مجله‌ی غدد درون‌ریز و متابولیسم ایران
دانشگاه علوم پزشکی و خدمات بهداشتی ـ درمانی شهید بهشتی
دوره‌ی بیست و چهارم، شماره‌ی ۱، صفحه‌های ۱۱ ـ ۱ (فروردین ـ اردیبهشت ۱۴۰۱)

مقاله‌ی پژوهشی

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

دکتر جواد زارعی[۱] مهدیه ایزدی[۲]، دکتر امیر عباس عزیزی[۲]، دکتر صدیقه نوح جاه[۱]

۱) مرکز تحقیقات دیابت، پژوهشکده سلامت، دانشگاه علوم پزشکی جندی شاپور اهواز، اهواز، ایران، ۲) گروه فناوری اطلاعات سلامت، دانشکده پیراپزشکی، دانشگاه علوم پزشکی جندی شاپور اهواز، اهواز، ایران، **نشانی مکاتبه با نویسنده‌ی مسئول:** اهواز، بلوار گلستان، شهر دانشگاهی، دانشگاه علوم پزشکی جندی شاپور، مرکز تحقیقات دیابت، کدپستی ۱۵۷۹۴-۶۱۳۵۷، دکتر صدیقه نوح جاه:
e-mail:s_nouhjah@yahoo.com

چکیده

مقدمه: دیابت بارداری با عوارض متعدد کوتاه‌مدت و درازمدت در مادر و کودک همراه است. شناسایی عوامل خطرزای آن می‌تواند به تشخیص به‌موقع و پیشگیری از عوارض مرتبط با آن کمک کند. هدف از این مطالعه طراحی و مقایسه مدل‌های پیش‌بینی ابتلا به دیابت بارداری با استفاده از الگوریتم‌های هوش مصنوعی بود. مواد و روش‌ها: برای پیش‌بینی دیابت بارداری از الگوریتم‌های درخت تصمیم، و شبکه عصبی مصنوعی استفاده گردید. جامعه پژوهش ۱۲۷۰ زن باردار تحت پوشش مراکز بهداشتی درمانی شهر اهواز بودند که ۸۱۶ مورد آن‌ها سالم و ۴۵۴ مورد مبتلا به دیابت بارداری بودند. جهت ارزیابی کارآیی مدل‌ها؛ حساسیت، ویژگی، دقت و صحت محاسبه گردید. در نهایت از الگوریتم طبقه‌بندی **AdaBoost** برای تقویت مدل پیشنهادی استفاده گردید. یافته‌ها: پس از انجام تحلیل مولفه اساسی، نه متغیر برای مدل‌سازی اولیه انتخاب شدند. که در مدل شبکه عصبی مصنوعی، سطح زیر منحنی راک و حساسیت به ترتیب ۸۳/۲ درصد و ۸۵/۱ درصد بود، و برای مدل درخت تصمیم نیز سطح زیر منحنی راک و حساسیت به ترتیب ۰/۸۲۶ و ۸۴ درصد به‌دست آمد. پس از حذف متغیرها با وزن کمتر و تقویت مدل پیشنهادی، سطح زیر منحنی راک و حساسیت افزایش پیدا کرد (۰/۸۶۱ و ۹۲/۱ درصد). پنج متغیر شامل: قند خون ناشتا در اولین معاینه بارداری، سابقه دیابت بارداری در بارداری‌های قبلی، نمایه توده بدنی، سن مادر و سابقه خانوادگی دیابت، بالاترین دقت را در پیش‌بینی ابتلا به دیابت بارداری داشتند. نتیجه‌گیری: نتایج این مطالعه نشان داد که الگوریتم‌های هوش مصنوعی از دقت و کارآیی قابل توجهی برخوردارند و می‌توانند با پیش‌بینی زودرس دیابت بارداری در پیشگیری از پیامدهای منفی آن مؤثر باشند.

واژگان کلیدی: دیابت بارداری، یادگیری ماشین، هوش مصنوعی، درخت تصمیم، شبکه عصبی مصنوعی
دریافت مقاله: ۱۴۰۱/۶/۲۷ـ دریافت اصلاحیه: ۱۴۰۱/۸/۱۴ ـ پذیرش مقاله: ۱۴۰۱/۸/۳۰

مقدمه

دیابت یکی از مهم‌ترین چالش‌های بهداشت عمومی در جهان است که با شیوع بالا و عوارض متعدد بار زیادی را بر نظام سلامت تحمیل می‌کند.[۱] یکی از عومل خطر مهم برای ابتلا به دیابت نوع دو؛ سابقه دیابت بارداری (GDM)[i] است.[۲] GDM به درجات متغیری از عدم تحمل گلوکز که برای اولین بار در بارداری ایجاد یا تشخیص داده شود، اطلاق می‌شود.[۳] GDM به عنوان شایع‌ترین عارضه متابولیک دوران بارداری، می‌تواند منجر به عوارضی در طول بارداری و بعد از آن، برای مادر و جنین/نوزاد شود.[۴]

از جمله این عوارض می‌توان به ماکروزومی جنین، افزایش احتمال زایمان سزارین، سخت‌زایی شانه[ii]، افت قند خون نوزادان، و افزایش خطر چاقی و دیابت نوع ۲ در کودکان این مادران اشاره نمود.[۵] در مقایسه با کشورهای توسعه یافته، شیوع GDM در کشورهای فقیر و در حال

i -Gestational Diabetes(GDM)

ii- Shoulder Dystocia

توسعه؛ به دلایل عوامل مختلفی مثل فقر غذایی، سبک زندگی، شیوع چاغی و ضعف در دسترسی به غربالگری و تشخیص زودهنگام، بالاتر است.[۶،۷] علی‌رغم اهمیت GDM، توافقی در مورد رویکردی واحد جهت تشخیص، زمان مناسب انجام تست غربالگری و حتی درمان دیابت بارداری حاصل نشده است.[۸] آزمایش تحمل گلوکز خوراکی ۷۵ گرمی، به عنوان آزمون استاندارد طلایی غربالگری دیابت بارداری در هفته ۲۸-۲۴ بارداری، در بسیاری کشورها مورد استفاده قرار می‌گیرد.[۹] اما استفاده از این روش در عمل با محدودیت‌ها و مشکلاتی در اجرا همراه است.[۱۰] با توجه به اختلاف نظرها در مورد غربالگری و تشخیص دیابت بارداری، مطالعات اخیر بر تشخیص زودرس GDM براساس ویژگی‌های مادر، یافته‌های بالینی مادر، شاخص‌های زیستی در هفته‌های اول بارداری و طراحی مدل‌های پیش‌بینی با استفاده از ترکیب چند عامل خطر متمرکز شده‌اند.[۱۱،۱۲]

در سال‌های اخیر هم‌زمان با فناوری اطلاعات درحیطه پزشکی، یکی از مواردی که به شدت در زمینه کمک به تشخیص بیماری‌ها مورد توجه قرار گرفته است، استفاده از هوش مصنوعی در حوزه پزشکی است.[۱۳] یکی از مزایای الگوریتم‌های هوش مصنوعی، هم‌چون یادگیری ماشین،[i] توانایی این مدل‌ها در شناسایی روابط پنهان و غیر خطی بین داده‌ها است.[۱۴،۱۵] الگوریتم‌های یادگیری ماشین به عنوان یکی از زیر شاخه‌های هوش مصنوعی امکان طبقه‌بندی، خوشه‌بندی، انتخاب ویژگی و پیش‌بینی را بر اساس تحلیل داده‌ها و تبدیل آن‌ها به اطلاعات را فراهم می‌آورد. الگوریتم-های یادگیری ماشین تاکنون در زمینه‌های مختلف پژوهش-های سلامت، مانند تشخیص انواع سرطان، بیماری‌های قلبی و دیابت به کارگرفته شده و با نتایج مطلوبی همراه بوده است.[۱۶،۱۷]

