پیش‌بینی دیابت بارداری با استفاده از شبکه عصبی مصنوعی و درخت تصمیم

زهره صیادی
استاد راهنما
Seniorengineers۹۶@gmail.com

فاطمه امینی
کاردانی نرم افزار کامپیوتر
Fatemehamini۲۸۸۲@gmail.com

مهرآنا مسکوچی
کاردانی نرم افزار کامپیوتر
mehranamoschochi@gmail.com

چکیده
دیابت بارداری یا GDM یکی از مهم‌ترین چالش‌ها در زمینه بهداشت عمومی محسوب می‌شود، در این مقاله با مقایسه شبکه عصبی یا ANN و درخت تصمیم به بررسی و پیش‌بینی‌های آن پرداخته‌ایم.
با استفاده از متغیرهایی همچون سن مادر، سابقه بارداری قبلی، گروه خونی، سابقه دیابت در خانواده، وضعیت جنین و غیره نتایج پیش‌بینی باداری در دیابت را با استفاده از منابع ذکر شده مورد بررسی قرار داده‌ایم. در پیش‌بینی ابتلا و عدم ابتلا به دیابت بارداری، مدل شبکه عصبی دارای نرخ دسته‌بندی اشتباه کمتر نسبت به مدل‌های دیگر بود. می‌توان نتیجه‌گیری کرد که این مدل پیش‌بینی‌های صحیح‌تر و نزدیک‌تری به واقعیت نسبت به سایر مدل‌ها دارد.

واژگان کلیدی: دیابت بارداری، GDM، پرسپترون، شبکه عصبی، درخت تصمیم

مقدمه

دیابت یکی از مهم‌ترین چالش‌های بهداشت عمومی در جهان است و همچنین به عنوان یکی از شایعترین عارضه‌های پزشکی در بارداری شناخته می‌شود [۱]. دیابت بارداری (GDM) به عنوان یک اختلال متابولیسم مزمن موثر بر کربوهیدرات‌ها، پروتئین و چربی‌ها تعریف می‌شود و به دلیل کمبود یا مقاومت به انسولین بروز می‌کند. هنگامی که این اختلال در دوران بارداری شروع شود یا تشخیص داده شود، به عنوان دیابت بارداری (GDM)نامیده می‌شود [۲]. دیابت بارداری می‌تواند منجر به عوارض متعددی در طول بارداری و بعد از آن برای مادر و نوزاد شود. از جمله این عوارض برای مادر می‌توان به افزایش احتمال زایمان سزارین، پره اکلامپسی، پلی هیدرآمنیوس، عفونت باکتریایی و قارچی، صدمات زایمانی ناشی از ماکروزومی جنین، نفروپاتی، رتینوپاتی و یکی از عوامل خطر مهم برای ابتلا به دیابت نوع ۲ است[۳]. از جمله عوارض جنینی این اختلال می‌توان به هیپوگلسیمی، هیپوکلسمی، هیپریبلی روبینمی و افزایش مرگ و میر در پری ناتال اشاره کرد. بر اساس تحقیقات و مطالعات انجام شده اگر مادر در کنترل قند خود موفق شود، عوارض فوق به میزان زیادی کاهش می‌یابد. عوارضی مثل هیپوگلسیمی، مرگ داخل رحمی و هیپربیلی روبینمی قابل کنترل می‌باشد و عوارضی مانند دیستوشی شانه، ماکروزومی و زایمان سزارین ۵۰ درصد کاهش پیدا می‌کند [۴].

مطالعات نشان‌دهنده افزایش شیوع GDM در سال‌های اخیر هستند که به دلیل افزایش شیوع و عوارض متعدد آن برای کودک و مادر، بار زیادی را به نظام سلامت تحمیل می‌کند [۱]. اختلال در تست تحمل گلوگز در ۳ تا ۵ درصد حاملگی‌ها رخ می‌دهد و ۹۰ درصد از تمام بارداری‌های دارای عارضه با دیابت، در اثر دیابت بارداری می‌باشد. شیوع GDM به عوامل مختلفی مانند شیوع چاقی، ضعف در دسترسی به غربالگری، سبک زندگی و فقر غذایی بستگی دارد [۵,۶] که متاسفانه آمار افراد بارداری که دچار GDM می‌شوند در کشورهای فقیر و در حال توسعه در مقایسه با کشورهای توسعه یافته بیشتر است [۷]. فراوانی GDM در نقاط مختلف جهان بین ۱ تا ۱۴ درصد گزارش شده است. در کشورهای آسیایی به صورت متوسط ۱۰/۹ درصد، در اروپا ۵/۲ درصد، در آمریکا ۱ تا ۳ درصد [۸] و در ایران در مجموع ۹/۴ درصد ارزیابی شده‌است [۹].

آزمایش تحمل گلوکز خوراکی ۷۵ گرمی به عنوان آزمون استاندارد طلایی غربالگری GDM در بازه زمانی بین هفته‌های ۲۸ تا ۲۴ بارداری در بیشتر کشورها مورد استفاده قرار می‌گیرد [۱۰]. برخی مطالعات نشان می‌دهند آزمون گلوکز خوراکی ۱۰۰ گرم برای غربالگری GDM نشان می‌دهد قند خون دو ساعته با نقطه برش مساوی یا بزرگتر ۱۵۳ dl/mg بهترین حساسیت و ویژگی را در تشخیص GDM دارد و این روش پیشنهاد شده است[۱۱]. کار گروه مطالعات بارداری انجمن بین المللی دیابت و سازمان جهانی بهداشت در سال ۲۰۱۳، داشتن قند خون ناشتا FBS بیشتر یا مساوی ۹۲ را در هفته ۲۶ تا ۳۰ بارداری را به عنوان یکی از معیارهای تشخیص دیابت معرفی کرد [۱۲,۱۳]. با توجه به افزایش شیوع این اختلال و عوارضی که مادر و جنین متوجه آن هستند تشخیص به موقع این اختلال دارای اهمیت فراوانی است زیرا همان‌طور که در فوق ذکر شد در صورت تشخیص زودرس بسیاری از این عوامل خطر را می‌توان کنترل نمود. پس در این صورت استفاده از روش‌هایی که بتواند ابتلای فرد را در ماه‌های اولیه بارداری به دیابت بارداری را در هفته‌های پایانی بارداری پیش‌بینی کند، بسیار کاربردی و مفید است. مطالعات اخیر راجع به تشخیص زودرس دیابت بارداری بر اساس یافته‌های بالینی مادر، ویژگی‌های مادر، شاخص‌های زیستی در هفته‌های اول بارداری و طراحی مدل‌های پیش‌بینی با استفاده از چند عامل خطر متمرکز شده‌اند [۱۴].

