مجله تخصصی اپیدمیولوژی ایران
https://irje.tums.ac.ir/
دانشگاه علوم پزشکی و خدمات بهداشتی درمانی تهران
IrEA انجمن علمی اپیدمیولوژیست های ایران

مقاله پژوهشی

پیش‌بینی وقوع زایمان زودرس و تعیین عوامل خطر آن به‌صورت شخصی با استفاده از مدل یادگیری ماشینی تفسیرپذیر

رامین فرخی[۱]، سمانه حسین زاده[۲]، عباس حبیب اللهی[۳]، اکبر بیگلریان[۴]

۱- دانشجوی کارشناسی ارشد گروه آمار زیستی و اپیدمیولوژی، دانشکده سلامت اجتماعی، دانشگاه علوم توانبخشی و سلامت اجتماعی، تهران، ایران.
۲- استادیار گروه آمار زیستی و اپیدمیولوژی، دانشکده سلامت اجتماعی، دانشگاه علوم توانبخشی و سلامت اجتماعی، تهران، ایران.
۳- استادیار نوزادان، دفتر سلامت خانواده و جمعیت وزارت بهداشت درمان و آموزش پزشکی، تهران، ایران.
۴- استاد گروه آمار زیستی و اپیدمیولوژی، مرکز تحقیقات عوامل اجتماعی موثر بر سلامت، پژوهشکده سلامت اجتماعی، دانشگاه علوم توانبخشی و سلامت اجتماعی، تهران، ایران.

اطلاعات مقاله

تاریخ دریافت
۱۴۰۲/۱۰/۰۶
تاریخ پذیرش
۱۴۰۳/۰۲/۰۵
نویسنده رابط
اکبر بیگلریان
ایمیل نویسنده رابط
abiglaria@uswr.ac.ir
نشانی نویسنده رابط
مرکز تحقیقات عوامل اجتماعی موثر بر سلامت، پژوهشکده سلامت اجتماعی، دانشگاه علوم توانبخشی و سلامت اجتماعی، تهران، ایران
واژگان کلیدی: بارداری، زایمان زودرس، یادگیری ماشینی، تفسیرپذیری، مدل-آگنوستیک

چکیده

مقدمه و اهداف: شناسایی زنان بارداری که در معرض زایمان زودرس هستند و همچنین تعیین عوامل خطر موثر برآن، امری ضروری است که در سلامت نوزادان تاثیرگذار است. این مطالعه با هدف به‌کارگیری نوعی از مدل یادگیری ماشینی تفسیرپذیر برای پیش‌بینی زایمان زودرس انجام شد.

روش‌کار: این مطالعه به‌صورت مقطعی انجام شد و از داده‌های ۱۴۹۳۵۰ مورد تولد شهر تهران در سال ۱۳۹۹ از مجموعه داده شبکه مادران و نوزادان ایران (IMaN) استفاده گردید. در این مطالعه، عوامل مختلف وابسته به مادر و جنین مانند متغیرهای جمعیت‌شناختی مادر نوزاد، وضعیت سلامت و سوابق بیماری مادر، شرایط بارداری، زایمان و خطرات آن استفاده شد. پس از پیش‌پردازش و آماده‌سازی داده‌ها، از مدل‌های یادگیری ماشینی شبکه عصبی چندلایه، جنگل تصادفی و XGBoost برای پیش‌بینی وقوع زایمان زودرس استفاده گردید. مدل‌ها بر اساس معیارهای دقت، حساسیت، ویژگی و سطح زیر منحنی راک ارزیابی شدند. برای تحلیل داده‌ها از زبان برنامه‌نویسی پایتون نسخه ۳-۱۰-۰ استفاده شد.

یافته‌ها: ۸/۶۷ درصد از رخداد زایمان‌ها، زودرس بودند. بالاترین دقت پیش‌بینی (۰/۹۰) مربوط به الگوریتم XGBoost بود. در تفسیر خروجی مدل برای یک خانم باردار با مقادیر مشخص برای متغیرها، مهم‌ترین متغیر، با امتیاز اهمیت ۴۶ درصد، متغیر چندقلویی و پس از آن عوامل خطر زایمان، با امتیاز اهمیت ۴۱ درصد بود و متغیرهای دیگر نظیر بیماری اعصاب و روان، پره اکلامپسی و بیماری قلبی عروقی در رده‌های بعدی اهمیت برای این فرد خاص بودند.

نتیجه‌گیری: استفاده از روش یادگیری ماشینی تفسیرپذیر توانست وقوع زایمان زودرس را پیش‌بینی نماید. این روش می‌تواند توصیه‌های پیشگیرانه مختص هر زن باردار را، مبتنی بر عوامل خطر و با هدف جلوگیری از زایمان زودرس ارائه کند.

مقدمه

سازمان بهداشت جهانی[۱] (WHO) زایمان زودرس[۲] را هر زایمانی قبل از ۳۷ هفته کامل بارداری یا کمتر از ۲۵۹ روز از اولین روز آخرین قاعدگی[۳] خانم (LMP) تعریف می‌کند (۱). براورد انجام شده در سال ۲۰۱۴ نشان می‌دهد که در دنیا ۱۴/۸ میلیون نوزاد نارس متولد شده‌اند. نرخ تولد زودرس ۱۰/۶ درصد در سراسر جهان بوده که از ۸.۷ درصد تا ۱۳/۴ درصد نوزادان متولد شده، در مناطق مختلف، متغیر گزارش شده است (۲).

