

بهبود پیش‌بینی دیابت نوع ۲ با الگوریتم‌های طبقه‌بندی: بررسی و تحلیل ویژگی‌های مؤثر

زهرا محمدی فرد

پارسا ذاکر حیات

محمد گل زوری

کارشناسی ارشد

کارشناسی ارشد

کارشناسی ارشد

Zahra.mohammadifard@edu.ikiu.ac.com

Parsa.zakerhayat@edu.ikiu.ac.com

Mohammadgol.zoori@edu.ikiu.ac.com

چکیده - دیابت نوع ۲ یکی از بیماری‌های متابولیکی شایع است که در صورت تشخیص دیر هنگام می‌تواند منجر به بروز عوارض حادی مانند مشکلات قلبی و عروقی، آسیب‌های عصبی و نارسایی کلیه منجر شود. لذا، تشخیص زودهنگام این بیماری بسیار حیاتی است چراکه می‌تواند از بروز شدید عوارض جلوگیری کرده و کیفیت زندگی بیمار را بهبود دهد. با توسعه فناوری‌های داده‌کاوی و هوش مصنوعی، الگوریتم‌های یادگیری ماشین به ابزاری برای تحلیل داده‌های بالینی و پیش‌بینی بیماری دیابت بدل شده‌اند. در این مقاله عملکرد سه الگوریتم پراهمیت جنگل تصادفی^۱، ماشین بردار پشتیبان^۲ و شبکه‌های عصبی مصنوعی^۳ بر روی مجموعه‌ای از داده‌ها مورد بررسی و مقایسه قرار گرفته است. پیش‌پردازش داده‌ها در چند مرحله شامل: حذف مقادیر گم‌شده، نرمال‌سازی و انتخاب ویژگی‌های مؤثر که گام‌های اولیه در تحلیل داده‌ها هستند، صورت گرفته است. عملکرد هر الگوریتم با در نظر گرفتن معیارهایی چون حساسیت^۴، دقت^۵، اختصاص پذیری^۶ و دقت پیش‌بینی^۷ مورد بررسی قرار گرفته است. باتوجه به نتایج به دست آمده الگوریتم جنگل تصادفی عملکرد بهتری در تشخیص بیماران مبتلا به دیابت نوع ۲ ارائه می‌دهد.

کلیدواژه - داده‌کاوی، تشخیص زودهنگام، دیابت نوع ۲، یادگیری ماشین.

1- مقدمه

با توجه به پیچیدگی‌های بالینی و اقتصادی ناشی از دیابت نوع ۲، فناوری‌های نوین مانند داده‌کاوی و هوش مصنوعی می‌تواند در بهبود تشخیص زودهنگام و مدیریت این بیماری مؤثر واقع شوند. الگوریتم‌هایی مانند ماشین بردار پشتیبان، شبکه‌های عصبی مصنوعی و جنگل تصادفی به دلیل توانایی پردازش حجم زیاد داده‌های بالینی و تشخیص الگوهای پنهان در داده‌های متابولیکی، کاربرد زیادی در تشخیص دیابت نوع ۲ پیدا کرده‌اند [2].

این مقاله به منظور بررسی و مقایسه عملکرد الگوریتم‌های ذکر شده در پیش‌بینی و مدیریت دیابت نوع ۲ تهیه شده است. از مجموعه داده دیابت سرخپوستان پیمایا (PIMA) همراه با نرم‌افزار وکا جهت اجرای شبیه‌سازی استفاده شده و مراحل پیش‌پردازش داده‌ها - شامل حذف مقادیر

دیابت نوع ۲ به عنوان یکی از چالش‌های جدی سلامت جهانی، به دلیل شیوع رو به رشد و عوارض شدید ناشی از آن همواره توجه متخصصان بهداشت را جلب کرده است. با استناد به گزارش‌های اخیر سازمان‌های بین‌المللی مانند فدراسیون بین‌المللی دیابت در سال ۲۰۱۹، حدود ۹.۳٪ از جمعیت بزرگسال جهان (معادل ۴۶۳ میلیون نفر) به این بیماری مبتلا بوده‌اند و پیش‌بینی می‌شود این درصد تا سال ۲۰۴۵ به ۱۰.۹٪ (حدود ۷۰۰ میلیون نفر) برسد. رشد شهرنشینی، تغییرات سبک زندگی، افزایش چاقی و عدم آگاهی از وضعیت بیماری از مهمترین عوامل افزایش شیوع دیابت می‌باشند؛ به طوری که حدوداً نیمی از افراد مبتلا، از بیماری خود مطلع نیستند [1].