در سال‌های اخیر علاوه بر پیشرفت روزافزون فناوری اطلاعات در بخش پزشکی، یکی دیگر از مواردی که شایان توجه است کاربرد هوش مصنوعی AI در پزشکی است. یکی از کاربرد الگوریتم‌های هوش مصنوعی مانند یادگیری ماشین، تشخیص روابط پنهان و غیر خطی داده‌ها است. الگوریتم‌های هوش مصنوعی مانند شبکه عصبی ANN به عنوان یکی از زیر مجموعه‌های AI امکان خوشه بندی،

انتخاب ویژگی، طبقه بندی و پیش‌بینی را بر اساس تحلیل داده‌ها و تبدیل آن‌ها به اطلاعات فراهم می‌آورد. شبکه‌های عصبی مصنوعی، ماشین‌های بردار پشتیبانی، K نزدیک ترین همسایه، یادگیری عمیق، درخت تصمیم، جنگل تصادفی و الگوریتم‌های تکاملی مهم ترین رویکردهای هوش مصنوعی هستند که برای تشخیص روابط پنهان داده‌ها مورد استفاده قرار می‌گیرند. درخت تصمیم به دلیل فهم آسان‌تر سادگی بیشتر و شبکه عصبی مصنوعی به دلیل دقت بالاتر نسبت به سایر الگوریتم‌های دیگر مورد استفاده قرار می‌گیرند. الگوریتم‌های یادگیری ماشین تاکنون در زمینه‌های مختلف پژوهش‌های سلامت، مثل تشخیص انواع سرطان، دیابت و بیماری‌های قلبی استفاده شده و نتیجه‌ی مطلوب و موفقیت آمیزی داشته است[۱۵,۱۶].

از میان مدل‌های هوش مصنوعی که برای رده‌بندی داده‌ها مورد استفاده قرار می‌گیرند، می‌توان به درخت تصمیم و شبکه عصبی مصنوعی اشاره کرد. این مدل‌ها در مطالعات و پژوهش‌های بسیاری برای تشخیص ابتلا و عدم ابتلا به بیماری استفاده شده‌اند. در تحقیقاتی که به بررسی عملکرد مدل‌های شبکه عصبی مصنوعی و درخت تصمیم در پیش‌بینی بیماری قلبی پرداخته‌اند، حساسیت، ویژگی و صحت پیش‌بینی مدل شبکه عصبی مصنوعی به ترتیب ۸۱/۱، ۷۸/۷، ۸۰/۲ درصد و برای درخت تصمیم به ترتیب ۸۱/۷، ۰/۷۶، ۷۹/۳ به دست آمد. در این مطالعه شبکه عصبی مصنوعی صحت بیشتر، دقت بالاتر و نرخ خطای کمتری نسبت به مدل درخت تصمیم در پیش‌بینی بیماری‌های قلبی دارد [۱۷]. یکی دیگر از کاربردهای مهم هوش مصنوعی کمک به تشخیص زودهنگام دیابت بارداری (GDM) در طی شیوع بیماری همه‌گیر COVID-۱۹ در سال ۲۰۱۹ بوده است. در آن زمان، خطر بیماری برای زنان باردار از یک سو و اجرای محدودیت‌های گسترده و ممنوعیت‌های سختگیرانه از سوی دیگر، اختلالاتی در فرآیند دریافت خدمات بهداشتی معمول در دوران بارداری ایجاد کرد. مطالعات انجام شده در ایران نیز نشان‌دهنده تأثیرات منفی شیوع COVID-۱۹ بر دریافت خدمات بهداشتی معمول در دوران بارداری نشان می‌دهند [۱۸,۱۹].

بنابراین، با توجه به اهمیت تشخیص زودرس GDM و قابلیت‌ها و توانایی‌های الگوریتم‌های هوش مصنوعی اشاره شد؛ هدف از مطالعه مقایسه عملکرد دو الگوریتم شبکه عصبی مصنوعی و درخت تصمیم جهت پیش‌بینی و تشخیص احتمال ابتلا به دیابت بارداری است.

مواد و روش

در مدل پیش‌بینی شبکه عصبی مصنوعی و درخت تصمیم، ۴۲۰ زن باردار مورد بررسی قرار گرفتند. از این تعداد، ۱۵۳ نفر (۲۶/۵٪) اولین بارداری خود را تجربه می‌کردند و ۱۳ نفر (۳/۱٪) دو یا چند قلو باردار بودند. در این نمونه، ۲۸۷ نفر (۶۸/۲٪) غیر مبتلا و ۱۳۴ نفر (۳۱/۸٪) مبتلا به دیابت بارداری بودند. آزمون t متغیرهایی که در دو گروه (غیر مبتلا و مبتلا به GDM) تفاوت معنی‌داری داشتند مشخص کرد.
(جدول شماره ۱)

در این تحقیق، داده‌های به دست آمده از دو مطالعه کوهورت در زمینه دیابت بارداری که بین سال‌های ۱۳۹۳ تا ۱۳۹۷ در مراکز بهداشتی درمانی شهر اهواز انجام شده‌اند، مورد تجزیه و تحلیل قرار گرفته‌اند.

طرح اول: این طرح با هدف تعیین میزان بروز دیابت بارداری بر اساس معیارهای کارگروه بین‌المللی انجمن مطالعه دیابت و بارداری (IADPSG) طراحی شده بود. در این طرح، ۷۵۰ خانم باردار مورد بررسی قرار گرفته بودند.