شبکه‌های عصبی مصنوعی (ANN)[ii]، جنگل تصادفی[iii]، یادگیری عمیق[iv]، ماشین‌های بردار پشتیبانی[v]، الگوریتم‌های تکاملی[vi]، K نزدیک ترین همسایه[vii]، و درخت تصمیم[viii] مهم‌ترین رویکردهای هوش مصنوعی هستند که برای کشف روابط پنهان بین داده‌ها مورد استفاده قرار گرفته‌اند.[۱۸] ANN به دلیل دقت بالا و درخت تصمیم به علت سادگی و فهم آسان‌تر بیش از سایر الگوریتم‌ها مورد استفاده قرار می‌گیرند.[۱۹،۲۰]

مسئله مهم دیگری که استفاده از هوش مصنوعی در کمک به تشخیص زودهنگام و ارائه مراقبت از افراد در معرض خطر؛ مثل مادران باردار را پررنگ می‌کند، رخ داد همه‌گیری COVID-19 در سال ۲۰۱۹ بود. خطر بالای این بیماری برای مادران باردار از یک سو و اعمال قرنطینه‌های سراسری و محدودیت‌های شدید از سوی دیگر، سبب اختلال در دریافت مراقبت‌های معمول در زمان بارداری گردید.[۲۳-۱۲] در ایران نیز برخی از مطالعات انجام شده حاکی از تاثیر منفی همه‌گیری COVID-19 بر دریافت مراقبت‌های معمول دوران بارداری است.[۲۴،۲۵]

بنابراین، با توجه به اهمیت تشخیص زودهنگام دیابت بارداری و هم‌چنین قابلیت‌های الگوریتم‌های هوش مصنوعی، هدف از این مطالعه استفاده از دو الگوریتم مبتنی بر هوش مصنوعی ANN و درخت تصمیم برای پیش‌بینی خطر ابتلا به GDM به منظور کمک به تشخیص زودرس بیماری بود.

مواد و روش‌ها

داده‌های جمع‌آوری شده از دو طرح کوهورت دیابت بارداری ادغام، پردازش و آنالیز گردید. این دو طرح بین سالهای ۱۳۹۳ تا ۱۳۹۷ در مراکز بهداشتی درمانی شهر اهواز انجام شده بودند.[۶،۲۶] در طرح اول که با هدف تعیین میزان بروز دیابت بارداری، بر اساس معیارهای کارگروه بین‌المللی انجمن مطالعه دیابت و بارداری(IADPSG)،[ix] طراحی‌شده بود.[۲۷] در این طرح ۷۵۰ خانم باردار موردبررسی قرار گرفته بودند. در مطالعه دوم؛ پیامدهای متابولیک دیابت بارداری (۲۶۰ مورد) با گروه سالم (۲۶۰ مورد) مورد مقایسه قرار گرفته بودند. در پژوهش حاضر در مجموع داده‌های ۱۲۷۰ زن باردار مورد استفاده قرار گرفت که ۸۱۶ مورد آن‌ها سالم و ٤٥٤ مورد مبتلا به دیابت بارداری بودند.

در هر دو مطالعه دیابت بارداری بر اساس نتایج غربالگری دیابت بارداری در هفته ۲۸-۲۴ بارداری با استفاده از تست تحمل گلوکز خوراکی ۷۵ گرمی و معیارهای IADPSG تعریف شده بود. بر اساس این معیارها، وجود

i- Machine Learning
ii -Artificial Neural Network(ANN)
iii - Random Forest
iv -Deep Learning
v -Support Vector Machine
vi -Evolutionary Algorithm
vii-k-Nearest Neighbors
viii -Decision Tree

ix -International Association of the Diabetes and Pregnancy Study Groups (IADPSG)

حداقل یک قند غیرطبیعی به‌عنوان دیابت بارداری تعریف می‌شود: گلوکز ناشتای پلاسما مساوی یا بیشتر از ۹۲ میلی‌گرم/گلوکز یک‌ساعته مساوی یا بیشتر از ۱۸۰/گلوکز پلاسمای دوساعته مساوی یا بیشتر از ۱۵۳. [۲۸]

متغیرهای موردبررسی در این دو مطالعه شامل قند خون ناشتای اولین معاینه بارداری، سن مادر، وضعیت شغلی، قومیت مادر، سطح سواد، تعداد بارداری، سن ازدواج، سن اولین بارداری، سابقه دیابت بارداری در بارداری‌های قبلی، سابقه سقط جنین، مقدار هموگلوبین در اولین معاینه بارداری (در هفته ۱۰-٦ بارداری)، گروه خونی مادر، وزن ماه اول بارداری، وزن ماه نهم بارداری، نمایه توده بدنی[i] بر اساس وزن اولین معاینه بارداری (در اولین مرحله از مراقبت معمول دوران بارداری در هفته ۱۰-٦ بارداری)، فشارخون سیستولی، فشارخون دیاستول، سابقه دیابت در خانواده، بود. پس از اخذ داده‌های این دو مطالعه، تحلیل و مدل سازی داده‌ها در پنج مرحله زیر صورت گرفت:

مرحله اول: کنترل کیفی داده‌ها

در این مرحله فایل داده‌های حاصل از دو مطالعه قبلی با یکدیگر ادغام و ازنظر کیفیت داده‌ها مورد ارزیابی قرار گرفتند. در این مرحله، هدف شناسایی داده‌های پرت[ii]،داده‌های پرت [iii] و داده‌های ناموجود[iv] بود. کنترل کیفی داده‌ها با کمک مجریان دو مطالعه کوهورت انجام شد.

مرحله دوم: انتخاب متغیرهای اصلی برای مطالعه

در مرحله دوم، متغیرهای اصلی برای مدل‌های پیش‌بینی تعیین شدند. ابتدا بر اساس بررسی مطالعات قبلی انجام‌شده در خصوص دیابت بارداری و مشاوره با افراد متخصص طی پنج جلسه نشست خبرگان[v] به مدت ٤٥ الی ۱۲۰ دقیقه، ۱۸ متغیر از مجموع ۳۹ متغیر انتخاب شدند. این متغیرها شامل سن مادر، قند خون ناشتا در اولین معاینه بارداری، قند خون غربالگری در هفته ٢٤ تا ۲۸ بارداری، شغل مادر، تحصیلات مادر، تعداد بارداری، تعداد زایمان، سابقه سقط جنین، وزن قبل از بارداری، وزن ماه نهم بارداری، هموگلوبین اولین معاینه بارداری، نمایه توده بدنی(BMI) [vi]، گروه خون، فشارخون سیستولی، فشارخون دیاستول، سابقه دیابت در خانواده، سابقه قبلی دیابت بارداری (در بارداری‌های قبلی) و قومیت، بودند. در گام بعدی با استفاده از آزمون‌های آماری تی مستقل[vii] و مجذور کای[viii] برای سطح معناداری متغیرها در دو گروه دیابت بارداری و مادران سالم انجام شد. هدف انتخاب متغیرهایی برای ورود به مدل‌سازی بود که از نظر آماری در دو گروه معنی‌دار بودند.