^۱ Type 2 diabetes

^۲ Random Forest

^۳ Support Vector Machine (SVM)

^۴ Artificial Neural Network (ANN)

^۵ Recall

^۶ Accuracy

^۷ Specificity

^۸ Precision

پارامتر C در این الگوریتم تعیین کننده این است که مدل تا چه حد باید داده‌ها را دقیقاً طبقه‌بندی کند یا مقداری خطا را بپذیرد تا از بیش‌برازش جلوگیری شود. به‌طور کلی، ماشین بردار پشتیبان با داشتن دقت بالا، به پزشکان و پژوهشگران در تشخیص زودهنگام دیابت کمک کرده و می‌تواند به‌عنوان یک ابزار تصمیم‌گیری مفید در کنار روش‌های پزشکی سنتی مورد استفاده قرار می‌گیرد [5].

2-2- شبکه‌های عصبی مصنوعی

شبکه‌های عصبی مصنوعی به‌عنوان یک روش پیشرفته یادگیری ماشین درپیش‌بینی دیابت نوع ۲ پر کاربرد هستند. این الگوریتم‌ها از ساختار مغز انسان الهام گرفته‌اند و می‌توانند روابط پیچیده و غیرخطی میان ویژگی‌هایی مانند سطح گلوکز خون، شاخص توده بدنی و سن فرد را شبیه‌سازی کنند. و با دقت بالا الگوهای پنهان در داده‌ها را استخراج کنند [6].

ویژگی‌هایی مانند سطح گلوکز خون، فشار خون، شاخص توده بدنی، سطح انسولین، سن و سابقه خانوادگی، همگی به عنوان ورودی‌های مدل شبکه عصبی استفاده می‌شوند و پس از پردازش در لایه‌های مختلف، نتیجه نهایی به‌صورت احتمال ابتلا یا عدم ابتلا به دیابت ارائه می‌شود. به دلیل توانایی بالا در مدل‌سازی الگوهای پیچیده، شبکه‌های عصبی می‌توانند پیش‌بینی دقیقی از ابتلا به دیابت نوع ۲ ارائه دهند. مزیت مهم شبکه‌های عصبی، توانایی آن‌ها در یادگیری از داده‌های حجیم و کشف وابستگی‌های پیچیده‌ای است که روش‌های سنتی قادر به تشخیص آن‌ها نیستند این روش به‌ویژه در شرایطی که داده‌ها وابستگی‌های غیرخطی دارند، عملکرد مؤثر و مفیدی دارند [7].

2-3- الگوریتم جنگل تصادفی

جنگل تصادفی یک الگوریتم یادگیری ماشین جمع‌سپاری است که از مجموعه‌ای از درخت‌های تصمیم برای انجام پیش‌بینی استفاده می‌کند. در این الگوریتم، برای هر داده ورودی، چندین درخت تصمیم به‌طور تصادفی ساخته می‌شوند و سپس از رأی اکثریت درخت‌ها برای پیش‌بینی خروجی استفاده می‌شود. هر درخت تصمیم به‌طور مستقل روی یک زیرمجموعه تصادفی از داده‌ها آموزش می‌بیند و در نهایت، پیش‌بینی نهایی بر اساس رأی‌گیری اکثریت درخت

گمشته، نرمال‌سازی و انتخاب ویژگی‌های مؤثر - صورت گرفته است. معیارهایی مانند دقت، حساسیت، اختصاص‌پذیری و ارزیابی‌های عددی خطا برای بررسی عملکرد الگوریتم‌ها به کار گرفته شده‌اند. هدف اصلی این مقاله ارائه یک مدل بهینه به منظور تشخیص زودهنگام دیابت نوع ۲ و کمک به تصمیم‌گیری‌های درمانی در جهت کاهش عوارض و افزایش کیفیت زندگی بیماران می‌باشد [3].

