و آزمون لحاظ نشده باشند، مدل‌های تشخیصی ممکن است دقت کمتری داشته و نتایج ضعیفی ارائه دهند. بنابراین، عدم توجه به این عوامل محیطی می‌تواند محدودیت مهمی در تعمیم‌پذیری و دقت نتایج تحقیق باشد. همچنین نمونه‌های مورد استفاده در این تحقیق ممکن است تنوع کافی نداشته باشند. زیرا داده‌های استفاده شده تنها مربوط به یک منطقهٔ خاص است و ممکن است نتایج به‌طور کامل قابل تعمیم به دیگر مناطق نباشد.

یک راهکار آتی برای بهبود این تحقیق می‌تواند شامل جمع‌آوری و تحلیل داده‌های مرتبط با عوامل محیطی و شرایط زندگی افراد باشد. به این صورت که در مراحل جمع‌آوری داده‌ها، اطلاعاتی مانند تغذیه، سطح فعالیت بدنی، میزان استرس، وضعیت اقتصادی-اجتماعی و دسترسی به مراقبت‌های بهداشتی نیز در نظر گرفته شوند. با استفاده از این داده‌های اضافی، مدل‌های هوش مصنوعی می‌توانند آموزش ببینند و نتایج دقیق‌تر و جامع‌تری ارائه دهند. همچنین، به‌کارگیری الگوریتم‌های پیشرفته‌تر و ترکیب آنها با تکنیک‌های فراابتکاری می‌تواند به بهبود دقت تشخیص دیابت کمک کند. افزایش همکاری با مراکز درمانی و تحقیقاتی نیز می‌تواند منجر به ایجاد بانک‌های داده‌ای بزرگ‌تر و متنوع‌تر شود که نتایج را قابل تعمیم‌تر خواهد کرد.

تعارض منافع

مؤلفان اظهار می‌کنند که منافع متقابلی از تألیف و انتشار این مقاله وجود ندارند.

سپاسگزاری

نویسندگان مقاله از تمامی محققانی که در زمینه مربوط به این مقاله پژوهش نمودند و نقش موثری در شکل‌گیری پژوهش ما داشتند، سپاسگزاری می‌نمایند.

References

1. Khan RMM, Chua ZJY, Tan JC, Yang Y, Liao Z, & Zhao Y. From pre-diabetes to diabetes: diagnosis, treatments and translational research. *Medicina*. 2019; 55(9): 546.
2. Janiesch C, Zschech P, & Heinrich K. Machine learning and deep learning. *Electronic Markets*. 2021; 31(3):685-695.
3. Azgomi H, & Sohrabi MK. MR-MVPP: A map-reduce-based approach for creating MVPP in data warehouses for big data applications. *Information Sciences*. 2021; 570: 200-224.
4. Teymournezhad K, Azgomi H, & Asghari A. Detection of counterfeit banknotes by security components based on image processing and GoogleNet deep learning network. *Signal, Image and Video Processing*. 2022; 16(6):1505-1513.
5. Asghari A, Azgomi H, & Darvishmofarahi Z. Multi-Objective edge server placement using the whale optimization algorithm and Game theory. *Soft Computing*. 2023; 1-15.
6. Sen PC, Hajra M, Ghosh M. *Supervised Classification Algorithms in Machine Learning: A Survey and Review*. In: Mandal, J., Bhattacharya, D. (eds) Emerging Technology in Modelling and Graphics. Advances in Intelligent Systems and Computing, vol 937. Springer, Singapore, 2020.
7. Chirici G, Mura M, McInerney D, Py N, Tomppo EO, Waser LT, Travaglini D, Ronald ME. A meta-analysis and review of the literature on the k-Nearest Neighbors technique for forestry applications that use remotely sensed data. *Remote Sensing of Environment*. 2016; 176:282-294.
8. Alimoradi M, Azgomi H, & Asghari A. Trees social relations optimization algorithm: A new Swarm-Based metaheuristic technique to solve continuous and discrete optimization problems. *Mathematics and Computers in Simulation*. 2022; 194:629-664.
9. Daliri A, Asghari A, Azgomi H, & Alimoradi M. The water optimization algorithm: a novel metaheuristic for solving optimization problems. *Applied Intelligence*. 2022; 52(15):17990-18029.
10. Delahaye D, Chaimatanan S, & Mongeau M. Simulated annealing: From basics to applications. *Handbook of metaheuristics*. 2019; 1-35.
11. Calisir D, & Dogantekin E. An automatic diabetes diagnosis system based on LDA-Wavelet Support Vector Machine Classifier. *Expert Systems with Applications*. 2011; 38(7):8311-8315.