طرح دوم: در این مطالعه، پیامدهای متابولیک دیابت بارداری در ۲۶۰ مورد با گروه سالم (۲۶۰ مورد) مورد مقایسه قرار گرفته بودند.

در مجموع، داده‌های ۱۲۷۰ خانم باردار مورد استفاده قرار گرفت که ۸۱۶ مورد آن‌ها سالم و ۴۵۴ مورد مبتلا به دیابت بارداری بودند. دیابت بارداری بر اساس نتایج

غربالگری دیابت بارداری در هفته ۲۴ تا ۲۸ بارداری با استفاده از تست تحمل گلوکز خوراکی ۷۵ گرمی و معیارهای IADPSG تعریف شده بود. بر اساس این معیارها، وجود حداقل یک قند غیر طبیعی به عنوان دیابت بارداری تعریف می‌شود:

-گلوکز ناشتای پلاسما:

$$\geq ۹۲\ mg/dL$$

-گلوکز یک ساعته:

$$\geq ۱۸۰\ mg/dL$$

-گلوکز دو ساعته:

$$\geq ۱۵۳\ mg/dL$$

متغیرهای مورد بررسی در این دو مطالعه شامل موارد زیر بوده است:

-قند خون ناشتای اولین معاینه بارداری
-سن مادر
-وضعیت شغلی مادر
-قومیت مادر
-سطح سواد مادر
-تعداد بارداری‌ها
-سن ازدواج
-سن اولین بارداری
-سابقه دیابت بارداری در بارداری‌های قبلی
-سابقه سقط جنین
-مقدار هموگلوبین در اولین معاینه بارداری
-گروه خونی مادر
-وزن مادر در ماه اول بارداری
-وزن مادر در ماه نهم بارداری
-نمایه توده بدنی بر اساس وزن اولین معاینه بارداری
-فشارخون سیستولی و دیاستولی
-سابقه دیابت در خانواده

متغیرها	افراد مبتلا	افراد غیر مبتلا	آماره آزمون	P-value
هماتوکریت هفته ۶ تا ۱۰ بارداری	۴۰/۳۶ ± ۲۷/۴	۳۷/۲۸ ± ۳/۱	-۲/۰۳	۰/۰۴۳
قند خون ناشتا هفته ۶ تا ۱۰ بارداری	۸۸/۶۱ ± ۹/۴	۷۶/۵۸ ± ۸/۶	-۱۳/۹۹	< ۰/۰۰۱
فشارخون سیستول ماه اول بارداری	۱۰۳/۴۶ ± ۹/۶	۱۰۱/۱ ± ۱۲/۹	-۲/۰۴	۰/۰۰۶
فشارخون دیاستول ماه اول بارداری	۶۴/۹۹ ± ۷/۵	۶۲/۹۹ ± ۷/۵	-۲/۷۵	۰/۰۰۶
شاخص توده بدنی اولین مراجعه	۲۶/۶۶ ± ۴/۹	۲۳ /۶۳ ± ۳/۲	-۸/۱۱	< ۰/۰۰۱

جدول شماره ۱- مقایسه میانگین () متغیرهای کمی پیشگو در دو گروه مبتلا و غیرمبتلا

روش تحقیق

از میان مدل‌های هوش مصنوعی که برای رده‌بندی داده‌ها مورد استفاده قرار می‌گیرند، می‌توان به درخت تصمیم و شبکه عصبی مصنوعی اشاره کرد. این مدل‌ها در مطالعات و پژوهش‌های بسیاری برای تشخیص ابتلا و عدم ابتلا به بیماری استفاده شده‌اند. در تحقیقاتی که به بررسی عملکرد مدل‌های شبکه عصبی مصنوعی و درخت تصمیم در پیش‌بینی بیماری قلبی پرداخته‌اند، حساسیت، ویژگی و صحت پیش‌بینی مدل شبکه عصبی مصنوعی به ترتیب ۸۱/۱، ۷۸/۷، ۸۰/۲ درصد و برای درخت تصمیم به ترتیب ۸۱/۷، ۰/۷۶، ۷۹/۷۳ به دست آمد. در این مطالعه شبکه عصبی مصنوعی صحت بیشتر، دقت بالاتر و نرخ خطای کمتری نسبت به مدل درخت تصمیم در پیش‌بینی بیماری‌های قلبی دارد. یکی دیگر از کاربردهای مهم هوش مصنوعی کمک به تشخیص زودهنگام دیابت بارداری (GDM) در طی شیوع بیماری همه‌گیر COVID-۱۹ در سال ۲۰۱۹ بوده است. در آن زمان، خطر بیماری برای زنان باردار از یک سو و اجرای محدودیت‌های گسترده و ممنوعیت‌های سختگیرانه از

سوی دیگر، اختلالاتی در فرآیند دریافت خدمات بهداشتی معمول در دوران بارداری ایجاد کرد. مطالعات انجام شده در ایران نیز نشان‌دهنده تأثیرات منفی شیوع COVID-۱۹ بر دریافت خدمات بهداشتی معمول در دوران بارداری نشان می‌دهند.

شبکه عصبی مصنوعی (ANN)

شبکه عصبی مصنوعی، سیستمی پردازشی دینامیک است که برگرفته از سیستم‌های عصبی زیستی طراحی شده است. این سیستم با تحلیل داده‌های تجربی، دانش یا قوانین پنهان در داده‌ها را استخراج کرده و به ساختار شبکه انتقال می‌دهد. واحد اصلی این ساختار، پردازنده‌های اطلاعاتی هستند که با انجام محاسبات بر روی داده‌های اولیه و خام، قوانین کلی را فرا می‌گیرند.