مرحله سوم: پیاده‌سازی

در مرحله سوم مدل‌های پیش‌بینی دیابت بارداری با استفاده از الگوریتم‌های (شبکه عصبی، درخت تصمیم و...)، در محیط MATLAB ایجاد شدند. قبل از مدل‌سازی ابتدا با انجام روش تحلیل مولفه‌های اساسی(PCA)[ix] در نرم‌افزار MATLAB نسخه (R2022a) انجام گردید. دلیل انجام این کار تعداد بالای متغیرهای معنی‌دار و هم‌چنین وجود متغیرهای چند حالته (مثل سطح تحصیلات، گروه خون و ...) بود که می‌توانست پیچیدگی مدل را افزایش و بالطبع دقت مدل‌های پیشنهادی را کاهش دهد. بنابراین، جهت ساده‌تر شدن مدل‌های پیش‌بینی، پژوهش‌گران معمولاً از روش‌هایی مانند تحلیل مولفه اساسی جهت کاستن متغیرهای پیش‌بینی استفاده می‌کنند. لذا با استفاده روش PCA نمره عاملی هر کدام از عوامل خطرساز دیابت بارداری تعیین شد. در این روش اولین مولفه بدست آمده، بیشترین واریانس متغیرها را تعیین می‌کند و بر اساس آن وزن هر یک از عوامل مرتبط تعیین می‌گردد.[۲۹]

با انجام PCA متغیرهایی که در دیابت بارداری دارای وزن کمتری بودند، حذف شدند. پس از انجام PCA در نهایت ۹ متعیر سن مادر، سابقه دیابت بارداری در بارداری‌های قبلی، سابقه سقط‌جنین، هموگلوبین اولین معاینه بارداری، نمایه توده بدنی، فشارخون سیستولی، سابقه دیابت در خانواده، گروه خونی و تحصیلات مادر برای مدل‌سازی انتخاب شدند.

الگوریتم‌های ANN، و درخت تصمیم برای مدل‌سازی در نظر گرفته شدند. ANN یکی از الگوریتم‌هایی است که به‌طور گسترده در داده‌کاوی مورد استفاده قرار گرفته و با استفاده از آن ارتباطات غیرخطی و اثرات متقابل پیچیده، بین عوامل مختلف بررسی می‌شود. در این الگوریتم از چارچوب

i -Body mass index
ii- Outlyers
iii -Noise
iv -Missing Data
v -Expert panel
vi - Body Mass Index

vii -Independent t-test
viii -Chi square test
ix -Principal Component Analysis (PCA)

نورون‌های مغز انسان الهام گرفته شده است، و در صورت آموزش مناسب ANN ابزار قدرتمندی برای یافتن روابط پنهان بین داده‌های ورودی و خروجی است.[۱۹،۲۰] علاوه بر ANN درخت تصمیم یکی دیگر از الگوریتم‌هایی است که به علت سادگی و فهم آسان‌تر آن محبوبیت زیادی دارد. این روش با به‌کارگیری تکنیک‌های ساده یک الگوی رده‌بندی برای مشاهدات موجود معرفی می‌کند. با وجود این‌که در این روش از تکنیک‌های ساده استفاده می‌شود، اما می‌تواند در زمینه پیش‌بینی و تشخیص به‌خوبی روش‌های پیچیده‌تر عمل نمایند.[۳۰،۳۱]

هم‌چنین از پراستفاده‌ترین روش‌های درخت تصمیم می‌توان به C4.5، C5، CHAID[i] و C&R Tree[ii] اشاره کرد که که در این پژوهش از الگوریتم C4.5 برای درخت تصمیم استفاده شده است.

مرحله چهارم: ارزیابی عملکرد مدل‌های پیش‌بینی

در این مرحله، عملکرد مدل‌های استخراج‌شده مورد ارزیابی قرار گرفت. برای ارزیابی عملکرد مدل‌ها از معیارهای دقت[iii]، حساسیت[iv]، و ویژگی[v] و صحت[vi]، استفاده گردید.

دقت مدل، نشان‌دهنده این است که چه درصدی از مشاهدات توسط الگوریتم به کار رفته به درستی پیش‌بینی شده است. در واقع دقت به میزان توانایی الگوریتم در پیش‌بینی صحیح خروجی یا پیامد مدل (بر اساس درستی آموزش داده شده)، اشاره دارد. حساسیت به احتمال مثبت شدن صحیح نتیجه آزمون وقتی که نمونه دارای بیماری (GDM) است، اشاره می‌کند. ویژگی مدل به احتمال منفی شدن صحیح نتیجه آزمون وقتی که نمونه سالم (فاقد GDM) است، اشاره می‌کند. صحت مدل، توانایی الگوریتم در پیش‌بینی درست تعداد موارد صحیح ابتلا به بیماری (GDM) را نشان می‌دهد.[۳۲]

برای ارزیابی دقیق‌تر عملکرد مدل‌ها، مجموعه داده به زیرمجموعه‌های آموزش و تست تقسیم گردید. سپس مدل پیش‌بینی با مجموعه داده‌ی آموزش ایجاد و دقت آن با مجموعه داده‌ی تست مقایسه گردید. برای کاهش میزان تورش از روش K-Fold Cross استفاده شد که نتیجه آن در قالب دسته‌بندی ده گروه مشخص شده است. در این روش مجموعه داده اصلی به صورت تصادفی در K زیرگروه داده‌ای جداگانه قرار می گیرند. مدل پیش‌بینی در هر بار با K-1 زیرگروه ایجاد و با یک زیرگروه باقیمانده تست می‌شود. این فرایند آموزش و تست، K بار تکرار می‌گردد. دقت کلی مدل پیش‌بینی در این تکنیک از معادله‌ی زیر محاسبه می‌شود.

$$CVA = \frac{1}{K}\sum_{i}^{k} = 1^{Ai}$$

در این معادله، CVA[vii] ، بیانگر میانگین دقت مدل‌های پیش‌بینی و A معیار ارزیابی حساسیت، ویژگی، دقت، F- measure می‌باشد. در این مطالعه K برابر ۱۰ در نظر گرفته شده است.[۳۳]

از منحنی ROC[viii] برای مقایسه‌ی کارآیی مدل‌های پیش‌بینی، استفاده گردید. منحنی ROC یک نمودار گرافیکی است که جهت بیان نرخ مثبت صحیح در برابر نرخ مثبت غلط به کار می‌رود. محور عمودی این منحنی حساسیت و محور افقی آن همان ویژگی است. این منحنی یکی از ابزارهای سنجش کارآیی یک آزمون است. هر چه مساحت زیر منحنی ROC بیشتر باشد، دقت مدل پیش‌بینی در تفکیک صحیح مقادیر پیامد بیشتر خواهد بود.

مرحله پنجم: بهبود مدل‌های پیش‌بینی

بر اساس خروجی حاصل از ارزیابی مدل‌ها در مرحله چهارم، و محاسبه نسبت شانس (OR[ix])، متغیرهایی که تاثیر کمتری در پیش‌بینی دیابت بارداری داشتند، کنار گذاشته شدند و مدل سازی با استفاده از نه متغیر موثرتر صورت گرفت. برای محاسبه OR از روش رگرسیون لوجستیک استفاده گردید. در نهایت نیز ٤ متغیر با وزن کمتر کنار گذاشته شدند و ٥ متغیر برای استفاده در مدل پیشنهادی انتخاب گردیدند. مدل پیشنهادی ترکیبی از مدل ANN و درخت تصمیم بود. هم‌چنین، از الگوریتم تقویت‌کننده AdaBoost برای بهبود مدل‌سازی استفاده گردید. اساساً، AdaBoosting یکی از الگوریتم تقویت‌کننده موفقیت‌آمیز برای تقسیم‌بندی دوتایی و بهترین نقطه شروع برای درک مفهوم تقویت‌کنندگی است.[۳٤] علاوه بر این،

i- Chi-squared Automatic Interaction Detector
ii -Classification and Regression
iii -Accuracy
iv- Sensitivity
v -Specificity
vi- Precision

vii- Cross-validation Accuracy
viii- Receiver Operating Characteristics
ix -Odd Ratio

روش‌های تقویت‌کننده مدرن بر اساس AdaBoost ساخته می‌شوند؛ از جمله ماشین‌های تقویت گرادیان تصادفی. به طور کلی، AdaBoost با درختان تصمیم گیری کوچک مورد استفاده قرار می گیرد.[۳۵] بدین‌صورت که درخت اول ایجاد می شود و از عملکرد آن روی هر نمونه آموزشی، برای سنجش میزان توجه درخت بعدی به نمونه‌ها استفاده می شود. بنابراین، درخت باید به تمام نمونه‌های آموزشی توجه کند و به داده‌های آموزشی که پیش‌بینی آن‌ها دشوار است، وزن بیشتری بدهد، در حالی که به داده‌هایی که پیش‌بینی آن‌ها آسان است وزن کمتری بدهد.