در ادامه، مراحل مختلف پردازش داده، الگوریتم‌های بررسی شده و نتایج به دست آمده تشریح خواهند شد.

2- الگوریتم‌های بررسی شده

2-1- ماشین بردار پشتیبان

یکی از محبوب‌ترین الگوریتم‌های یادگیری ماشین، ماشین بردار پشتیبان است که برای طبقه‌بندی و پیش‌بینی در تشخیص دیابت نوع ۲ استفاده می‌شود. این الگوریتم ویژگی‌هایی مانند سطح گلوکز خون، فشار خون، شاخص توده بدنی، سطح انسولین، سن و سابقه خانوادگی را بررسی کرده و با یافتن یک مرز تصمیم‌گیری بهینه، بیماران را از افراد سالم جدا می‌کند.

ماشین بردار پشتیبان ابرصفحه ایجاد می‌کند که داده‌های دو کلاس را با بیشترین فاصله ممکن تفکیک می‌کند. در صورتی که داده‌ها به‌صورت خطی قابل جداسازی نباشند، از توابع کرنل مانند چندجمله‌ای یا گوسی برای نگاشت داده‌ها به فضای بعد بالاتر استفاده می‌شود. این الگوریتم روی مجموعه‌های داده کوچک هم عملکرد قابل قبولی دارد و در برابر بیش‌برازش مقاوم است [4].

یکی از ویژگی‌های کلیدی و متمایز این الگوریتم، انتخاب بهینه ابرصفحه‌ای است که داده‌های مربوط به بیماران دیابتی و غیردیابتی را به بهترین شکل از هم جدا می‌کند. در حالتی که داده‌ها به‌صورت خطی قابل تفکیک نباشند، ماشین بردار پشتیبان با استفاده از ترفندهایی مانند تبدیل داده‌ها به فضای با ابعاد بالاتر و به‌کارگیری کرنل‌ها، مرز تصمیم‌گیری بهتری ایجاد می‌کند.

کرنل‌هایی مانند کرنل گوسی یا چندجمله‌ای امکان نگاشت داده‌ها به فضایی با بعد بالاتر را فراهم می‌کنند تا بتوان مرز تصمیم‌گیری مناسبی ایجاد کند. انتخاب مقدار مناسب برای

3- روش شناسی

3-1- مجموعه داده

در این پژوهش، از مجموعه داده دیابت سرخپوستان پیما برای پیش‌بینی دیابت نوع ۲ استفاده شده است. این مجموعه داده شامل ۷۶۸ نمونه است که هر نمونه دارای ۸ ویژگی کلیدی مرتبط با دیابت می‌باشد. داده‌ها از بانک اطلاعاتی مؤسسه ملی دیابت و بیماری‌های گوارشی و کلیوی^۹ تهیه شده و در بسیاری از مطالعات علمی به‌عنوان یک منبع معتبر برای تحلیل دیابت به کار رفته است.

برای اجرای آزمایش‌ها و پیاده‌سازی مدل‌های یادگیری ماشین، از نرم‌افزار وکا^{۱۰} استفاده شده است. این ابزار به دلیل قابلیت‌های قدرتمند در پردازش داده‌ها و اجرای الگوریتم‌های مختلف، انتخاب شده است. داده‌ها در وکا بارگذاری شده و پس از انجام مراحل پیش‌پردازش، برای آموزش مدل‌ها مورد استفاده قرار گرفته‌اند.

ویژگی‌های موجود در مجموعه داده عبارت‌اند از: تعداد دفعات بارداری^{۱۱}، سطح گلوکز خون^{۱۲}، فشارخون^{۱۳}، ضخامت پوست^{۱۴}، میزان انسولین^{۱۵}، شاخص توده بدنی، سابقه خانوادگی دیابت^{۱۶} و سن.

پس از بارگذاری این داده‌ها در نرم‌افزار وکا، ابتدا پیش‌پردازش انجام شده تا مقادیر گمشده برطرف شوند و داده‌ها برای یادگیری مدل‌ها بهینه شوند. سپس، الگوریتم‌های مختلف یادگیری ماشین روی داده‌ها اعمال شده و نتایج ارزیابی شدند.