بهترین مدل ANN

بهترین مدل ANN با یک لایه مخفی و ۶ نورون در لایه مخفی ساخته شد. تابع فعال‌سازی در لایه مخفی، تانژانت هایپربولیک و در لایه خروجی، تابع softmax بود. همچنین مقدار SSE (به مجموع مربعات خطاها در یک مدل آماری گفته می‌شود. این معیار برای اندازه‌گیری فاصله بین داده‌های واقعی و پیش‌بینی شده توسط یک مدل استفاده می‌شود) مجموعه آزمایشی از مجموعه آموزشی کمتر بوده (۶۲,۲۳ در مقابل ۹۱,۰۳) و خطای نسبی در دو مجموعه به هم نزدیک بود (۰,۷ در مقابل ۰,۶) که هر دو مورد برازش مناسب این مدل را نشان می‌دهند. آنالیز حساسیت متغیرهای FBS (قند خون

پیش‌بینی دیابت بارداری با استفاده ازANN

برای پیش‌بینی دیابت بارداری، یک شبکه عصبی مصنوعی (ANN) با استفاده از نرم‌افزار Matlab پیاده‌سازی شد. Matlab به دلیل قابلیت‌های گسترده در محاسبات ریاضی و برنامه‌نویسی علمی، انتخاب مناسبی برای این کار بود. در این مدل، چهار پارامتر کلیدی از پرونده‌های بیماران به عنوان ورودی‌های شبکه در نظر گرفته شدند. نورون‌های لایه ورودی، مشخصات بالینی بیماران را نمایندگی می‌کنند که شامل اطلاعاتی مانند

بنابراین، با توجه به اهمیت تشخیص زودرس GDM و قابلیت‌ها و توانایی‌های الگوریتم‌های هوش مصنوعی اشاره شد؛ هدف از مطالعه مقایسه عملکرد دو الگوریتم شبکه عصبی مصنوعی و درخت تصمیم جهت پیش‌بینی و تشخیص احتمال ابتلا به دیابت بارداری است.

ANNها به عنوان سیستم‌های هوشمندی شناخته می‌شوند که قابلیت پردازش اطلاعات مشابه مغز انسان را دارند. این شبکه‌ها در طبقه‌بندی، خوشه‌بندی، تشخیص و پیش‌بینی در داده‌های بزرگ (Big data) که اغلب دارای روابط غیرخطی هستند، کاربرد دارند. شبکه‌های عصبی از تعداد زیادی نورون‌های (عناصر پردازشی) به هم پیوسته تشکیل شده‌اند که به صورت موازی کار می‌کنند [۲۰].

ناشتا)، BMI (شاخص توده بدنی) Hc (اندازه دور سر جنین) و Hb (هموگلوبین) سه ماهه اول بارداری را مهم‌ترین عوامل خطر دیابت بارداری در مدل شبکه عصبی نشان داد.

فرمول تابع تانژانت هایپربولیک:

$$\boldsymbol{tanh(x) = \frac{e^{2x} - 1}{e^{2x} + 1}}$$

فرمول softmax:

$$\boldsymbol{\sigma\left(\underset{z}{\rightarrow}\right)_i = \frac{e_i^z}{\sum_{j=1}^{K} e_i^z}}$$

سطح قند خون، BMI، و دیگر شاخص‌های مهم هستند. پس از آموزش دقیق و ارزیابی عملکرد مدل، نتایج حاصل از ANN با داده‌های واقعی مقایسه شدند. این مقایسه نشان داد که مدل ANN توانایی بالایی در پیش‌بینی وضعیت دیابت بارداری دارد. **جدول شماره ۲** نتایج پس از مقایسه عملکرد مدل شبکه عصبی مصنوعی را نشان می‌دهد که بر اساس آن، می‌توان به اثربخشی مدل ANN در پیش‌بینی دیابت بارداری پی برد.

مدل مورد استفاده	صحت	حساسیت	ویژگی	سطح زیر منحنی (ROC)
شبکه عصبی مصنوعی	٪۸۳	٪۵۸	٪۹۴	٪۸۷

جدول شماره ۲

شبکه پرسپترون

پرسپترون، نوعی از شبکه‌های عصبی مصنوعی است که توسط روزنبلات در سال ۱۹۵۸ معرفی شد. این شبکه بر اساس واحد محاسباتی پرسپترون ساخته شده و معماری آن شامل سه لایه ورودی، میانی (پنهان) و خروجی است. پرسپترون با دریافت بردار ورودی‌ها و انجام ترکیب خطی آن‌ها، در صورتی که نتیجه کمتر از یک آستانه مشخص باشد، خروجی آن (۱-) و در غیر این صورت (۱) خواهد بود. یادگیری در این شبکه‌ها از طریق الگوریتم‌های یادگیری خاصی صورت می‌گیرد این الگوریتم ها با تنظیم وزن موجود در ارتباطات بین نورون ها، شبکه را آموزش می‌دهند. نورون ها در شبکه عصبی مصنوعی به هم متصل هستند و هر اتصال آن‌ها در شبکه دارای وزنی است که میزان تأثیر نورون بر دیگری را نشان می‌دهد. مدل پرسپترون چند لایه کاربرد (MLP) در حل مسائلی مانند تخمین تابع و شناسایی الگو بهتر عمل می‌کند و به همین دلیل در این مطالعه از این مدل استفاده شده است [۲۰,۲۱].

فرمول پرسپترون:

$$\mathbf{f(x)} = \begin{cases} \mathbf{1}, & \mathbf{if\ w.x + b > 0} \\ \mathbf{0}, & \mathbf{otherwise} \end{cases}$$

آموزش شبکه

در فرآیند آموزش، پس از معرفی داده‌های ورودی مقادیر خروجی به ازای ورودی ها به سیستم معرفی شده، و آموزش شبکه شروع می‌شود. با استفاده از داده‌های (ورودی و خروجی) واقعی، می‌توان از ANN برای گرفتن حاصل ورودی های جدید بهره برد. درANN ها، هر واحد خروجی تابعی خطی از ورودی‌های آن است و پارامترهای مربوط به توابع خطی برای کاهش خطا در آموزش قابل تنظیم هستند. شبکه را می‌توان با مجموعه‌ای از نمونه‌ها آموزش داد تا به بهترین مقادیر پارامترها دست یافت و به کمترین میزان خطا رسید [۲۲].

درخت تصمیم

درخت تصمیم (Decision Tree) یک الگوریتم یادگیری ماشین است که برای حل مسائل دسته‌بندی (Classification) و رگرسیون (Regression) استفاده می‌شود. این الگوریتم با ساختار درختی وارونه‌ای که شبیه به فلوچارت است، عمل می‌کند و می‌تواند تفکر انسان را در سطوح مختلف تقلید کند [۲۳].