زایمان زودرس (PTB) یک نگرانی جدی در حوزه‌ی سلامت عمومی است که بر خانواده‌ها و جامعه تأثیر منفی می‌گذارد (۳). این یکی از علل اصلی مرگ و میر و معلولیت نوزادان در سراسر جهان و همچنین دومین علت عمده فوت کودکان زیر پنج سال است (۴). شناسایی زنان باردار که در معرض خطر زایمان زودرس هستند و همچنین شناسایی عوامل خطر موثر در زایمان زودرس در این زنان، می‌تواند به ارائه دهندگان مراقبت‌های بهداشتی کمک کند تا مداخلات مناسب را برای

[1]World Health Organization
[2]Preterm Birth
[3]Period Last Menstrual

[Downloaded from irje.tums.ac.ir on 2026-03-24]
[DOI: 10.18502/ijre.v20i1.17622]

جلوگیری از زایمان زودرس و بهبود نتایج برای مادران و نوزادان انجام دهند. مدل‌های یادگیری ماشینی در سال‌های اخیر برای پیش‌بینی زایمان زودرس با استفاده از عوامل مختلف وابسته به مادر و جنین مورد استفاده قرار گرفته‌اند (۷-۵). با این حال، ماهیت جعبه سیاه برخی از این مدل‌ها، از جمله یادگیری عمیق[1] و مدل‌های یادگیری ترکیبی[2] می‌تواند کاربرد بالینی و قابلیت تفسیر آنها را محدود کند. برای پرداختن به این موضوع، مدل‌های یادگیری ماشینی قابل تفسیر، توسعه داده شده‌اند که بینشی در مورد مکانیسم‌های اساسی پیش‌بینی‌ها ارائه می‌دهند. یکی از این روش‌ها، توضیحات محلی تفسیرپذیرِ مستقل از مدل (LIME) است که توضیحات قابل تفسیری را برای پیش‌بینی‌های فردی تولید می‌کند (۸, ۹). در مطالعه‌ای که از مدل یادگیری ماشینی توضیح‌پذیر مبتنی بر LIME برای پیش‌بینی سطح شدت بیماران مبتلا به کووید۱۹ استفاده شد، از الگوریتم‌های یادگیری ماشینی شبکه عصبی مصنوعی و درخت‌های تصمیم استفاده گردید و برای تفسیرپذیری مدل پیش‌بینی ارائه شده، از روش مدل آگنوستیک محلی یعنی LIME استفاده گردید (۱۰). مطالعه دیگری که برای مقایسه توضیحات ارائه شده توسط روش‌های هوش مصنوعی قابل توضیح در تجزیه و تحلیل پرونده الکترونیک سلامت طراحی شده بود، برای تفسیر مدل پیش‌بینی مرگ و میر در سرطان ریه از روش مدل آگنوستیکِ LIME استفاده شد (۱۱). برای پیش‌بینی زایمان زودرس در خانم‌هایی که سرکلاژ دهانه رحم داشتند، از چهار مدل درخت تصمیم، جنگل تصادفی، کا-نزدیکترین همسایه و شبکه‌ی عصبی مصنوعی استفاده شد و در آن برای رفع عدم توازن توزیع کلاس‌ها، از روش بیش‌نمونه‌گیری اقلیت مصنوعی[3] استفاده گردید (۱۲). در مطالعه‌ای دیگر، برای پیش‌بینی زایمان زودرس خودبه‌خودی در سه ماهه‌ی اول، مشخصات مادر و سابقه پزشکی او، در ابزار آنلاین بنیاد پزشکی جنین [4] وارد گردید. تحلیل رگرسیون لجستیک چند متغیره نشان داد، سیگار کشیدن و سابقه زایمان زودرس به طور معناداری با زایمان زودرس ناگهانی در ارتباط است (۱۳). استفاده از سیستم جامع هوشمند تلفن همراهِ توسعه داده شده، برای دریافت مراقبت‌های بهتر به خانم‌هایی که در معرض خطر عوارض بارداری هستند مورد مطالعه قرار گرفت و استفاده شد تا با پیش‌بینی زودهنگام خطر زایمان زودرس، سلامت مادر و جنین را به طور قابل توجهی بهبود بخشد (۱۴). در مطالعه‌ای دیگر، دو روش رگرسیون لجستیک و ماشین بردار پشتیبان[5] در پیش‌بینی وقوع زایمان زودرس استفاده شد و نتایج نشان داد سوابق پزشکی برای پیش‌بینی وقوع زایمان زودرس قابل استفاده بود و عمکرد مدل با استفاده از معیارهای دقت، حساسیت و ویژگی مورد ارزیابی قرار گرفت (۱۵).

اما هدف مطالعه حاضر، توسعه مدل‌های یادگیری ماشینی با استفاده از مجموعه داده‌های سامانه مادران و نوزادان ایرانی (IMaN) برای پیش‌بینی زایمان زودرس است. همچنین از LIME برای تفسیر مدل‌های موسوم به جعبه سیاه و شناسایی عوامل خطر زایمان زودرس در پیش‌بینی‌های فردی استفاده می‌شود. این رویکرد به‌طور بالقوه می‌تواند دقت پیش‌بینی زایمان زودرس را بهبود بخشد و بینش و شهود بیشتری را در مورد فرایند پیش‌بینی به پزشکان و متخصصین این حوزه ارائه نماید.

روش کار

داده‌ها: در این مطالعه که به‌صورت مقطعی انجام شد، از مجموعه داده شبکه مادران و نوزادان ایران (IMaN) استفاده شد که شامل ۱۴۹۳۵۰ مورد تولد از شهر تهران در سال ۱۳۹۹ بود. برای دسترسی به اطلاعات مفید و ضروری در مورد شاخص‌های سلامت مادر و نوزاد و همچنین استفاده برای برنامه‌ریزی و سیاست‌گذاری در نظام سلامت، «دفتر سلامت نوزادان» و «اداره سلامت جمعیت، خانواده و مدارس» با همکاری سایر دفاتر مرتبط و اکثر کارشناسان سلامت مادر و نوزادان در دانشگاه‌ها، پرسشنامه‌ها و نرم‌افزارهای متنوعی را برای ثبت وقایع تولد در بیمارستان‌ها طراحی کردند. IMaN یکی از بزرگترین منابع اطلاعاتی برای پایش سلامت مادران و

[1] Deep Learning
[2] Ensemble Learning Methods
[3] Over-sampling Technique Synthetic Minority (SMOTE)
[4] Fetal Medicine Foundation (FMF) The
[5] Support Vector Machine (SVM)

[Downloaded from irje.tums.ac.ir on 2026-03-24]
[DOI: 10.18502/ijre.v20i1.17622]