^۹ National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK)

^{۱۰} Weka

^{۱۱} Pregnancies

^{۱۲} Glucose

^{۱۳} Blood Pressure

^{۱۴} Skin Thickness

^{۱۵} Insulin

^{۱۶} Diabetes Pedigree Function

ها انجام می‌شود. این ساختار باعث افزایش دقت مدل و کاهش مشکل بیش‌برازش در مقایسه با یک درخت تصمیم منفرد می‌شود. جنگل تصادفی معمولاً نسبت به استفاده از یک درخت تصمیم تنها، دقت بسیار بالاتری دارد [8].

یکی از ویژگی‌های کلیدی جنگل تصادفی، توانایی آن در پردازش داده‌های پیچیده و غیرخطی است. همچنین، این روش به دلیل استفاده از چندین درخت تصمیم، حساسیت کمتری نسبت به داده‌های پرت دارد و می‌تواند با ترکیب پیش‌بینی‌های چندین مدل، دقت بالاتری در تشخیص بیماران دیابتی داشته باشد. همین ویژگی، جنگل تصادفی را به گزینه‌ای مناسب برای تشخیص دیابت نوع ۲ تبدیل کرده است. این الگوریتم توانایی بالایی در شناسایی وابستگی‌های پیچیده بین ویژگی‌های مختلف دارد و می‌تواند اهمیت نسبی هر ویژگی را تعیین کند [8].

از مزایای مهم جنگل تصادفی می‌توان به دقت بالا در مقایسه با بسیاری از الگوریتم‌های دیگر، به‌ویژه زمانی که داده‌های پیچیده و حجیم وجود دارند، اشاره کرد. این الگوریتم توانایی مقابله با داده‌های ناقص و پرت را دارد و همچنین تفسیر مدل را تسهیل می‌کند، زیرا می‌توان دلیل پیش‌بینی‌ها را از طریق بررسی مجموعه‌ای از درخت‌ها تحلیل کرد. علاوه بر این، جنگل تصادفی می‌تواند اهمیت نسبی هر ویژگی را مشخص کند، که در تشخیص دیابت نوع ۲ کمک می‌کند تا تأثیر هر عامل بر پیش‌بینی بیماری بهتر درک شود.

4-2- فاصله علمی و نوآوری مطالعه حاضر

با وجود مطالعات متعدد در زمینه پیش‌بینی دیابت نوع ۲، همچنان چالش‌هایی مانند انتخاب ویژگی‌های مؤثر، بهبود دقت مدل‌ها و کاهش هزینه‌های محاسباتی باقی مانده است. این پژوهش با مقایسه چندین الگوریتم یادگیری ماشین، از جمله بردار ماشین پشتیبان، جنگل تصادفی و شبکه‌های عصبی مصنوعی، به ارزیابی دقیق‌تر عملکرد آن‌ها می‌پردازد.

نوآوری این مطالعه شامل مقایسه جامع مدل‌ها، بهبود پیش‌پردازش داده‌ها و ارزیابی عملکرد بر اساس معیارهای متنوع است. هدف این پژوهش ارائه یک مدل بهینه برای پیش‌بینی دیابت نوع ۲ است که می‌تواند به تشخیص زودهنگام و بهبود تصمیم‌گیری‌های پزشکی کمک کند.

3-2- پیش پردازش داده ها

در این مقاله، مجموعه داده‌ی دیابت نوع ۲ پس از بررسی اولیه، تحت فرآیند پیش‌پردازش قرار گرفت تا کیفیت و دقت مدل‌های یادگیری ماشین بهبود یابد. ابتدا مقادیر گم‌شده شناسایی و با روش‌های مناسب جایگزین شدند. سپس، ویژگی‌های پرت اصلاح یا حذف گردیدند تا داده‌ها ساختار بهتری پیدا کنند. در گام بعدی، برای یکسان‌سازی مقیاس ویژگی‌ها، داده‌ها نرمال‌سازی یا استانداردسازی شدند. در نهایت، مجموعه داده‌ی پردازش‌شده در نرم‌افزار وکا بارگذاری شد تا الگوریتم‌های مختلف یادگیری ماشین روی آن آموزش داده شده و ارزیابی شوند.