اجزای اصلی درخت تصمیم:

-*گره ریشه (Root Node)*: نقطه شروع درخت که از آن تقسیم‌بندی داده‌ها آغاز می‌شود.

-*گره‌های تصمیم (Decision Nodes)*: نقاطی که تقسیم‌بندی داده‌ها بر اساس ویژگی‌های مختلف صورت می‌گیرد.

-*گره‌های برگ (Leaf Nodes)*: نقاط پایانی که نتیجه یا پیش‌بینی نهایی را نشان می‌دهند [۲۳].

چگونگی کارکرد:

درخت تصمیم با استفاده از داده‌های آموزشی، قوانینی را برای تقسیم‌بندی داده‌ها بر اساس ویژگی‌های مختلف یاد می‌گیرد. هر گره تصمیم، یک سوال بله/خیر را بر اساس یک ویژگی مطرح می‌کند و بر اساس پاسخ، به یکی از شاخه‌های زیرین هدایت می‌شود. این فرآیند تا رسیدن به گره برگ و تعیین نتیجه ادامه می‌یابد.

مزایا:

-سادگی و قابل فهم بودن

-توانایی کار با داده‌های هم نوعی و غیر هم نوعی

-نیازی به نرمال‌سازی داده‌ها ندارد.

معایب:

-ممکن است به راحتی بیش‌برازش (Overfitting) ایجاد کند.

-تغییرات کوچک در داده‌ها می‌تواند تغییرات بزرگی در ساختار درخت ایجاد کند [۲۴].

درخت تصمیم به دلیل سادگی و قابل فهم بودن، یکی از الگوریتم‌های محبوب در حوزه یادگیری ماشین است. این روش با استفاده از تکنیک‌های ساده، الگویی برای رده‌بندی داده‌ها ارائه می‌دهد و علی‌رغم سادگی، در پیش‌بینی و تشخیص، به اندازه روش‌های پیچیده‌تر کارآمد است. در این پژوهش، از الگوریتم C۴.۵ برای ساخت درخت تصمیم استفاده شده است که یکی از روش‌های رایج در این زمینه است، در کنار سایر روش‌هایی مانند CHAID، C۵ و C&R Tree.

یافته‌ها

داده‌های دو مطالعه قبلی ادغام شدند و برای کیفیت، داده‌های پرت و ناموجود شناسایی شدند. ***۱۸ متغیر*** اصلی برای مدل‌های پیش‌بینی دیابت بارداری انتخاب شدند، که شامل سن مادر، قند خون، تحصیلات، تعداد بارداری، وزن، هموگلوبین، BMI، فشارخون، سابقه دیابت در خانواده، و قومیت بود. این متغیرها با استفاده از آزمون‌های آماری تی مستقل و مجذور کای برای اهمیت آماری در دو گروه مورد بررسی قرار گرفتند. در مرحله سوم، مدل‌های پیش‌بینی دیابت بارداری با استفاده از الگوریتم ***درخت تصمیم*** در محیط ***MATLAB*** ایجاد شدند. قبل از مدل‌سازی، ***تحلیل مولفه‌های اساسی (PCA)*** در MATLAB انجام شد تا تعداد متغیرهای معنی‌دار کاهش یابد و پیچیدگی مدل‌ها کمتر شود، که این امر به نوبه خود می‌توانست دقت مدل‌ها را افزایش دهد. با استفاده ازPCA ، وزن هر یک از عوامل خطرساز دیابت بارداری تعیین و متغیرهای کم اهمیت حذف شدند. در نهایت، متغیرهایی مانند ***سن مادر، سابقه دیابت بارداری، سابقه سقط جنین، هموگلوبین اولین معاینه بارداری، نمایه توده بدنی، فشارخون سیستولی، سابقه دیابت در خانواده، گروه خونی، و تحصیلات مادر*** برای مدل‌سازی انتخاب شدند.

در مرحله بعد، عملکرد مدل‌های پیش‌بینی با استفاده از معیارهای ***دقت، حساسیت، ویژگی، و صحت*** مورد بررسی قرار گرفت. ***دقت*** مدل به درصد پیش‌بینی‌های صحیح توسط الگوریتم اشاره دارد، ***حساسیت*** به احتمال تشخیص صحیح بیماری (GDM)، ***ویژگی*** به احتمال تشخیص صحیح نمونه‌های سالم، و ***صحت*** به تعداد موارد صحیح ابتلا به بیماری اشاره دارد.

برای ارزیابی دقیق‌تر، داده‌ها به دو زیرمجموعه آموزش و تست تقسیم شدند. مدل‌ها با داده‌های آموزشی ساخته شده و دقت آن‌ها با داده‌های تست مقایسه شد. برای کاهش خطای تورش، از روش ***K-Fold Cross Validation*** استفاده شد که در آن داده‌ها به ده گروه تقسیم و مدل‌ها با (K-۱) گروه آموزش داده و با یک گروه تست می‌شوند. این فرایند (K) بار تکرار می‌شود و دقت کلی مدل از معادله زیر محاسبه می‌شود:

$$\mathrm{CVA} = \frac{۱}{K}\sum_{i}^{K} = ۱^{Ai}$$

که در آن ***CVA*** میانگین دقت مدل‌های پیش‌بینی و ***A*** معیارهای ارزیابی مانند حساسیت، ویژگی، دقت و F measure را نشان می‌دهد. در این مطالعه، (K) برابر با ۱۰ در نظر گرفته شده است.

در نهایت در نتیجه مطالعه برای پیش‌بینی ابتلا به دیابت بارداری(GDM) ، از مجموع ۱۳ متغیر معنی‌دار، با استفاده از نرم‌افزار MATLAB و روشPCA ، تعدادی از متغیرها حذف شدند. در نهایت، نه متغیر با وزن بالاتر برای طراحی مدل‌های پیش‌بینی اولیه انتخاب شدند که شامل: سن مادر، سطح تحصیلات مادر، سابقه دیابت بارداری در بارداری‌های قبلی، سابقه دیابت در خانواده، BMI، هموگلوبین و قند خون ناشتا در اولین معاینه بارداری، فشارخون سیستولی و سابقه سقط جنین بودند. مدل‌سازی با استفاده از الگوریتم درخت تصمیم C۴.۵ و تابع ناخالصی انجام شد که در آن، متغیر قند ناشتا در گره ریشه و متغیر BMI در گره بعدی قرار گرفتند. مدل درخت تصمیم با عمق ۹ گره ساخته شد.