یافته‌ها

میانگین سنی مادران مورد مطالعه ۵/۵۰±۲۸/۶۵ سال بود. مادرانی که دو بارداری یا بیشتر را تجربه کرده بودند ۶۶/۱ درصد و مادران دارای اضافه وزن یا چاق ۵۶/۶ درصد بودند. از مجموع ۱۸ متغیر انتخاب شده توسط خبرگان، ۱۳ متغیر در دو گروه مادران مبتلا به دیابت بارداری و سالم دارای رابطه معنی‌دار از نظر آماری بودند (جدول ۱ و ۲). همان‌طور که جدول ۱ نشان می‌دهد انجام آزمون تی مستقل برای متغیرهای کمی نشان داد که هر هفت متغیر کمی در دو گروه زنان مبتلا به دیابت بارداری و مادران سالم دارای رابطه معنادار آماری بودند. در جدول ۲، نیز آزمون‌های مجذور کای برای تعیین سطح معنی‌داری متغیرهای کیفی در گروه زنان مبتلا به دیابت بارداری و گروه سالم استفاده شده است.

جدول ۱- متغیرهای کمی معنی‌دار از نظر آماری در دو گروه مادران مبتلا به دیابت بارداری و مادران سالم

متغیر	مادران مبتلا به دیابت بارداری (N=۴۵۴)	گروه سالم (N=۸۱۶)	P value*
	میانگین (انحراف معیار)	میانگین (انحراف معیار)	
سن	۲۹/۹۰(۵/۲۸)	۲۷/۹۶(۵/۴۹)	۰/۰۰۱
وزن در اولین ماه بارداری	۶۹/۹۹(۱۲/۶۲)	۶۴/۹۵(۱۱/۶۸)	۰/۰۰۱
وزن در آخرین ماه بارداری	۸۱/۲۲(۱۲/۹۵)	۷۷/۷۱(۱۱/۷۸)	۰/۰۰۱
نمایه توده بدنی در اولین معاینه بارداری	۲۷/۵۰(۴/۹۱)	۲۵/۳۵(۴/۶۷)	۰/۰۰۱
قند خون ناشتا در اولین معاینه بارداری	۹۴/۶۸(۱۲/۵۹)	۸۰/۸۳(۷/۱۷)	۰/۰۰۱ .
هموگلوبین در اولین معاینه بارداری	۱۲/۱۶(۰/۷۶)	۱۱/۴۲(۰/۸۷)	۰/۰۰۱
فشارخون سیستولی در اولین معاینه بارداری	۱۱۱/۸۹(۱۳/۴۴)	۱۱۰/۰۷(۹/۷۷)	۰/۰۰۶

*آزمون تی مستقل

از مجموع ۱۳ متغیر معنی‌دار از نظر آماری، با استفاده از نرم‌افزار MATLAB و روش PCA تعدادی از متغیرها حذف شدند و نه متغیر که وزن بالاتری داشتند برای طراحی مدل‌های پیش‌بینی اولیه انتخاب شدند. این متغیرها شامل سن مادر، سطح تحصیلات مادر، سابقه قبلی دیابت بارداری (در بارداری‌های قبلی)، سابقه دیابت در خانواده، BMI، هموگلوبین در اولین معاینه بارداری، قند خون ناشتا در اولین معاینه بارداری، فشارخون سیستولی و سابقه سقط جنین بود. برای پیش‌بینی ابتلا به GDM مدل‌سازی با دو الگوریتم درخت تصمیم و ANN صورت گرفت. در برازش مدل‌های درخت تصمیم، در هر دو حالت با توجه به تابع ناخالصی و شاخص، درخت تصمیم C4.5 با عمق نه گره ساخته شد و متغیر قند ناشتا در گره ریشه و متغیر BMI در گره بعدی قرار گرفت.

مهم‌ترین عوامل خطر دیابت بارداری در مدل درخت تصمیم قند ناشتا، BMI ، سابقه GDM و سابقه خانوادگی دیابت بود. نتایج آنالیز حساسیت و اهمیت نسبی متغیرها در مدل ANN نشان داد که متغیرهای قند خون ناشتا در اولین معاینه بارداری، BMI، سابقه دیابت و سابقه GDM مهم‌ترین متغیرها در پیش‌بینی و شناسایی خطر ابتلا به دیابت بارداری بودند. هم‌چنین انجام رگرسیون لجستیک و محاسبه OR برای متغیرهای مورد استفاده بر روی داده‌ها

نیز نشان داد که به ترتیب سه متغیر قند خون ناشتا در اولین معاینه بارداری، سابقه خانوادگی دیابت، و سابقه قبلی دیابت بارداری (در بارداری‌های قبلی)، بیشترین وزن را در احتمال ابتلا به دیابت بارداری دارند (جدول۳).

جدول ۲- متغیرهای کیفی معنی‌دار از نظر آماری در دو گروه مادران مبتلا به دیابت بارداری و مادران سالم

متغیر		مادران مبتلا به دیابت بارداری (N=۴۵۴)	گروه سالم (N=۸۱۶)	P value*
		فراوانی (درصد فراوانی)	فراوانی (درصد فراوانی)	
سطح تحصیلات	بی‌سواد	(٪۲۹)	۱۳(٪۱/٦)	
	ابتدایی	۵۲(٪۱۱/۵)	۱۲٤(٪۱۵/۲)	
	راهنمایی	۱۳٤(٪۲۹/۵)	۳۳۱(٪٤۰/٦)	۰/۰۰۱
	متوسطه	۱٤۳(٪۳۱/۵)	۱٦۸(٪ ۲۰/٦)	
	دانشگاهی	۱۱٦(٪۲۵/٦)	۱۸۰(٪۲۲/۱)	
تعداد بارداری	۱	۱۳۰(٪۲۸/٦)	۳۰۰(٪۳٦/۸)	
	۲	۱۸٦(٪٤۱)	۳٤۰(٪٤۱/۷)	
	۳	۸۳(٪۱۸/۳)	۱۱۵(٪۱٤/۱)	۰/۰۰۲
	٤	۳٦(٪۷/۹)	٤۳(٪۵/۳)	
	بیشتر از ۵	۱۹(٪٤/۲)	۱۸(٪۲/۲)	
سابقه GDM† در بارداری‌های قبلی	بلی	٤۹(٪۱۵/۱)	۱۰(٪۱/۹)	۰/۰۰۱
	خیر	۲۷٦(٪۸٤/۹)	۵۱٤(٪۹۸/۱)	
سابقه خانوادگی دیابت	بلی	۹٦(٪۲۸/۳)	۱۰۸(٪۱۳/۲)	۰/۰۰۱
	خیر	۲۹٤(٪٦٤/۸)	۷۰۸(٪۸٦/۸)	
گروه خون	A	۱۰۱(٪۲۲/۲)	۲۵٤(٪۳۱/۱)	
	B	۸۰(٪۱۷/٦)	۱٦۵(٪۲۰/۲)	۰/۰۰۱
	AB	٤۰(٪۸/۸)	۸٤(٪۱۰/۳)	
	O	۲۳۳(٪۵۱/۳)	۳۱۳(٪۳۸/٤)	
سابقه سقط جنین	بلی	۹٦(٪۲۸/۳)	۱۱۸(٪۱۷/٤)	۰/۰۰۱
	خیر	۲٤۳(٪۷۱/۷)	۵٦۰(٪۸۲/٦)	