نوزادان است. این سامانه تقریباً تمام موارد تولد (زنده و مرده)، اطلاعات جمعیتی و اطلاعات مربوط به سلامت مادران و نوزادان را به صورت الکترونیکی در داخل و خارج از بیمارستان‌های سراسر کشور ثبت می‌کند. در این شبکه، اطلاعاتی مانند مشخصات جمعیتی مادر نوزاد، نوع زایمان، سابقه‌ی سقط، هفته‌های بارداری، سابقه پزشکی مادر، عوامل خطر بارداری و زایمان، تاریخ تولد و ویژگی‌های تولد نوزاد (مانند وزن بدن) و متغیرهای مرتبط دیگر ثبت می‌شود (۱۶). طبق برنامه اعتباربخشی بیمارستانی اجرا شده در ایران، روایی و پایایی پایگاه داده IMaN، با بررسی تصادفی سوابق پزشکی به صورت دوره‌ای ارزیابی می‌شود (۱۷). لازم به ذکر است این مطالعه با کد اخلاق IR.USWR.REC.۱۴۰۱.۰۴۹ برای اجرا، پذیرفته شد و در آن، تمام مادرانی که سن حاملگی برابر یا بیشتر از ۲۸ هفته داشتند وارد مطالعه شدند و تولدهایی که به‌صورت مرده به دنیا آمده بودند از مطالعه حذف شدند. برای تحلیل داده‌ها از زبان برنامه‌نویسی پایتون نسخه ۳.۱۰.۰ استفاده شد. در ادامه برای تبیین بیش‌تر، توضیحات مختصری درباره پیش‌پردازش، الگوریتم‌های یادگیری ماشینی، عملکرد مدل‌های یادگیری ماشینی و تفسیر این مدل‌ها آورده شده است.

پیش‌پردازش داده‌ها: در ابتدا، متغیرهایی که نسبت مقدار گمشده آنها بیش از ۲۰٪ بود، حذف شدند. بعد از پالایش اولیه داده‌ها، جانهی به‌صورت میانگین، برای متغیر سن مادر در جهت از دست ندادن اطلاعات ارزشمند، به‌طور خاص انجام گردید. همچنین به دلیل متوازن نبودن توزیع متغیر پاسخ، از روش بیش نمونه‌برداری اقلیت مصنوعی (SMOTE[1]) استفاده گردید (۱۸). برای اطمینان از کیفیت داده‌ها، مقادیر پرت بررسی شده و با استفاده از روش دامنه میان چارکی (IQR) حذف شدند. علاوه براین، برای انتخاب متغیرهای مهم و موثر برای پیش‌بینی زایمان زودرس از روش حذف ویژگی بازگشتی با اعتبار متقابل (RFECV[2]) استفاده شد (۱۹). همچنین، برای تعیین معنی‌داری ارتباط متغیرهای پیش‌بینی‌کننده و متغیر وقوع زایمان زودرس، از آزمون کای-دو با سطح معنی‌داری ۰/۰۵ استفاده گردید. برای اجرای فرایند آموزش و همچنین آزمایش، داده‌ها به نسبت ۷۰٪ برای آموزش و ۳۰٪ برای آزمایش تقسیم شدند.

الگوریتم‌های یادگیری ماشینی: برای پیش‌بینی وقوع زایمان زودرس از مدل‌های یادگیری ماشینی ۱- XGBoost ۲- Random Forest ۳- ANN-MLP استفاده گردید. این مدل‌ها به نسبت گفته شده در قسمت پیش‌پردازش مورد آموزش و آزمون قرار داده شدند تا مشخص گردد کدام مدل برای پیش‌بینی زایمان زودرس بهترین عملکرد را دارند. از روش جستجوی شبکه‌ای[3] برای تنظیم کردن هایپرپارامترها برای بهبود عملکرد مدل استفاده گردید.

XGBoost مبتنی بر تقویت گرادیان است، روشی است که چندین الگوریتم ضعیف (معمولاً درختان تصمیم) را برای ساخت یک مدل ترکیبی[4] کارا تلفیق می‌کند. الگوریتم با اضافه کردن مکرر درخت‌های تصمیم در حالی که برای باقیمانده‌های تکرارهای قبلی بهینه می‌شود، تابع زیان را به‌کمینه می‌کند (۲۰).

شبکه عصبی مصنوعی، یک مدل یادگیری ماشینی است که از سه لایه، شامل گره‌های متصل به هم، تشکیل شده است: یک لایه ورودی، یک لایه پنهان و یک لایه خروجی. طی فرایند آموزش، شبکه عصبی مصنوعی با تنظیم وزن‌ها درصدد یادگیری نحوه پیش‌بینی خروجی می‌باشد. هر متغیر ورودی به عنوان یک گره ظاهر می‌شود. لایه پنهان شامل چندین گره است که در مرحله تنظیم مدل تعیین شده‌اند. لایه خروجی شامل چندین گره است. تعداد گره‌های لایه خروجی برابر با تعداد کلاس‌های پیش‌بینی است. یک پیوند وزنی بین این لایه‌ها وجود دارد و لایه پنهان، مجموع حاصل‌ضرب مقدار وزن مرتبط و متغیر ورودی به اضافه اریبی را دریافت می‌کند. این مقدار برای تعیین کلاس، وارد تابع فعال‌سازی می‌شود (۲۱).