بر اساس نتایج، ***مهم‌ترین عوامل خطر*** برای ابتلا بهGDM عبارت بودند از: قند خون ناشتا، BMI، سابقهGDM و سابقه خانوادگی دیابت. همچنین، آنالیز حساسیت و اهمیت نسبی متغیرها در مدل ANN نشان داد که این متغیرها مهم‌ترین عوامل در پیش‌بینی و شناسایی خطر ابتلا به دیابت بارداری هستند. علاوه بر این، انجام رگرسیون لجستیک و محاسبه OR نشان داد که قند خون ناشتا، سابقه خانوادگی دیابت و سابقه دیابت بارداری در بارداری‌های قبلی، بیشترین تأثیر را در احتمال ابتلا به دیابت بارداری دارند. این یافته‌ها می‌توانند در طراحی استراتژی‌های پیشگیری و درمانی مؤثر برایGDM به کار روند.

در جدول مقایسه حساسیت ویژگی صحت و دقت دو مدل درخت تصمیم و ANN را نشان می‌دهد.

بر اساس نتایج به دست آمده از مدل‌های ***درخت تصمیم و ANN***، چهار متغیر شامل فشارخون سیستولی، میزان هموگلوبین در اولین معاینه بارداری، سطح تحصیلات مادر، و سابقه سقط جنین، که تأثیر کمتری در پیش‌بینی دیابت بارداری داشتند، از مطالعه کنار گذاشته شدند. مدل‌سازی جدید با پنج متغیر باقی‌مانده انجام شد.

(جدول شماره۳)

مدل	حساسیت (Sensitivity)	ویژگی (Specificity)	صحت (Precision)	دقت (Accuracy)
درخت تصمیم	٪۸۴	٪۸۳	٪۸۷/۳	٪۸۹/۳

جدول شماره ۳- مقایسه نشان‌دهنده حساسیت، صحت، و دقت مدل درخت تصمیم است.

بحث

تشخیص به موقع دیابت حاملگی برای کنترل و درمان آن حیاتی است و می‌تواند به کاهش عوارض کمک کند. ترکیب مدل‌های پیش‌بینی با سیستم‌های کمکی تصمیم‌گیری می‌تواند در مراقبت‌های بهداشتی مفید باشد. در این مطالعه، دو مدل پیش‌بینی، ANN و درخت تصمیم، مورد بررسی قرار گرفتند و نشان دادند که هر دو مدل قابلیت پیش‌بینی مناسبی دارند. با اینکه مطالعات کمتری نسبت به دیابت نوع دوم انجام شده، اما این مدل‌ها در بین محققان محبوبیت دارند و استفاده از چندین مدل مختلف می‌تواند به انتخاب بهترین مدل برای پیش‌بینی بیماری کمک کند [۲۵]. در مدل پیش‌بینی درخت تصمیم، فرآیند انتخاب متغیرهای مهم برای پیش‌بینی دیابت حاملگی (GDM) از طریق چهار مرحله دقیق و مبتنی بر داده‌ها صورت گرفت. این مراحل شامل مشورت با خبرگان، انجام آزمون‌های آماری، استفاده از روش PCA برای کاهش ابعاد داده‌ها و محاسبه نسبت شانس (OR) بود. در ابتدا، ۹ متغیر شامل سن مادر، سطح تحصیلات، سابقه دیابت حاملگی، قند خون ناشتا، سابقه دیابت در خانواده، BMI، هموگلوبین و سابقه سقط جنین مورد بررسی قرار گرفتند. پس از آن، پنج متغیر با اهمیت بیشتر برای مدل نهایی انتخاب شدند که شامل

قند خون ناشتا، BMI، سابقهGDM، سابقه دیابت در خانواده و سن مادر است.
مطالعات متعددی که از الگوریتم‌های یادگیری ماشین و مدل‌های هوش مصنوعی برای پیش‌بینی GDM استفاده کرده‌اند، نشان داده‌اند که این روش‌ها می‌توانند به طور مؤثری در پیش‌بینی این بیماری کمک کنند. به ویژه، قند خون ناشتا به عنوان یکی از مهم‌ترین متغیرها در پیش‌بینی GDM شناخته شده است. کاهش تعداد متغیرها و تمرکز بر متغیرهای کلیدی می‌تواند به بهبود عملکرد مدل‌های پیش‌بینی کمک کند و در نتیجه، مراقبت‌های بهداشتی مؤثرتری را برای زنان باردار فراهم آورد. این تحقیقات می‌توانند به توسعه روش‌های جدید و بهینه‌تر برای تشخیص و مدیریت GDM کمک کنند.
در تحقیقات مختلف، نتایج گوناگونی در مورد مدل‌های پیش‌بینی بیماری‌ها به دست آمده است. مطالعات نشان داده‌اند که میزان دقت و حساسیت این مدل‌ها متفاوت است و به عواملی مانند حجم نمونه، نوع و تعداد متغیرهای مورد استفاده، و دقت داده‌ها بستگی دارد. به طور خاص، در مطالعات انجام شده در ایران، مدل‌های مبتنی بر شبکه‌های عصبی مصنوعی (ANN) و درخت تصمیم‌گیری مورد بررسی قرار گرفته‌اند. در یک مطالعه، مدل ANN عملکرد بهتری نسبت به درخت تصمیم‌گیری داشت، در حالی که در مطالعه دیگری، درخت تصمیم‌گیری دقت بالاتری نشان داد. این تفاوت‌ها نشان‌دهنده اهمیت انتخاب الگوریتم مناسب و تنظیمات دقیق برای هر مجموعه داده خاص هستند[۲۵,۲۶].
علاوه بر این، طراحی مدل‌های پیش‌بینی کننده با حساسیت و دقت بالا چالش‌برانگیز است، به ویژه زمانی که با داده‌های محدود و نیاز به آزمایش‌های خاص مواجه هستیم. عوامل خطر دیابت بارداری نیز در حال افزایش هستند و به چالش‌های موجود در طراحی مدل‌های پایدار می‌افزایند [۲۷].
اما در این تحقیق مدل شبکه عصبی مصنوعی یا ANN دارای بالاترین صحت بود به این معنا که نسبت به مدل‌های دیگر میزان دسته‌بندی درست بیشتری داشت. این مدل دارای توان بیشتری در تشخیص درست افراد غیر مبتلا دارد.
همچنین حساسیت مدل‌های شبکه عصبی مصنوعی و تحلیل ممیزی آن به این معناست که این مدل دارای توان بیشتری در تشخیص درست افرادی است که به دیابت بارداری مبتلا هستند.
یکی از مزایای مدل پیش‌بینی درخت تصمیم، استفاده از داده‌های استاندارد و متغیرهای قابل اندازه‌گیری در ابتدای بارداری است که هزینه‌های اضافی را بر بیمار و سیستم بهداشتی تحمیل نمی‌کنند. با این حال، یکی از محدودیت‌ها، استفاده از تعداد محدودی از الگوریتم‌ها برای مدل‌سازی است که می‌تواند بر نتایج تأثیر بگذارد [۲۸].
و یکی از محدودیت‌های اصلی مدل پیش‌بینیANN، ناقص بودن اطلاعات و وجود داده‌های گمشده در پرونده‌های موجود بوده است. این مسئله باعث شده است که بسیاری از پرونده‌ها به دلیل نبود داده‌های کامل از مطالعه حذف شوند، که در نتیجه به کاهش حجم نمونه منجر شده است [۲۹].