*آزمون مجذور کای، † در مادران با سابقه دو بارداری یا بیشتر

جدول ۳- محاسبه OR برای متغیرهای مورد استفاده با استفاده از رگرسیون لجستیک

متغیرها	سطح معنی‌داری	نسبت شانس	نسبت شانس با فاصله اطمینان ۹۵٪	
			حد پایین	حد بالا
سن مادر	۰/۰۰۱	۱/۰٦٦	۱/۰۳٤	۱/۱۰۳
سطح تحصیلات	۰/۰۵۲	۱/۱۷۵	۰/۹۹۹	۱/۳۸۳
سابقه خانوادگی دیابت	۰/۰۰۱	٦/۰۰۲	۲/٦۷۱	۱۳/٤۸۸
سابقه قبلی دیابت بارداری (در بارداری‌های قبلی)،	۰/۰۰۱	۲/٦۱۱	۱/۷٤۲	۳/۹۱٤
هموگلبین در اولین معاینه بارداری	۰/۰٦٦	۱/۱۳٤	۰/۹۹۲	۱/۲۹۵
نمایه توده بدنی (BMI)	۰/۰۰۱	۱/۱۰۱	۱/۰٦۱	۱/۱٤۳
سابقه سقط جنین	۰/٤٤٤	۱/۱٦٤	۰/۷۸۹	۱/۷۱۹
قند خون ناشتا در اولین معاینه بارداری	۰/۰۰۱	۱۵۸/۰۷۰	۳۸/۸۷٦	٦٤۲/۷۱۲
فشارخون سیستولی	۰/۲۵۹	۱/۰۰۹	۰/۹۹٤	۱/۰۲٤

در ادامه بر اساس نتایج حاصل از دو مدل درخت تصمیم و ANN، چهار متغیر فشارخون سیستولی، مقدار هموگلوبین در اولین معاینه بارداری، سطح تحصیلات مادر و سابقه سقط جنین، که از اهمیت کمتری در پیش‌بینی خطر ابتلا به دیابت بارداری داشتند، از مطالعه کنار گذاشته شدند و مدل‌سازی با استفاده از پنج متغیر باقی مانده صورت گرفت. سپس مدل پیش‌بینی از ترکیب ANN و درخت تصمیم‌گیری C4.5 ایجاد گردید. برای تقویت مدل پیشنهادی از الگوریتم AdaBoost استفاده گردید. برای ارزیابی دقیق‌تر کارآیی مدل پیشنهادی مجموعه داده به زیرمجموعه‌های آموزش (۷۰ درصد داده‌ها) و تست (۳۰ درصد داده‌ها) تقسیم گردید. سپس مدل‌های پیش‌بینی با مجموعه داده‌ی آموزش ایجاد و دقت آن با مجموعه داده‌ی تست ارزیابی گردد.

برای ارزیابی دقیق‌تر کارآیی مدل پیشنهادی پیش‌بینی GDM با استفاده از روش K-Fold Cross Validation در ۱۰ گروه مدل‌سازی صورت گرفت. جدول ٤ نتایج حاصل از ارزیابی روش K-Fold Cross Validation را نشان می‌دهد.

در جدول ٥ مقایسه حساسیت، ویژگی، صحت و دقت دو مدل درخت تصمیم و ANN را با مدل پیشنهادی نشان می‌دهد. همان‌طور که جدول ٤ نشان می‌دهد؛ عملکرد مدل پیشنهادی از دو مدل درخت تصمیم و ANN بیشتر است.

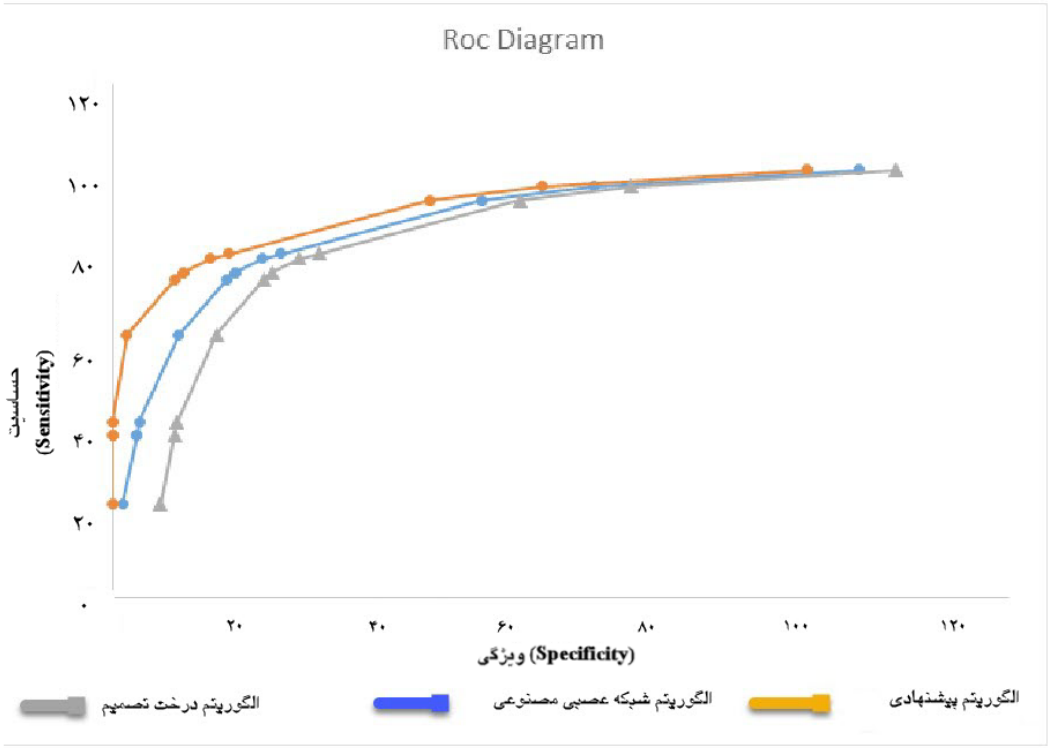
جدول ۴- مقایسه متغیرهای مستقل با متغیر وابسته (قند خون ناشتا) در پیش‌بینی دیابت بارداری با روش K-Fold Cross Validation

گروه داده‌ای	صحت (Accuracy)	پوشش (Recall)	دقت (Precision)
گروه داده‌ای ۱	۰/۹٥	۰/۸۰	۰/۹۷
گروه داده‌ای ۲	۰/۹۳	۰/٦۱	۰/۹٤
گروه داده‌ای ۳	۰/۹٤	۰/٦٦	۰/۹۰
گروه داده‌ای ٤	۰/۹۷	۰/۸۸	۰/۹۹
گروه داده‌ای ٥	۰/۹٦	۰/۸٥	۰/۹۸
گروه داده‌ای ٦	۰/۹٥	۰/۸٤	۰/۹۷
گروه داده‌ای ۷	۰/۹۳	۰/۷۰	۰/۹۱
گروه داده‌ای ۸	۰/۹٦	۰/۸۷	۰/۹۸
گروه داده‌ای ۹	۰/۹۷	۰/۸۰	۰/۹٥
گروه داده‌ای ۱۰	۰/۹٤	۰/۸٦	۰/۹۷
میانگین	۰/۹٥	۰/۷۹	۰/۹٦

جدول ۵- مقایسه حساسیت، ویژگی، دقت و صحت بین مدل‌های پیش‌بینی

مدل	حساسیت (Sensitivity)	ویژگی (Specificity)	صحت(Precision)	دقت (Accuracy)
	درصد	درصد	درصد	درصد
درخت تصمیم	٪۸٤	٪۸۳	٪۸۷/۳	٪۸۹/۳
ANN	٪۸٥/۱	٪۸٦/٥	٪۸۹/٦	٪۹۱/۳
مدل پیشنهادی	٪۹۲/۱	٪۹٤/۱	٪۹٥/۸	٪۹٦/۱

سطح زیر منحنی مدل‌های طراحی شده و مقایسه آن‌ها در نمودار ۱ و جدول ٦ ارائه شده است. با توجه به این‌که بیشترین مساحت زیر منحنی مربوط به مدل پیشنهادی است می‌توان نتیجه گرفت که کارآیی مدل پیشنهادی در تشخیص دیابت بارداری بیشتر از دو مدل درخت تصمیم و ANN است. هم‌چنین کارآیی مدل ANN از درخت تصمیم بیشتر بود.