جنگل تصادفی، یک روش ترکیبی[5] است که تعمیم‌پذیری را با ترکیب چند مدل یادگیری بهبود می‌بخشد (۱۷). منطق پشت یک مدل ترکیبی این است که مجموعه‌ای از مدل‌های ساده ممکن است، عملکرد بهتری نسبت به مدل‌هایی که بیش‌برازش شده‌اند، ارائه دهند. جنگل تصادفی ترکیب‌هایی از درختان

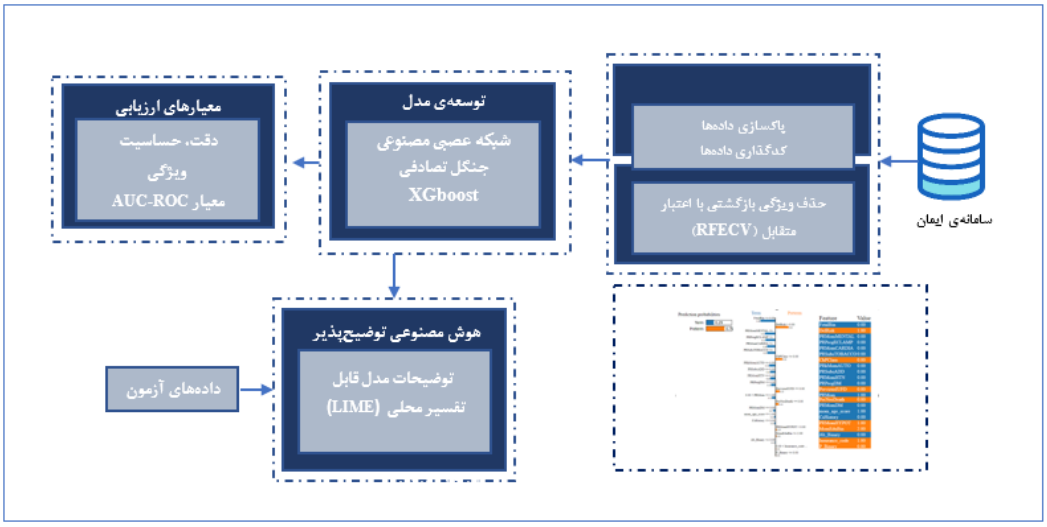
[1] Synthetic Minority Oversampling Technique
[2] Recursive Feature Elimination and Cross-Validation Selection
[3] Grid Search Method
[4] Ensemble Model
[5] Ensemble

[Downloaded from irje.tums.ac.ir on 2026-03-24]
[DOI: 10.18502/ijre.v20i1.17622]

تصمیم را اجرا می‌کند. هر درخت، فرآیند تصمیم‌گیری را توصیف می‌کند، به‌طوری‌که فرایند تصمیم‌گیری یک شاخه، با مقایسه مقدار یک ویژگی در هر گره تصمیم با یک آستانه صورت می‌پذیرد. ساختار درخت تصمیم و آستانه‌ها[1] در مرحله یادگیری تعیین می‌شوند. جنگل تصادفی، چندین درخت تصمیم را می‌سازد که بر روی نمونه‌های آموزشی، آموزش داده شده‌اند و پیش‌بینی‌ها را برای تولید یک خروجی مجموعه ترکیب می‌کند (۲۱).

عملکرد مدل‌های یادگیری ماشینی: مدل‌ها بر اساس معیارهای دقت، حساسیت، ویژگی و سطح زیر منحنی موسوم به راک[2] (AUC)، ارزیابی شدند. دقت به توانایی مدل در پیش‌بینی‌های درست در کل موارد بررسی شده اشاره دارد. حساسیت به توانایی مدل در تشخیص مثبت درست اشاره دارد. ویژگی به توانایی آزمون برای تشخیص منفی درست مربوط می‌شود (۴). AUC نیز یک معیار جامع شامل حساسیت و ویژگی مدل است.

تفسیر مدل یادگیری ماشینی: تفسیرپذیری به قابلیت یک الگوریتم یادگیری ماشینی برای ارائه توضیحات برای پیش‌بینی‌ها یا تصمیمات یا خروجی‌های آن اشاره دارد. تفسیرپذیری برای افزایش اعتماد و اطمینان به نتایج مدل، دارای اهمیت است(۲۲). راه‌های مختلفی برای دستیابی به تفسیرپذیری وجود دارد، مانند استفاده از مدل‌های ساده‌تر یا استفاده از تکنیک‌های تفسیرپذیری در مدل‌های جعبه سیاه. توضیحات محلی تفسیرپذیرِ مستقل از مدل (LIME)، یک تکنیک محبوب برای تفسیر پیش‌بینی‌های هر مدل یادگیری ماشینی است (۱۰). تفسیرپذیری به ویژه در کاربردهای پرمخاطره مانند مراقبت‌های بهداشتی و پزشکی بسیار حائز اهمیت می‌باشد(۲۳). در شکل ۱ چارچوب مفهومی مدل یادگیری ماشینی تفسیرپذیر برای پیش‌بینی وقوع زایمان زودرس آمده است.



شکل شماره ۱- چارچوب مفهومی مدل یادگیری ماشینی تفسیرپذیر برای پیش‌بینی وقوع زایمان زودرس

یافته‌ها

در این مطالعه، زایمان زنان باردار به‌طور کلی به دو گروه زایمان به‌موقع و زودرس گروه‌بندی گردید. درصد وقوع زایمان زودرس در شهر تهران ۸/۶۷ درصد، و درصد وقوع زایمان به‌موقع ۹۱/۳ درصد بود. مشخصات فردی و بالینی زنان باردار برای این دو گروه و رابطه تک‌متغیره آن‌ها در جدول ۱ گزارش گردید. بیش‌ترین تعداد بارداری گزارش شده، ۷۵۹۲۳ مورد بود که در سن ۳۷-۳۸ هفتگی بوده است. توزیع سن بارداری خانم‌های باردار مبتنی‌بر تعداد هفته بارداری در شکل ۲ آمده است.

[1]Threshold

[2]Receiver Operating Characteristic (ROC)

[Downloaded from irje.tums.ac.ir on 2026-03-24]

[DOI: 10.18502/ijre.v20i1.17622]