نتیجه‌گیری

مطالعات اخیر نشان می‌دهند که الگوریتم‌های هوش مصنوعی، به‌ویژه شبکه‌های عصبی مصنوعی(ANN)، توانایی بالایی در پیش‌بینی زودهنگام دیابت بارداری دارند و می‌توانند به بهبود مدیریت این بیماری کمک کنند. با اینکه استفاده از این مدل‌ها در تشخیص‌های بالینی هنوز محدود است، تجربیات مثبت برخی مراکز درمانی در سراسر جهان نشان می‌دهد که ادغام الگوریتم‌های پیشرفته هوش مصنوعی در سیستم‌های پشتیبانی تصمیم‌گیری پزشکی می‌تواند مفید باشد. همچنین، پیشنهاد می‌شود که در راستای توسعه پرونده الکترونیک سلامت، از این الگوریتم‌ها استفاده شود تا دقت تشخیص و پیش‌بینی بیماری‌ها افزایش یابد.
در مقایسه با درخت تصمیم، مدل ANN در پیش‌بینی دیابت بارداری دارای نرخ خطای کمتر و سطح زیر منحنی ROC بالاتری بوده است، که نشان‌دهنده دقت بیشتر این مدل در پیش‌بینی‌های صحیح است. با توجه به

کمبود مطالعات گذشته که مدل‌های آماری را در زمینه دیابت بارداری مقایسه کرده‌اند؛ البته، با بررسی و آزمایش مدل‌های پیش‌بینی دیگر، می‌توان درستی و کارآیی آن‌ها را در پیش‌بینی دیابت بارداری ارزیابی کرد. این کار می‌تواند به شناسایی روش‌های موثرتر برای تشخیص زودهنگام و مدیریت بهینه این بیماری کمک کند و در نهایت به ارتقاء سلامت مادران و نوزادان منجر شود.

فهرست منابع:

جواد زارعی، مهدیه ایزدی، امیر عباس عزیزی، و صدیقه نوح جاه، "پیش بینی زودهنگام دیابت بارداری با استفاده از الگوریتم های درخت تصمیم و شبکه عصبی مصنوعی،" مجله غدد درون ریز و متابولیسم ایران، vol. ۲۴، no. ۱ ، pp. ۱–۱۱، ۱۴۰۱، [Online]. Available: https://sid.ir/paper/۱۱۳۰۳۲۰/fa

منصور رضایی، نگین فخری، سوده شهسواری، و فاطمه رجعتی، "مقایسه پیش بینی ابتلا به دیابت بارداری با مدل های رگرسیون لجستیک, تحلیل ممیزی, درخت تصمیم و شبکه عصبی مصنوعی،" مجله اپیدمیولوژی ایران، vol. ۱۵، no. ۴ ، pp. ۳۶۲–۳۷۰، ۱۳۹۸، [Online]. Available: https://sid.ir/paper/۳۷۹۵۵۶/fa

[۱] Lin X, Xu Y, Pan X, Xu J, Ding Y, Sun X, et al. Global, regional, and national burden and trend of diabetes in ۱۹۵ countries and territories: an analysis from ۱۹۹۰ to ۲۰۲۵. Sci Rep ۲۰۲۰; ۱۰: ۱۴۷۹۰.

[۲] Dechereney A, Pernoll M. Current Obstetric& Gynecologic Diagnosis& Treatment. ۸th ed,Appleton& Lange, Mc Graw Hill Com; ۱۹۹۴. P: ۴۶۸-۷۰.

[۳] Nouhjah S, Shahbazian H, Latifi SM, Azizi Malamiri R,Ghodrati N. Body mass index growth trajectories from birth through ۲۴ months in Iranian infants of mothers with gestational diabetes mellitus. Diabetes Metab Syndr ۲۰۱۹; ۱۳:۴۰۸-۱۲.

[۴] Manshori A, Rezaeian M, Bagheri H, Aminzadeh F, Goujani R. Assessment of the appropriate cut-off point in glucose challenge test based on the risk of gestational diabetes in pregnant women. The Iranian Journal Of Obstetrics, Gynecology And Infertility. ۲۰۱۵; ۱۸: ۱-۸.