نمودار ۱- مقایسه مساحت زیر منحنی ROC بین مدل پیشنهادی (proposed model) با دو مدل مدل درخت تصمیم (Decision Tree) وشبکه عصبی مصنوعی (ANN)

جدول ۶- مقایسه مساحت زیر منحنی ROC بین مدل پیشنهادی با دو مدل مدل درخت تصمیم و ANN

مدل پیشنهادی	ANN	DT	ROC curve
۰/۸۶۱	۰/۸۳۲	۰/۸۲۶	مساحت زیر منحنی ROC

بحث

تشخیص زودهنگام دیابت بارداری نقش مهمی در مدیریت آن داشته و امکان تصمیم‌گیری آگاهانه بالینی و مداخلات درمانی را فراهم نموده و می‌تواند به کاهش عوارض کوتاه‌مدت و طولانی‌مدت GDM منجر گردد.[۳۶-۳۸] هم‌چنین برخی از مطالعات نشان داده‌اند که با تلفیق مدل‌های پیش‌بینی GDM با سیستم‌های پشتیبانی از تصمیم‌گیری بالینی؛ می‌توان به صورت عملی از این مدل‌ها در مراکز بهداشتی و درمانی برای بهبود مراقبت بارداری استفاده کرد.[۳۹]

در مطالعه حاضر سه مدل درخت تصمیم، ANN و مدل تلفیقی (تقویت شده با الگوریتم AdaBoost) بر روی داده‌ها برازش و عملکرد مدل‌ها مقایسه شد. اعتبارسنجی الگوریتم‌های مورد استفاده نشان داد که سه مدل استفاده شده عملکرد قابل قبولی در پیش‌بینی GDM دارند. اگر چه در مقایسه با دیابت نوع دوم مطالعات محدودتری با استفاده از الگوریتم‌های یادگیری ماشین به پیش‌بینی دیابت بارداری انجام شده است و حتی در برخی از مطالعات انجام شده در مورد GDM خروجی مدل‌های مبتنی بر یادگیری ماشین از مدل‌های آماری مثل رگرسیون لوجستیک بهتر نبوده است،[۴۰] اما این مدل‌ها از جذابیت بیشتری بین پژوهش‌گران برخوردار هستند.[۴۱] در بیشتر مطالعات انجام شده از چندین

مدل مختلف برای پیش‌بینی GDM استفاده و عملکرد مدل‌های مختلف با هم مقایسه شده است.[۴۰،۴۲-۴۶] استفاده از چندین مدل امکان انتخاب بهترین مدل با عملکرد بالا در پیش‌بینی خطر ابتلا به بیماری را فراهم می‌کند.[۴۱]

در مطالعه حاضر انتخاب متغیرهای مطالعه در چهار مرحله شامل؛ استفاده از جلسه نشست خبرگان، انجام آزمون آماری برای تعیین سطح معناداری، روش PCA، محاسبه OR و انتخاب متغیرها با اهمیت‌تر بر اساس خروجی مدل‌های درخت تصمیم و ANN صورت گرفت. انتخاب متغیرهای اصلی برای مدل‌سازی به منظور اطمینان از عملکرد بهتر مدل‌ها و قابلیت کاربرد آن‌ها در عرصه بالینی صورت می‌گیرد.[۴۲] در این مطالعه برای مدل سازی اولیه از نه متغیر سن مادر، سطح تحصیلات مادر، سابقه قبلی دیابت بارداری (در بارداری‌های قبلی)، قند خون ناشتا در اولین معاینه بارداری، سابقه دیابت در خانواده، BMI، هموگلوبین در اولین معاینه بارداری و سابقه سقط جنین برای پیش‌بینی ابتلا به GDM استفاده شد. در نهایت برای مدل نهایی پنج متغیر قند خون ناشتا در اولین معاینه بارداری، BMI، سابقه GDM در بارداری‌های قبلی، سابقه خانوادگی دیابت، و سن مادر که دارای اهمیت بالاتری در پیش‌بینی خطر ابتلا به GDM بودند، انتخاب شدند. متغیر قند خون ناشتا در اولین معاینه بارداری بیش از سایر متغیرها در پیش‌بینی دیابت بارداری نقش داشت. زانگ[i] و همکاران (۲۰۲۲) در مطالعه‌ای به بررسی ۲۵ مطالعه مختلف انجام شده در ارتباط با پیش‌بینی GDM با استفاده هوش مصنوعی پرداختند. نتایج مطالعه آن‌ها نشان داد که چهار متغیر سن مادر، سابقه خانوادگی دیابت، BMI و قند خون ناشتا متداول‌ترین متغیرهای مورد استفاده در پیش‌بینی GDM بودند.[۴۱] در مطالعه یه[ii] و همکاران در چین[۴۰] و رضایی و همکاران در کرمانشاه[۴۶] قند خون ناشتا مهم‌ترین متغیر در پیش‌بینی GDM بود. بر اساس نتایج پژوهش حاضر، کاهش تعداد متغیرها و مدل‌سازی بر اساس متغیرهای مهم‌تر منجر به ارتقاء عملکرد مدل پیش‌بینی GDM گردید. برای مدل پیشنهادی مساحت زیر منحنی راک ۰/۸۶۱و حساسیت مدل ۹۲/۱ درصد به دست آمد. هم‌چنین عملکرد ANN در پیش‌بینی ابتلا به GDM، بر اساس شاخص‌های ارزیابی، بهتر از مدل درخت تصمیم بود. در مطالعات مشابه انجام شده نتایج مختلفی به دست آمده است. زانگ و همکاران در بررسی متاآنالیز ۲۵ مطالعه انجام شده در مجموع میانگین ۰/۸۴۹ را برای مساحت زیر منحنی راک و میانگین ۶۹درصد را برای حساسیت ذکر کرده بودند.[۴۱] در دو مطالعه مشابه داخلی انجام شده در ایران، مشابه با مطالعه حاضر از دو الگوریتم ANN و درخت تصمیم برای مدل‌سازی استفاده شده بود.[۴۴،۴۶] در مطالعه رضایی و همکاران در کرمانشاه بر روی داده‌های ۴۰۰ بیمار، مساحت زیر منحنی راک برای مدل ANN و برای درخت تصمیم به ترتیب ۰/۷۹و ۰/۷۴به دست آمد. هم‌چنین مشابه با مطالعه حاضر، در مجموع عملکرد مدل ANN بیشتر از درخت تصمیم بود.[۶۴] اما در مطالعه روحانی و همکاران بر روی داده‌های ۲۴۸ بیمار در تهران، در مجموع عملکرد مدل درخت تصمیم بهتر از مدل ANN بود. صحت مدل درخت تصمیم ۹۳ درصد و مدل ANN ، ۹۰ درصد بود.[۴۴] تفاوت در شاخص‌های عملکرد در مدل‌های پیش‌بینی در مطالعات مشابه مختلف، جدا از الگوریتم مورد استفاده برای مدل‌سازی، بستگی زیادی به حجم نمونه، تعداد متغیرهای انتخابی، نوع متغیرهای دموگرافیکی و بالینی انتخاب شده، ویژگی متغیرهای مورد مطالعه، تعداد داده‌های ناموجود در پایگاه داده و صحت داده‌ها دارد.[۴۱] لذا بالاتر بودن شاخص‌های عملکردی در مطالعه حاضر نسبت به برخی از مطالعات می‌تواند ناشی از موارد مذکور باشد. از طرف دیگر دستیابی به مدل‌های پیش‌بینی‌کننده‌ای با حساسیت و ویژگی قابل قبول دشوار است. استفاده از الگوریتم‌های پیچیده، اجرای مدل‌ها بر جمعیت‌های با حجم نمونه کم و غیرقابل تعمیم، نیاز به انجام آزمون‌های آزمایشگاهی خاص در زمان‌های ویژه، از جمله مشکلات این مدل‌ها هستند. هم‌چنین عوامل خطرزای دیابت بارداری متعدد بوده و بر اساس تحقیقات جدید بر تعداد این عوامل اضافه می‌شود و بر چالش‌های موجود در طراحی مدل‌های پایا می‌افزاید.[۴۱]