جدول شماره ۱- رابطه تک متغیره بین نوع زایمان و متغیرهای جمعیتی، مادری و نوزادی

متغیر	نوع زایمان	سطوح	تعداد	درصد	مقدار احتمال*
سن مادر	به موقع	کمتر از ۱۸	۱۹۳۰	۱/۲۹	<۰/۰۰۱
		بین ۱۹ و ۳۰	۶۳۹۸۲	۴۲/۹۰	
		بین ۳۱ و ۴۰	۶۵۷۵۳	۴۴/۰۸	
		بالاتر از ۴۰	۴۵۵۶	۳/۰۵	
		گمشده	۰	۰/۰	
	زودرس	کمتر از ۱۸	۱۶۵	۰/۱۱	
		بین ۱۹ و ۳۰	۵۰۹۵	۳/۴۲	
		بین ۳۱ و ۴۰	۶۸۸۸	۴/۶۲	
		بالاتر از ۴۰	۷۸۸	۰/۵۳	
		گمشده	۰	۰/۰	
سطح تحصیلات	به موقع	کم سواد	۲۳۷۱۷	۱۵/۹۰	<۰/۰۰۱
		دبیرستان و کاردانی	۶۱۹۴۹	۴۱/۵۳	
		لیسانس و بالاتر			
		گمشده	۵۰۵۵۵	۳۳/۸۹	
	زودرس	کم سواد	۲۴۷۸	۱/۶۶	
		دبیرستان و کاردانی	۵۹۸۳	۴/۰۱	
		لیسانس و بالاتر			
		گمشده	۴۴۷۵	۳/۰	
بیمه مادر	به موقع	دارد	۸۹۹۲۷	۶۰/۲۹	<۰/۰۰۱
		ندارد	۴۶۲۹۴	۳۱/۰۴	
		گمشده	۰	۰/۰	
	زودرس	دارد	۹۱۶۴	۶/۱۴	
		ندارد	۳۷۷۲	۲/۵۳	
		گمشده	۰	۰/۰	
عوامل خطر زایمان	به موقع	خیر	۱۳۰۷۶۸	۸۷/۶۷	<۰/۰۰۱
		بله	۵۴۵۳	۳/۶۶	
		گمشده	۰	۰/۰	
	زودرس	خیر	۸۷۵۱	۵/۸۷	
		بله	۴۱۸۵	۲/۸۱	
		گمشده	۰	۰/۰	
دیابت بارداری	به موقع	خیر	۱۲۷۲۷۳	۸۵/۳۳	<۰/۰۰۱
		بله	۸۹۴۸	۶/۰	
		گمشده	۰	۰/۰	
	زودرس	دارد	۱۵۹۷	۱/۰۷	
		ندارد	۱۱۳۳۹	۷/۶۰	
		گمشده	۰	۰/۰	
دیابت پیش از بارداری	به موقع	خیر	۱۳۵۵۱۳	۹۰/۸۵	<۰/۰۰۱
		بله	۷۰۸	۰/۴۷	
		گمشده	۰	۰/۰	

[Downloaded from irje.tums.ac.ir on 2026-03-24]

[DOI: 10.18502/ijre.v20i1.17622]

		بله	۱۴۱	۰/۰۹	
	زودرس	خیر	۱۲۷۹۵	۸/۵۸	
		گمشده	۰	۰/۰	
	به موقع	دارد	۲۲۹۰	۱/۵۴	<۰/۰۰۱
		ندارد	۱۳۳۹۳۱	۸۹/۷۹	
		گمشده	۰	۰/۰	
فشار خون مزمن		دارد	۶۲۷	۰/۴۲	
	زودرس	ندارد	۱۲۳۰۹	۸/۲۵	
		گمشده	۰	۰/۰	
	به موقع	دارد	۱۰۷۳	۰/۷۲	<۰/۰۰۱
		ندارد	۱۳۵۱۴۸	۹۰/۶۱	
		گمشده	۰	۰/۰	
بیماری قلبی عروقی		دارد	۱۴۹	۰/۱۰	
	زودرس	ندارد	۱۲۷۸۷	۸/۵۷	
		گمشده	۰	۰/۰	
	به موقع	دارد	۳۷۴	۰/۲۵	<۰/۰۰۱
		ندارد	۱۳۵۸۴۷	۹۱/۰۸	
		گمشده	۰	۰/۰	
بیماری خود ایمنی		دارد	۷۶	۰/۰۵	
	زودرس	ندارد	۱۲۸۶۰	۸/۶۲	
		گمشده	۰	۰/۰	
	به موقع	بله	۲۳۲۷۲	۱۵/۶۰	<۰/۰۰۱
		خیر	۱۱۲۹۴۹	۷۵/۷۲	
		گمشده	۰	۰/۰	
اختلالات تیروئید		بله	۲۳۶۶	۱/۵۹	
	زودرس	خیر	۱۰۵۷۰	۷/۰۹	
		گمشده	۰	۰/۰	
	به موقع	دارد	۲۲۵۸	۱/۵۱	<۰/۰۰۱
		ندارد	۱۳۳۹۶۳	۸۹/۸۱	
		گمشده	۰	۰/۰	
پره اکلامپسی		دارد	۱۲۳۳	۰/۸۳	
	زودرس	ندارد	۱۱۷۰۳	۷/۸۵	
		گمشده	۰	۰/۰	
	به موقع	دارد	۷۱	۰/۰۵	<۰/۰۰۱
		ندارد	۱۳۶۱۵۰	۹۱/۲۸	
		گمشده	۰	۰/۰	
سیگار و دخانیات		دارد	۳۸	۰/۰۳	
	زودرس	ندارد	۱۲۸۹۸	۸/۶۵	
		گمشده	۰	۰/۰	
	به موقع	دارد	۱۸۸	۰/۱۳	<۰/۰۰۱
		ندارد	۱۳۶۰۳۳	۹۱/۲۰	
		گمشده	۰	۰/۰	
اعتیاد مادر به مواد مخدر		دارد	۸۷	۰/۰۶	
	زودرس	ندارد	۱۲۸۴۹	۸/۶۱	

[Downloaded from irje.tums.ac.ir on 2026-03-24]

[DOI: 10.18502/ijre.v20i1.17622]