4- روش‌های ارزیابی مدل

در ارزیابی مدل‌های یادگیری ماشین برای پیش‌بینی دیابت نوع ۲، چهار معیار اصلی دقت، حساسیت، اختصاص‌پذیری و دقت پیش‌بینی نقش کلیدی در سنجش عملکرد مدل ایفا می‌کنند و تصویری دقیق از کیفیت پیش‌بینی‌ها ارائه می‌دهند.

دقت بیان می‌کند که مدل چه درصدی از نمونه‌های کل را به‌درستی پیش‌بینی کرده است. به عبارت دیگر، این معیار نسبت تمام پیش‌بینی‌های صحیح (بیماران تشخیص داده شده و افراد سالم) را به کل نمونه‌ها اندازه‌گیری می‌کند. اگرچه دقت می‌تواند نمای کلی خوبی از عملکرد مدل باشد، اما در شرایطی که تعداد بیماران و افراد سالم در مجموعه داده نامتوازن است، ممکن است تصویری گمراه‌کننده ایجاد کند [7].

در کنار دقت، حساسیت و اختصاص‌پذیری مکمل‌های مهمی برای ارزیابی مدل هستند. حساسیت نشان می‌دهد چه بخشی از بیماران واقعی را مدل به‌درستی شناسایی کرده است. مقدار بالای حساسیت اهمیت بالایی در مسائل پزشکی دارد؛ چرا که از دست دادن موارد واقعی بیماری می‌تواند خطرناک باشد.

$$Recall = \frac{TP}{TP+FN} \quad (1)$$

	مثبت	منفی
مثبت	TP	FN
منفی	FP	TN

اختصاص‌پذیری نیز درصد افرادی را نشان می‌دهد که واقعاً دیابت ندارند و مدل آن‌ها را به‌درستی به‌عنوان غیرمبتلا پیش‌بینی کرده است. هنگامی که مدل دارای اختصاص‌پذیری بالاست، خطای مثبت کاذب پایین می‌آید و افراد سالم به اشتباه بیمار تشخیص داده نمی‌شوند. دقت پیش‌بینی نیز به سهم خود نشان می‌دهد از میان مواردی که مدل به‌عنوان بیمار تشخیص داده، چه تعدادی واقعاً بیمار هستند. اگر این مقدار پایین باشد، مدل به‌دفعات افراد سالم را بیمار تشخیص داده است [9].

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

بنابراین، توجه به هر چهار معیار کمک می‌کند تا علاوه بر دید کلی از عملکرد مدل، جنبه‌های مهمی همچون توانایی مدل در تشخیص صحیح بیماران واقعی و اجتناب از برچسب‌زدن اشتباه به افراد سالم نیز به‌طور دقیق بررسی شود.

5- نتایج و بحث

با اعمال الگوریتم‌های مختلف یادگیری ماشین از جمله ماشین بردار پشتیبان، جنگل تصادفی و شبکه‌های عصبی مصنوعی بر روی مجموعه داده دیابت نوع ۲، مشخص شد که هر الگوریتم نقاط قوت و ضعف ویژه‌ای دارد. در اکثر سناریوها، جنگل تصادفی به دلیل ساختار جمع‌سپاری خود (تلفیق چندین درخت تصمیم) و ماشین بردار پشتیبان به واسطه تعریف دقیق مرزهای تصمیم‌گیری، در مقایسه با شبکه‌های عصبی مصنوعی نتایج بهتری از نظر دقت (Accuracy) نشان دادند. شبکه‌های عصبی مصنوعی نیز در صورت داشتن داده‌های کافی و تنظیم دقیق پارامترها (مانند تعداد لایه‌ها و نورون‌ها) می‌توانند دقتی قابل‌توجه ارائه دهند. در ادامه به مقایسه الگوریتم‌ها می‌پردازیم:

نمودار ۱ میزان اهمیت هر یک از ۸ ویژگی ورودی را در سه الگوریتم شبکه عصبی مصنوعی، ماشین بردار پشتیبان و جنگل تصادفی نمایش می‌دهد. همان‌طور که مشاهده می‌شود، ویژگی قند خون الگوریتم بیشترین تأثیر را دارد. همچنین، ویژگی‌های شاخص توده بدنی و سابقه خانوادگی دیابت نیز نقش مهمی در پیش‌بینی دیابت نوع ۲ ایفا می‌کنند.