یکی از نقاط قوت این مطالعه استفاده از داده‌های استاندارد دو مطالعه کوهورت و استفاده از متغیرهایی بود که به سادگی در ابتدای بارداری قابل اندازه‌گیری هستند، به طور معمول اندازه‌گیری می‌شوند و هزینه زیادی را بر بیمار و سیستم بهداشتی درمانی تحمیل نمی‌کنند. از جمله نقاط ضعف این مطالعه استفاده از تعداد محدودی از الگوریتم‌ها برای مدل‌سازی بود.

i- Zhang
ii- Ye

نتیجه‌گیری

نتایج این مطالعه نشان داد که الگوریتم‌های هوش مصنوعی، به ویژه شبکه عصبی مصنوعی، دقت قابل قبولی در پیش‌بینی زودرس ابتلا به دیابت بارداری دارند. استفاده از هوش مصنوعی می‌تواند مدیریت دیابت بارداری را بهبود ببخشد، اگر چه در تشخیص‌های بالینی از این مدل‌ها کمتر استفاده می‌شود. با توجه به تجربه برخی از مراکز درمانی در دنیا در استفاده از این الگوریتم‌ها در سیستم‌های پشتیبانی از تصمیم‌گیری پزشکی، پیشنهاد می‌گردد که هم‌زمان با برنامه‌های وزارت بهداشت برای توسعه پرونده الکترونیک سلامت از این الگوریتم در پرونده الکترونیک سلامت استفاده گردد.

سپاسگزاری: این مقاله برگرفته از پایان‌نامه کارشناسی ارشد مهدیه ایزدی رشته انفورماتیک پزشکی با عنوان ارائه مدل پیش‌بینی تشخیص دیابت بارداری با استفاده از الگوریتم‌های هوش مصنوعی با کد D-0003 است. لذا پژوهش‌گران از مرکز تحقیقات دیابت، دانشگاه علوم پزشکی جندی‌شاپور اهواز به خاطر حمایت مالی از این مطالعه تشکر می‌کنند. هم‌چنین پژوهش‌گران از آقای دکتر رضازاده به دلیل همکاری در انجام مطالعه نیز تشکر می‌کنند.

ملاحظات اخلاقی

به‌منظور رعایت اصول اخلاق در پژوهش، مطالعه حاضر در کمیته اخلاق در پژوهش دانشگاه علوم پزشکی جندی‌شاپور اهواز مطرح و با کد اخلاق IR.AJUMS.REC.1400.135 تصویب گردید.

تعارض منافع: نویسندگان اعلام می‌نمایند که تضاد منافعی در این پژوهش وجود ندارد.

References

1. Lin X, Xu Y, Pan X, Xu J, Ding Y, Sun X, et al. Global, regional, and national burden and trend of diabetes in 195 countries and territories: an analysis from 1990 to 2025. Sci Rep 2020; 10: 14790.
2. Vounzoulaki E, Khunti K, Abner SC, Tan BK, Davies MJ, Gillies CL. Progression to type 2 diabetes in women with a known history of gestational diabetes: systematic review and meta-analysis. BMJ 2020; 369: m1361.
3. Saeedi M, Cao Y, Fadl H, Gustafson H, Simmons D. Increasing prevalence of gestational diabetes mellitus when implementing the IADPSG criteria: a systematic review and meta-analysis. Diabetes Res Clin Pract 2021; 172: 108642.
4. Zaman F, Nouhjah S, Shahbazian H, Shahbazian N, Latifi SM, Jahanshahi A. Risk factors of gestational diabetes mellitus using results of a prospective population-based study in Iranian pregnant women. Diabetes Metab Syndr 2018; 12: 721-25.
5. Nouhjah S, Shahbazian H, Latifi SM, Azizi Malamiri R, Ghodrati N. Body mass index growth trajectories from birth through 24 months in Iranian infants of mothers with gestational diabetes mellitus. Diabetes Metab Syndr 2019; 13: 408-12.
6. Nouhjah S, Shahbazian H, Jahanfar S, et al. Early postpartum lipid profile in women with and without gestational diabetes mellitus: results of a prospective cohort study. Iranian Red Crescent Medical Journal 2017; 19: 1-10.
7. Forouhi NG, Wareham NJJM. Epidemiology of Diabetes 2019; 47: 22-7.
8. Minschart C, Beunen K, Benhalima K. An Update on Screening Strategies for Gestational Diabetes Mellitus: A Narrative Review. Diabetes, Metabolic Syndrome and Obesity: Targets and Therapy 2021; 14: 3047-76.
9. Koivunen S, Viljakainen M, Männistö T, Männistö T, Gissler M, Pouta A, et al. Pregnancy outcomes according to the definition of gestational diabetes. PloS one 2020; 15: e0229496.
10. Jamieson EL, Spry EP, Kirke AB, Atkinson DN, Marley JV. Real-world gestational diabetes screening: problems with the oral glucose tolerance test in rural and remote Australia. Int J Environ Res Public Health 2019; 16: 4488.
11. Muche AA, Olayemi OO, Gete YKJAoPH. Prevalence and determinants of gestational diabetes mellitus in Africa based on the updated international diagnostic criteria: a systematic review and meta-analysis. Arch Public Health 2019; 77: 1-20.
12. Li-Zhen L, Yun X, Xiao-Dong Z, Shu-Bin H, Zi-Lian W, Sandra A, et al. Evaluation of guidelines on the screening and diagnosis of gestational diabetes mellitus: systematic review. BMJ Open 2019; 9: e023014.
13. Malik M, Tariq MI, Kamran M, Naqvi, M.R. Artificial intelligence in medicine. Advances in Smart Vehicular Technology, Transportation, Communication and Applications. Springer Singapore 2021: 159-70.
14. Ray S. A quick review of machine learning algorithms. 2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon): IEEE 2019: 35-9.
15. Sarker IH. Machine learning: Algorithms, real-world applications and research directions. SN Computer Science 2021; 2: 1-21.
16. Palimkar P, Shaw RN, Ghosh A. Machine learning technique to prognosis diabetes disease: random forest classifier approach. Advanced Computing and Intelligent Technologies: Springer 2022: 219-44.
17. Padmaja B, Srinidhi C, Sindhu K,Vanaja K,Deepika NM,Rao Patro K . Early and accurate prediction of heart disease using machine learning model. Turkish Journal of Computer and Mathematics Education 2021; 12: 4516-28.
18. Bansal M, Goyal A, Choudhary A. A comparative analysis of K-Nearest Neighbour, Genetic, Support Vector Machine, Decision Tree, and Long Short Term Memory algorithms in machine learning. Decision Analytics Journal 2022; 3: 100071.
19. Maji S, Arora S. Decision tree algorithms for prediction of heart disease. Information and communication technology for competitive strategies: Springer 2019: 447-54.
20. Tougui I, Jilbab A, Mhamdi E. Heart disease classification using data mining tools and machine learning techniques. Health and Technology 2020; 10: 1137-44.
21. Goyal M, Singh P, Singh K, Shashank S, Agrawal N, Misra S. The effect of the COVID-19 pandemic on ma-