[۵] Cunningham FG. William's Obstetrics. ۲۲rd ed. New York: McGraw-Hill Com; ۲۰۰۵. p:۱۱۷۰-۸۲

[۶] James D, Weiner C, Steer P, Gonik B. High Risk Pregnancy. ۳rd ed. Philadelphia: WB Saunders; ۲۰۰۶. P: ۹۸۶-۹۰.

[۷] Forouhi NG, Wareham NJJM. Epidemiology of Diabet-es ۲۰۱۹; ۴۷: ۲۲-۷.

[۸] Engelgau MM, Herman WH, Smith PJ, German RR,Aubert RE. The epidemiology of diabetes and pregnancy in the US, ۱۹۸۸. Diabetes care. ۱۹۹۵; ۱۸: ۱۰۲۹-۳۳.

[۹] Sayehmiri F, Bakhtiyari S, Darvishi P, Sayehmiri K.Prevalence of Gestational Diabetes Mellitus in Iran: A Systematic Review and Meta-Analysis Study. The Iranian Journal Of Obstetrics, Gynecology And Infertility. ۲۰۱۳; ۱۵: ۱۶-۲۳.

[۱۰] Koivunen S, Viljakainen M, Männistö T, Männistö T, Gissler M, Pouta A, et al. Pregnancy outcomes acc-ording to the definition of gestational diabetes. PloS one ۲۰۲۰; ۱۵: e۰۲۲۹۴۹۶.

[۱۱] Manshori A, Rezaeian M, Bagheri H, Aminzadeh F, Goujani R.Assessment of the appropriate cut-off point in glucose challenge test based on the risk of gestational diabetes in pregnant women. The Iranian Journal Of Obstetrics, Gynecology And Infertility. ۲۰۱۵; ۱۸: ۱-۸.

[۱۲] Font-López KC, Marcial-Santiago AdR, Becerril-Cabrera JI. Validity of blood glucose fasting test as diagnostic for gestational diabetes during the first trimester of pregnancy.Ginecología y Obstetricia de México. ۲۰۱۸; ۸۶: ۲۳۳-۸.
[۱۳] Johnson RA, Wichern DW. Applied multivariate statistical analysis: Prentice hall Upper Saddle River, NJ; ۲۰۰۲.
[۱۴] Metzger BE, Coustan DR, Committee O. Summary and recommendations of the fourth international workshop-conference on gestational diabetes mellitus. Diabetes care. ۱۹۹۸; ۲۱: B۱۶۱.
[۱۵] Maji S, Arora S. Decision tree algorithms for prediction of heart disease. Information and communication tech-nology for competitive strategies: Springer ۲۰۱۹: ۴۴۷-۵۴.
[۱۶] Tougui I, Jilbab A, Mhamdi E. Heart disease class-ification using data mining tools and machine learning techniques. Health and Technology ۲۰۲۰; ۱۰: ۱۱۳۷-۴۴.
[۱۷] Khemphila A, Boonjing V, editors. Comparing performances of logistic regression, decision trees, and neural networks for classifying heart disease patients. Computer Information Systems and Industrial Management Applications (CISIM), ۲۰۱۰ International Conference on; ۲۰۱۰: IEEE.
[۱۸] Yadollahi P, Zangene N, Heiran A, Sharafi M, Heiran KN, Hesami k, et al. Effect of the COVID-۱۹ pandemic on maternal healthcare indices in Southern Iran: an interrupted time series analysis. BMJ Open ۲۰۲۲; ۱۲: e۰۵۹۹۸۳.
[۱۹] Joulaei H, Fatemi M , Hooshyar D, Karimi Rouzbahani A, Joulaei R, Foroozanfar Z. Analyzing delay in referral of pregnant women and children under five years old during the COVID-۱۹ pandemic: Fars Province, Iran.Health Care Women Int ۲۰۲۲: ۱-۱۷.
[۲۰] Hagan M. Neural network design, pws, USA; ۱۹۹۵.
[۲۱] Zarkogianni K, Vazeou A, Mougiakakou SG,Prountzou A, Nikita KS. An insulin infusion advisory system based on autotuning nonlinear model-predictive control. IEEE Trans Biomed Eng ۲۰۱۱;۵۸(۹):۲۴۶۷ -۷۷
[۲۲] Jaafar SFB, Ali DM. Diabetes mellitus forecast using artificial neural network (ANN). Asian Conference on Sensors and the International Conference on New Techniques in Pharmaceutical and Biomedical Research; ۲۰۰۵ Sep ۵-۷; Kuala Lumpur, Malaysia, Malaysia IEEE; ۲۰۰۵.
[۲۳] Kurt I, Ture M, Kurum AT. Comparing performances of logistic regression, classification and regression tree, and neural networks for predicting coronary artery disease. Expert systems with applications. ۲۰۰۸; ۳۴: ۳۶۶-۷۴.
[۲۴] Hirose Y, Yamashita K, Hijiya S. Back-propagation algorithm which varies the number of hidden units.Neural Networks ۱۹۹۱;۴(۱):۶۱-۶
[۲۵] O. Maimon and L. Rokach, Data Mining With Decision Trees: Theory and Applications, ۲nd ed. River Edge, NJ, USA: World Scientific Publishing Co., Inc., ۲۰۱۴. ISBN: ۹۷۸۹۸۱۴۵۹۰۰۷۵.
[۲۶] M. Krzywinski and N. Altman, "Points of Significance:Classification and regression trees," Nature Methods, ۲۰۱۷.
[۲۷] Du Y, Rafferty AR, McAuliffe FM, , Wei L, Mooney C. An explainable machine learning-based clinical deci-sion support system for prediction of gestational dia-betes mellitus. Sci Rep ۲۰۲۲; ۱۲: ۱۱۷۰
[۲۸] Zhang Z, Yang L, Han W, Wu Y, Zhang L, Gao C, et al.Machine learning prediction models for gestational Dia-betes mellitus: Meta-analysis. J Med Internet Res ۲۰۲۲;۲۴: e۲۶۶۳۴.
[۲۹] Rezaei M, Fakhri N, Rajati F, Shahsavari S. Comparison of gestational diabetes prediction with artificial neural network and decision tree models. Tehran University Medical Journal ۲۰۱۹; ۷۷: ۳۵۹-۶۷. [Farsi]