اهداف و اولویتهای پژوهش بستگی دارد: اگر شناسایی حداکثری بیماران دیابتی مدنظر باشد، جنگل تصادفی یا شبکه‌های عصبی گزینه بهتری خواهند بود؛ اما اگر پرهیز از اشتباه تشخیص بیماران سالم اهمیت بالاتری داشته باشد، تنظیم دقیق ماشین بردار پشتیبان می‌تواند کارآمدتر باشد. در بحث نهایی، تأکید می‌شود که پیش‌پردازش داده (شامل حذف مقادیر پرت و گمشده، نرمال‌سازی ویژگی‌ها و انتخاب ویژگی‌های مؤثر) نقش مهمی در بهبود نتایج ایفا می‌کند.

جدول ۱: بررسی معیارهای ارزیابی

معیار ارزیابی	شبکه عصبی	ماشین بردار پشتیبان	جنگل تصادفی
ضریب همبستگی	۰/۴۰۴	۰/۵۱۶۶	۰/۵۲۹
خطای میانگین	۰/۳۳۲۶	۰/۳۲۷۹	۰/۳۱۴۳
درصد خطا	۷۳٪/۱۱	۷۲٪/۰۸	۶۹٪/۰۹
خطای جذر میانگین	۰/۴۷۱۵	۰/۴۱۷۷	۰/۴۰۵۳
درصد خطای نسبی جذر	۹۸٪/۸۰	۸۷٪/۵۲	۸۴٪/۹۱
زمان ساخت	۰/۷۲ ثانیه	۰/۴ ثانیه	۰/۷۷ ثانیه

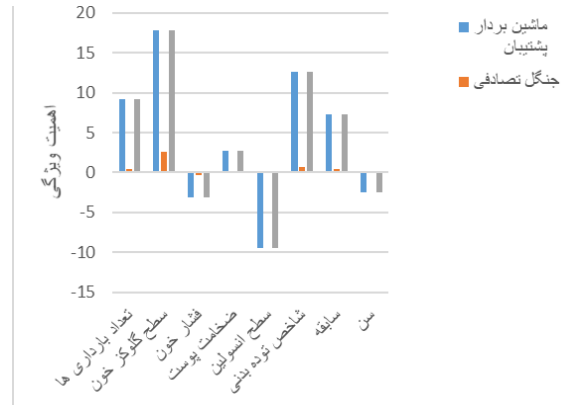
در جدول ۱، جنگل تصادفی بهترین عملکرد را با دقت بالاتر و کمترین خطای میانگین نشان داده است. ماشین بردار پشتیبان با وجود دقت کمتر، سریع‌ترین مدل بوده و برای کاربردهای نیازمند پردازش سریع مناسب است. در مقابل، شبکه عصبی مصنوعی ضعیف‌ترین عملکرد را داشته و بالاترین میزان خطا را ثبت کرده است.

۶- نتیجه‌گیری

در این پژوهش، عملکرد سه الگوریتم یادگیری ماشین شامل شبکه عصبی مصنوعی، ماشین بردار پشتیبان و جنگل تصادفی در پیش‌بینی دیابت نوع ۲ مورد ارزیابی قرار گرفت. برای این منظور، از مجموعه داده دیابت سرخپوستان پیمان‌فرمازار وکا جهت پردازش و تحلیل داده‌ها استفاده شد. در فرآیند ارزیابی، ۶۶ درصد از داده‌ها برای آموزش مدل‌ها و ۳۴ درصد باقی‌مانده برای تست و ارزیابی عملکرد آن‌ها به کار گرفته شد.