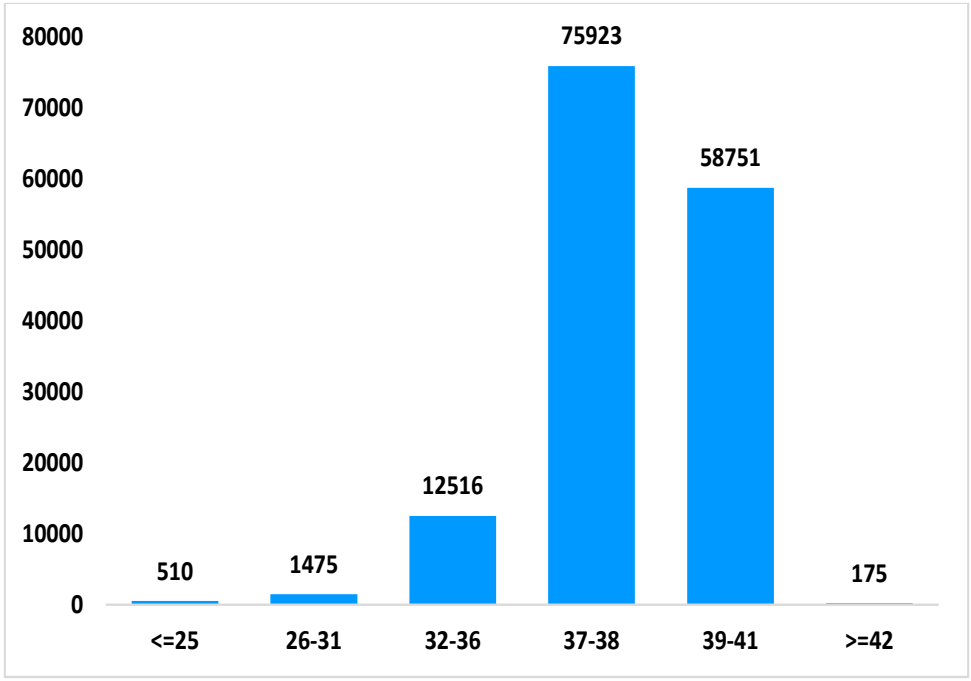
		گمشده	۰	۰/۰	
بیماری اعصاب و روان	به موقع	دارد	۸۰۵	۰/۵۴	<۰/۰۰۱
		ندارد	۱۳۵۴۱۶	۹۰/۷۹	
		گمشده	۰	۰/۰	
	زودرس	دارد	۱۱۰	۰/۰۷	
		ندارد	۱۲۸۲۶	۸/۶۰	
		گمشده	۰	۰/۰	
چند قلویی	به موقع	بله	۲۰۴۹	۱/۳۷	<۰/۰۰۱
		خیر	۱۳۴۱۷۲	۸۹/۹۵	
		گمشده	۰	۰/۰	
	زودرس	بله	۳۲۸۱	۲/۲۰	
		خیر	۹۶۵۵	۶/۴۷	
		گمشده	۰	۰/۰	
مرگ نوزاد قبلی	به موقع	بله	۱۴۴۸	۰/۹۷	<۰/۰۰۱
		خیر	۱۳۴۷۷۳	۹۰/۳۶	
		گمشده	۰	۰/۰	
	زودرس	بله	۲۸۲	۰/۱۹	
		خیر	۱۲۶۵۴	۸/۴۸	
		گمشده	۰	۰/۰	
سابقه مرده زایی	به موقع	خیر	۹۶۹	۰/۶۵	<۰/۰۰۱
		بله	۱۳۵۲۵۲	۹۰/۶۸	
		گمشده	۰	۰/۰	
	زودرس	خیر	۲۱۱	۰/۱۴	
		بله	۱۲۷۲۵	۸/۵۳	
		گمشده	۰	۰/۰	
سابقه زایمان قبلی	به موقع	بله	۷۵۵۸۶	۵۰/۶۸	<۰/۰۰۱
		خیر	۶۰۶۳۵	۴۰/۶۵	
		گمشده	۰	۰/۰	
	زودرس	بله	۶۸۹۷	۴/۶۲	
		خیر	۶۰۳۹	۴/۰۵	
		گمشده	۰	۰/۰	
سابقه سزارین	به موقع	بله	۲۲۰۲۹	۱۴/۷۷	<۰/۰۰۱
		خیر	۱۱۴۱۹۲	۷۶/۵۶	
		گمشده	۰	۰/۰	
	زودرس	بله	۲۴۳۶	۱/۶۳	
		خیر	۱۰۵۰۰	۷/۰۴	
		گمشده	۰	۰/۰	
سابقه سقط جنین	به موقع	بله	۳۱۴۶۹	۲۱/۱۰	<۰/۰۰۱
		خیر	۱۰۴۷۵۲	۷۰/۲۳	
		گمشده	۰	۰/۰	
	زودرس	بله	۳۵۳۰	۲/۳۷	
		خیر	۹۴۰۶	۶/۳۱	
		گمشده	۰	۰/۰	

[Downloaded from irje.tums.ac.ir on 2026-03-24]

[DOI: 10.18502/ijre.v20i1.17622]

	به موقع	بله	۶۸۷۲	۴/۶۱	<۰/۰۰۱
		خیر	۱۲۹۳۴۹	۸۶/۷۲	
		گمشده	۰	۰/۰	
شرکت در کلاس آمادگی زایمان		خیر	۳۸۴	۰/۲۶	
	زودرس	بله	۱۲۵۵۲	۸/۴۲	
		گمشده	۰	۰/۰	

* آزمون کای دو



شکل شماره ۲- توزیع سن بارداری خانم‌های باردار بر اساس تعداد هفته بارداری

همچنین عملکرد مدل‌های یادگیری ماشینی توسعه داده شده در جدول ۲ بر اساس معیارهای گفته شده آورده شده است. بر این اساس مدل XGboost بالاترین دقت را نسبت به سایر مدل‌ها به‌دست آورد.

جدول شماره ۲- معیارهای ارزیابی عملکرد مدل‌های یادگیری ماشینی

مدل	دقت	حساسیت	ویژگی	AUC
شبکه‌ی عصبی پرسپترون چند لایه	۰/۸۵۴۶	۰/۵۷۳۰	۰/۸۸۱۸	۰/۷۵۲۳
جنگل تصادفی (RF)	۰/۸۷۴۳	۰/۵۵۷۲	۹۴۹	۰/۷۶۷۲
XGboost	۰/۸۹۵۵	۰/۵۳۴۷	۹۳۳	۰/۷۶۶۱

شکل ۳ نمودار ROC مربوط به عمکرد سه مدل را به‌طور هم-زمان نمایش می‌دهد. این معیار حاکی از مشابه بودن عملکرد مدل‌ها در پیش‌بینی وقوع زایمان زودرس را دارند. برای نمایش نحوه استفاده از الگوریتم LIME بر روی مدل شبکه عصبی مصنوعی آموزش داده شده در این پژوهش، یک زن باردار از مجموعه داده آزمون، که زایمان زودرس را تجربه کرده است، به‌عنوان نمونه مورد بررسی قرار داده شد. مشخصات این زن باردار در جدول ۳ تنظیم و گزارش شده است.