[DOI: 10.18502/ijdl.v25i5.20342]

12. Saxena K, Khan Z, & Singh S. Diagnosis of diabetes mellitus using k nearest neighbor algorithm. *International Journal of Computer Science Trends and Technology (IJCST)*. 2014; 2(4), 36-43.
13. Nai-Arun N, & Moungmai R. Comparison of classifiers for the risk of diabetes prediction. *Procedia Computer Science*. 2015; 69, 132-142.
14. Perveen S, Shahbaz M, Guergachi A, & Keshavjee K. Performance analysis of data mining classification techniques to predict diabetes. *Procedia Computer Scienc*e. 2016; 82:115-121.
15. Sisodia D, & Sisodia DS. Prediction of diabetes using classification algorithms. *Procedia computer science*. 2018; 132:1578-1585.
16. Komal Kumar N, Vigneswari D, Vamsi Krishna M, & Phanindra Reddy GV. *An optimized random forest classifier for diabetes mellitus. In Emerging Technologies in Data Mining and Information Security: Proceedings of IEMIS*. 2018; 2:765-773, Springer Singapore.
17. Ali A, Alrubei MA, Hassan LFM, Al-Ja'afari MA, & Abdulwahed SH. (). Diabetes Diagnosis based on KNN. *IIUM Engineering Journal*. 2020; 21(1);175-181.
18. Priya KL, Kypa MSCR, Reddy MMS, & Reddy GRM. *A novel approach to predict diabetes by using Naive Bayes classifier. In 2020 4th International Conference on Trends in Electronics and Informatics* (ICOEI)(48184). 2020; (pp:603-607). IEEE.
19. Khaleel FA, & Al-Bakry AM. Diagnosis of diabetes using machine learning algorithms. *Materials Today: Proceedings*. 2023; 80(3):3200-3203.
20. Suyanto S, Meliana S, Wahyuningrum T, & Khomsah S. A new nearest neighbor-based framework for diabetes detection. *Expert Systems with Applications*. 2022; 199:116857.
21. Prasad BS, Gupta S, Borah N, Dineshkumar R, Lautre HK, & Mouleswararao B. Predicting diabetes with multivariate analysis an innovative KNN-based classifier approach. *Preventive Medicine*. 2023; 174, 107619.
22. Chou CY, Hsu DY, & Chou CH. Predicting the onset of diabetes with machine learning methods. *Journal of Personalized Medicine*. 2023; 13(3), 406.
23. Shabdiz M, Azarbar A, & Azgomi H. Pattern coloring By D2NN and FLSA-SVM based on Probabilistic ML model for diabetic patient. *Multimedia Tools and Applications*. 2024; 1-28.
24. Shabdiz M, Azarbar A, & Azgomi H. Hyperbolic decision-making based on RBF for diabetic/COVID-19 iris. *Engineering Applications of Artificial Intelligence*. 2024; 130:107589.
25. Reddy LR, Akhila V, Nivas LL, & Tharun P. Prediction of Diabetes by Machine Learning Algorithm. *In 2025 International Conference on Multi-Agent Systems for Collaborative Intelligence (ICMSCI)*. 2025; (pp. 1241-1244). IEEE.
26. Tiwari E, Gupta S, Pavulla A, Al-Maini M, Singh, R, Isenovic ER, & Suri JS. Artificial intelligence-based multiclass diabetes risk stratification for big data embedded with explainability: From machine learning to attention models. *Biomedical Signal Processing and Control*. 2025; 106:107672.

[DOI: 10.18502/ijdl.v25i5.20342]