ternal health due to delay in seeking health care: experience from a tertiary center. International Journal of Gynecology & Obstetrics 2021; 152: 231-35.
22. Sahin BM, Kabakci EN. The experiences of pregnant women during the COVID-19 pandemic in Turkey: A qualitative study. Women Birth 2021; 34: 162-69.
23. Townsend R, Chmielewska B, Barratt I, Kalafat E, Meulen J, Gurol-Urganci I, et al. Global changes in maternity care provision during the COVID-19 pandemic: A systematic review and meta-analysis. eClinicalMedicine 2021; 37: 100947.
24. Yadollahi P, Zangene N, Heiran A, Sharafi M, Heiran K N, Hesami k, et al. Effect of the COVID-19 pandemic on maternal healthcare indices in Southern Iran: an interrupted time series analysis. BMJ Open 2022; 12: e059983.
25. 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-19 pandemic: Fars Province, Iran. Health Care Women Int 2022: 1-17.
26. Shahbazian H, Nouhjah S, Shahbazian N, Jahanfar S, Latifi S M, Aleali A, et al. Gestational diabetes mellitus in an Iranian pregnant population using IADPSG criteria: Incidence, contributing factors and outcomes. Diabetes Metab Syndr 2016; 10: 242-46.
27. Contreras I, Vehi J. Artificial intelligence for diabetes management and decision support: literature review. J Med Internet Res 2018; 20: e10775.
28. Trujillo J, Vigo A, Duncan BB, Falavigna M,Wendland E M, Campos M A, et al. Impact of the International Association of Diabetes and Pregnancy Study Groups criteria for gestational diabetes. Diabetes Research and Clinical Practice 2015; 108: 288-95.
29. Kherif F, Latypova A. Principal component analysis. Machine Learning: Elsevier 2020: 209-25.
30. Odukoya O, Nwaneri S, Odeniyi I, Akodu B, Oluwole E, Olorunfemi G, et al. Development and Comparison of Three Data Models for Predicting Diabetes Mellitus Using Risk Factors in a Nigerian Population. Healthc Inform Res 2022; 28: 58-67.
31. Mirsharif M, Rouhani S. Data Mining Approach based on Neural Network and Decision Tree Methods for the Early Diagnosis of Risk of Gestational Diabetes Mellitus. Journal of Health and Biomedical Informatics 2017; 4: 59-68. [Farsi]
32. Karami T, Izadi M, Niaparast M. Comparison of logistic regression with some machine learning methods in classifying data. Andishe-ye-Amari 2021; 26: 47-59. [Farsi]
33. Wong T-T, Yeh P-YJIToK, Engineering D. Reliable accuracy estimates from k-fold cross validation. 2019; 32: 1586-94.
34. Wang F, Li Z, He F, Wang R, Yu W, Nie F. Feature learning viewpoint of AdaBoost and a new algorithm. 2019; 7: 149890-99.
35. Hu H, Siala M, Hebrard M, Huguet M-J. Learning optimal decision trees with maxsat and its integration in adaboost. IJCAI-PRICAI 2020, 29th International Joint Conference on Artificial Intelligence and the 17th Pacific Rim International Conference on Artificial Intelligence; 2020.
36. Xiong Y, Lin L, Chen Y, Salerno S, Li Y, Zeng X, et al. Prediction of gestational diabetes mellitus in the first 19 weeks of pregnancy using machine learning techniques. J Matern-Fetal Neonatal Med 2020; 35: 1-7.
37. Mennickent D, Rodríguez A, Farías-Jofré M , Araya J, Guzmán-Gutiérrez E. Machine learning-based models for gestational diabetes mellitus prediction before 24–28 weeks of pregnancy: A review. Artif Intell Med 2022; 132: 102378.
38. Artzi NS, Shilo S, Hadar E, Rossman H , Barbash-Hazan S, Ben-Haroush A, et al. Prediction of gestational diabetes based on nationwide electronic health records. Nature Medicine 2020; 26: 71-6.
39. Du Y, Rafferty AR, McAuliffe FM, , Wei L, Mooney C. An explainable machine learning-based clinical decision support system for prediction of gestational diabetes mellitus. Sci Rep 2022; 12: 1170.
40. Ye Y, Xiong Y, Zhou Q, Wu J , Li X , Xiao X . Comparison of machine learning methods and conventional logistic regressions for predicting gestational diabetes using routine clinical data: a retrospective cohort study. J Diabetes Res 2020; 2020: 4168340.
41. Zhang Z, Yang L, Han W, Wu Y, Zhang L, Gao C, et al. Machine learning prediction models for gestational Diabetes mellitus: Meta-analysis. J Med Internet Res 2022; 24: e26634.
42. Wu Y-T, Zhang C-J, Mol BW, Kawai A, Li C, Chen L, et al. Early prediction of gestational diabetes mellitus in the Chinese population via advanced machine learning. J Clin Endocrinol Metab 2021; 106: e1191-e205.
43. Xiong Y, Lin L, Chen Y, Salerno S, Li Y, Zeng X, et al. Prediction of gestational diabetes mellitus in the first 19 weeks of pregnancy using machine learning techniques. J Matern Fetal Neonatal Med 2020; 35: 2457-63.
44. Rouhani S, MirSharif M. Data mining approach for the early risk assessment of gestational diabetes mellitus. International Journal of Knowledge Discovery in Bioinformatics (IJKDB) 2018; 8: 1-11.
45. Hosseinpoor M. Predicting Gestational Diabetes Using an Intelligent Algorithm Based on Artificial Neural Network. Journal of Modern Medical Information Sciences 2022; 8: 126-39 [Farsi].
46. 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 2019; 77: 359-67. [Farsi]

Original Article

Early Prediction of Gestational Diabetes Using Decision Tree and Artificial Neural Network Algorithms

Zarei J[1], Izadi M[2], Azizi AA[2], Nouhjah S[1]

[1]Diabetes Research Center, Health Research Institute, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran
[2]Department of Health Information Technology, School of Allied Medical Sciences, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, I.R. Iran

e-mail:s_nouhjah@yahoo.com

Received: 18/09/2022 **Accepted**: 21/11/2022

Abstract

Introduction: Gestational diabetes is associated with many short-term and long-term complications in mothers and newborns; hence, the detection of its risk factors can contribute to the timely diagnosis and prevention of relevant complications. The present study aimed to design and compare Gestational diabetes mellitus (GDM) prediction models using artificial intelligence algorithms. Materials and Methods: In this study, Decision Tree and Artificial Neural Network algorithms were used to predict GDM. The research population encompassed 1270 pregnant women referred for primary care at urban healthcare centers in Ahvaz, of whom 816 persons were healthy, and 454 individuals were diagnosed with GDM. To evaluate the effectiveness of the GDM prediction models, their sensitivity, specificity, precision, and accuracy were calculated and compared. Finally, the AdaBoost classification algorithm was used to boost the two proposed models. Results: Following the Principal Component Analysis (PCA), nine cases were selected for primary modeling. In the Artificial Neural Network model, the area under the ROC curve and sensitivity were 83.2 and 85.1%, respectively, and the area under the ROC curve and sensitivity for the Decision Tree model were 0.826 and 84%, respectively. After removing variables with lower weights and reinforcing the proposed model, the level under the rock curve and sensitivity increased by 0.861 and 92.1%, respectively. In this regard, fasting blood sugar at the first pregnancy visit, history of gestational diabetes in previous pregnancies, body mass index, mothers' age, and family history of diabetes had the highest accuracy in predicting GDM. Conclusion: The findings of this study indicate that artificial intelligence algorithms are accurate and effective for the early prediction of gestational diabetes.

Keywords: Gestational diabetes, Artificial intelligence, Decision tree, Artificial neural network, Machine learning