نتایج به‌دست‌آمده نشان داد که جنگل تصادفی با ضریب همبستگی ۰/۵۲۹، کمترین میزان خطا را داشته و بهترین

از طرفی، ویژگی‌های انسولین و فشار خون در برخی مدل‌ها تأثیر کمتری داشته یا حتی تأثیر منفی دارند. این مقایسه نشان می‌دهد که میزان اهمیت ویژگی‌ها بسته به نوع الگوریتم متفاوت است و باید در انتخاب مدل مناسب برای پیش‌بینی دیابت به این موضوع توجه کرد.



نمودار ۱: مقایسه ویژگی‌ها در الگوریتم

در نمودار ۲، دقت و میزان خطای سه الگوریتم بررسی شده است. همان‌طور که دیده می‌شود، جنگل تصادفی بهترین عملکرد را داشته و کمترین میزان خطا را ثبت کرده است. ماشین بردار پشتیبان عملکرد متوسطی نشان داده و شبکه عصبی مصنوعی در این مطالعه دقت پایین‌تری داشته است.



نمودار ۲: مقایسه معیارها در الگوریتم

در تحلیل معیارهای مهم از جمله حساسیت و اختصاص پذیری، مشاهده شد که مدل‌های مبتنی بر درخت (جنگل تصادفی) و شبکه‌های عصبی، در شناسایی بیماران مبتلا (حساسیت) عملکرد مطلوبی دارند؛ اما الگوریتم ماشین بردار پشتیبان، در صورت انتخاب پارامترها و کرنل مناسب، می‌تواند در تشخیص موارد سالم (اختصاص‌پذیری) بهتر عمل کند. این موضوع نشان می‌دهد که انتخاب الگوریتم به

عملکرد را در مقایسه با دو الگوریتم دیگر ارائه داده است. ماشین بردار پشتیبان نیز عملکرد قابل قبولی داشت و نسبت به شبکه عصبی مصنوعی دقت بالاتری نشان داد. در مقابل، شبکه عصبی مصنوعی به دلیل حساسیت بالا به تنظیمات اولیه و احتمال نیاز به مجموعه داده‌های بزرگ‌تر، کمترین دقت را در پیش‌بینی دیابت نوع ۲ ثبت کرد. نتایج این مطالعه نشان می‌دهد که استفاده از روش‌های یادگیری ماشین، به‌ویژه جنگل تصادفی، می‌تواند ابزار قدرتمندی در پیش‌بینی و تشخیص زودهنگام دیابت نوع ۲ باشد. با این حال، چالش‌هایی مانند پیش‌پردازش داده‌ها، انتخاب ویژگی‌های بهینه و بهبود کارایی مدل‌ها همچنان مطرح است و می‌تواند در پژوهش‌های آینده مورد بررسی قرار گیرد.

7- منابع

- [1] Saeedi, P., et al., Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas. Diabetes research and clinical practice, 2019. **157**: p. 107843.
- [2] Mackenzie, S.C., C.A. Sainsbury, and D.J. Wake, Diabetes and artificial intelligence beyond the closed loop: a review of the landscape, promise and challenges. Diabetologia, 2024. **67**(2): p. 223-235.
- [3] Contreras, I. and J. Vehi, *Artificial intelligence for diabetes management and decision support: literature review*. Journal of medical Internet research, 2018. **20**(5): p. e10775.
- [4] Cortes, C. and V. Vapnik, *Support-vector networks*. Machine learning, 1995. **20**: p. 273-297.
- [5] Burges, C.J., *A tutorial on support vector machines for pattern recognition*. Data mining and knowledge discovery, 1998. **2**(2): p. 121-167.
- [6] Chen, J.H. and S.M. Asch, *Machine learning and prediction in medicine—beyond the peak of inflated expectations*. The New England journal of medicine, 2011. **364**(26): p. 2507.
- [7] Shen, J., et al., Artificial intelligence versus clinicians in disease diagnosis: systematic review. JMIR medical informatics, 2019. **7**(3): p. e10010.
- [8] Liaw, A. and M. Wiener, *Classification and regression by randomForest*. R news : (3)2 .2002 ,p. 18-22.
- [9] Manning, C.D., P. Raghavan, and H. Schütze, *Boolean retrieval. Introduction to information retrieval*, 2008: p. 1-18