[Downloaded from irje.tums.ac.ir on 2026-03-24]

[DOI: 10.18502/ijre.v20i1.17622]

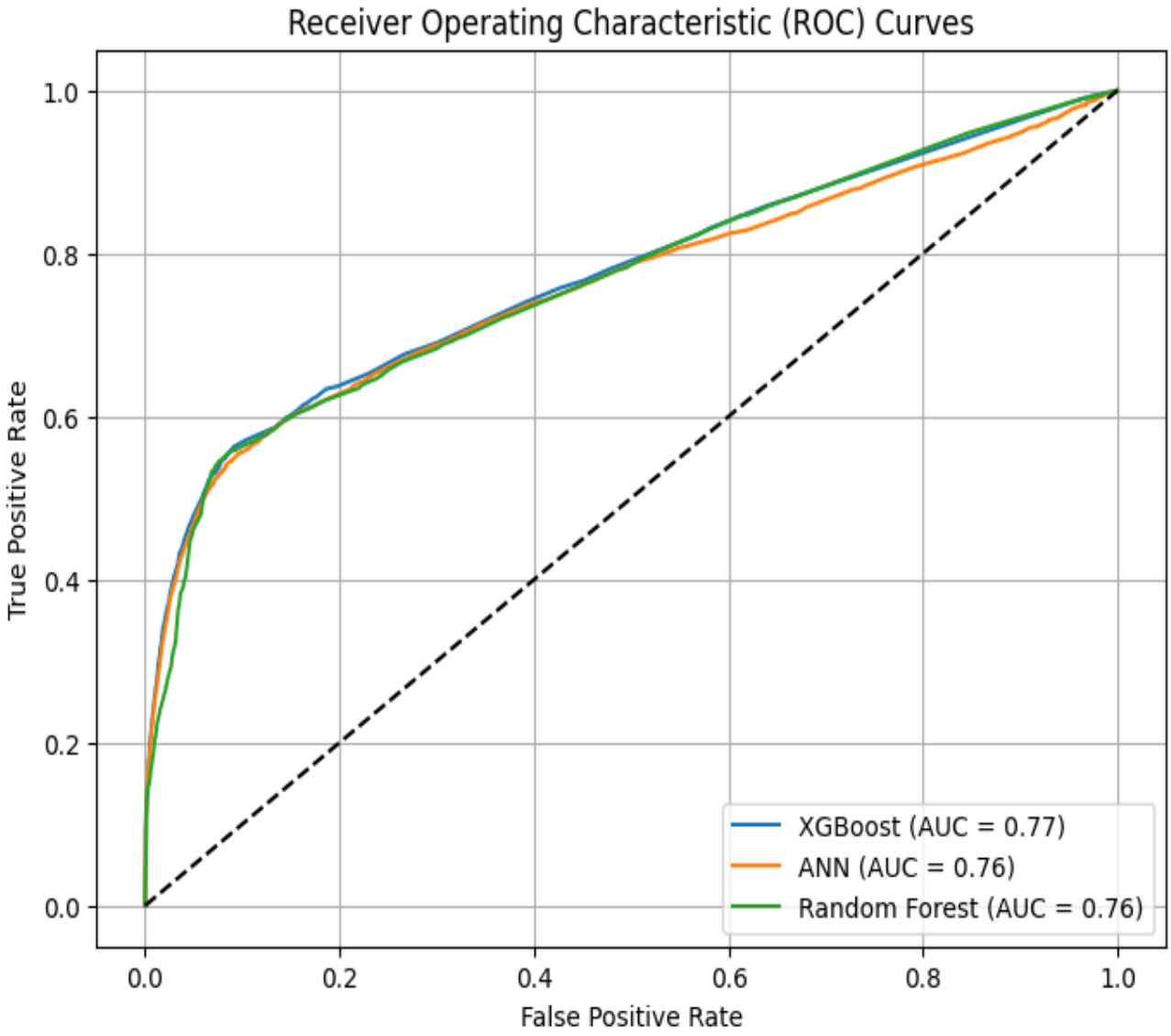
جدول شماره ۳- مشخصات زن باردار با تجربه زایمان زودرس

نام وارد شده به رایانه	نام متغیر (ویژگی)	مقدار	کد مربوط به مقدار
mom_age_score	سن مادر	۳۵ سال	۲
MomEduBin	سطح تحصیلات	کاردانی	۱
Insurance_code	بیمه‌ی مادر	بلی	۱
DelRisk	عوامل خطر زایمان	خیر	۰
PRMomHTN	فشار خون مزمن	بله	۱
PRMomCARDIA	بیماری قلبی و عروقی	خیر	۰
PRkMomAUTO	بیماری خود ایمنی	خیر	۰
PRMomDM	دیابت پیش از بارداری	خیر	۰
PRPregDM	دیابت بارداری	خیر	۰
PRMomHYPOT	اختلالات تیروئید	خیر	۰
PRMomMENTAL	بیماری اعصاب و روان	خیر	۰
PRPregECLAMP	پره‌اکلامپسی	بله	۱
PRSubsTOBACCO	سیگار و دخانیات	بله	۱
PRSubsADD	اعتیاد مادر به مواد مخدر	خیر	۰
Ab_Binary	سابقه‌ی سقط جنین	خیر	۰
CsHistory	سابقه‌ی سزارین	خیر	۰
PreviousIUFD	سابقه‌ی مرده زایی	خیر	۰
PreNeoDeath	مرگ نوزاد قبلی	خیر	۰
P_Binary	سابقه‌ی زایمان قبلی	خیر	۱
FetalBin	چند قلویی	خیر	۰
ChPClass	شرکت در کلاس‌های آمادگی برای زایمان	خیر	۰

این زن باردار که زایمان زودرس را تجربه کرده است، ۳۵ ساله با مدرک تحصیلی کاردانی، دارای بیمه درمانی، با فشار خون مزمن و پره‌اکلامپسی مثبت و دارای سابقه زایمان قبلی می‌باشد و مابقی متغیرها برای او خیر می‌باشد. مدل شبکه عصبی مصنوعی آموزش داده شده، وقوع زایمان زودرس را برای او پیش‌بینی کرده است. سابقه پزشکی، اطلاعات جمعیت‌شناختی و اطلاعات دیگر در رابطه با این زن باردار در شکل ۴-الف نیز نمایش داده شده است. تفسیر الگوریتم LIME بر روی این مدل، مطابق شکل ۴-ب است. نوارهای نارنجی رنگ، متغیرهایی را نشان می‌دهد که در پیش‌بینی وقوع زایمان زودرس نقش مثبت داشته‌اند. متغیرهای پره‌اکلامپسی، فشار خون مزمن و شرکت در کلاس‌های آمادگی برای زایمان (که این زن باردار شرکت نکرده است) با امتیازهای اهمیت نسبی ۰/۳۳، ۰/۳۰، و ۰/۲۲ نقش پررنگی را در پیش‌بینی وقوع زایمان زودرس در این زن باردار داشت. نوارهای آبی رنگ، متغیرهایی را نشان می‌دهد که در پیش‌بینی وقوع زایمان زودرس، نقش منفی داشت. متغیرهای چندقلویی، عوامل خطر زایمان (متغیر ترکیبی)، استعمال دخانیات و دیابت پیش از بارداری (که مقدار همه این متغیرها برای این زن باردار خیر بود) با امتیازهای اهمیت نسبی ۰/۴۴، ۰/۳۹، ۰/۲۳ و ۰/۲۲ نقش منفی در پیش‌بینی زایمان زودرس داشت.

[Downloaded from irje.tums.ac.ir on 2026-03-24]

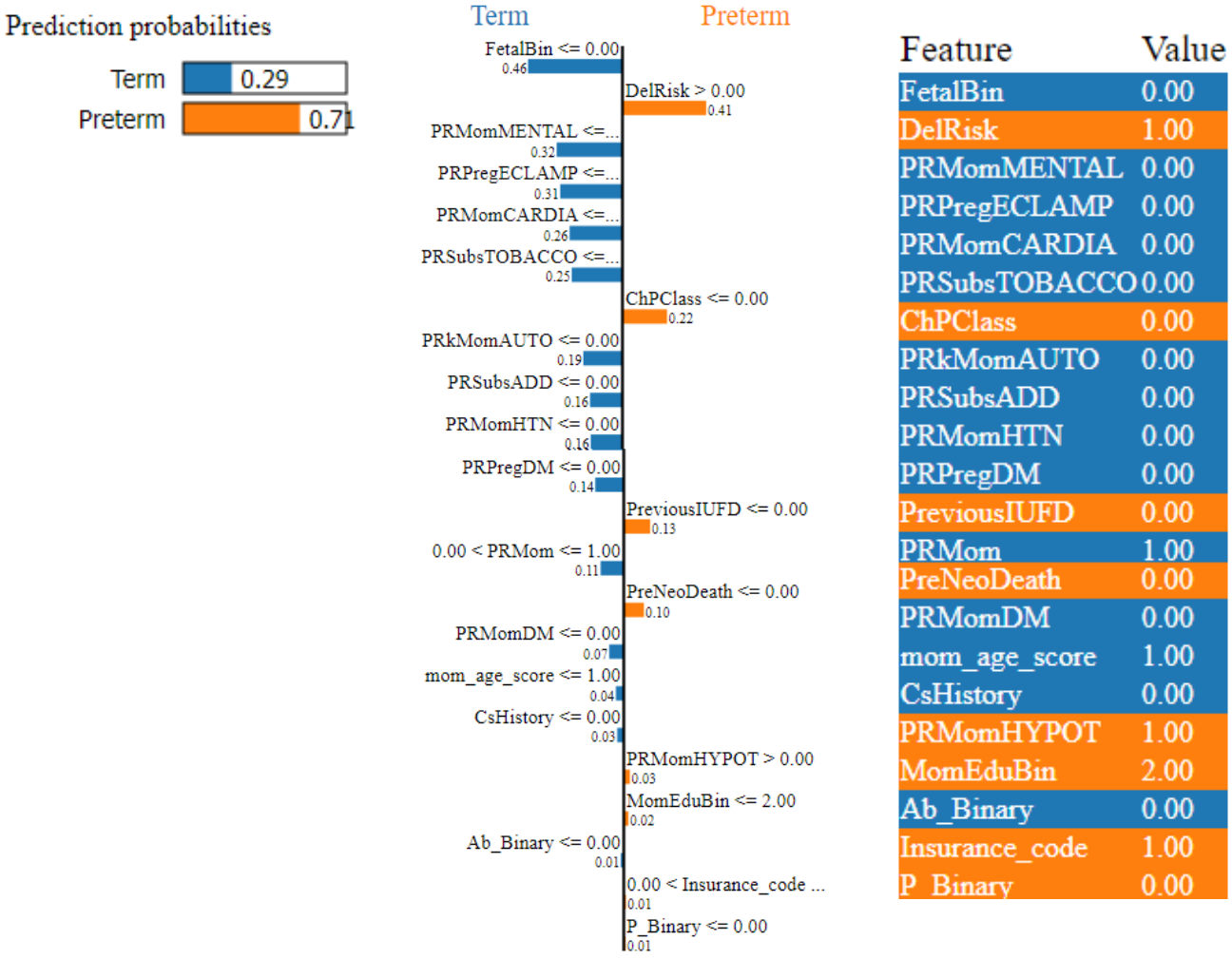
[DOI: 10.18502/ijre.v20i1.17622]



شکل شماره ۳- نمودار راک مربوط به عمکرد سه مدل یادگیری ماشینی توسعه داده شده

[Downloaded from irje.tums.ac.ir on 2026-03-24]

[DOI: 10.18502/ijre.v20i1.17622]



شکل شماره ۴- مقادیر مشخصات زن باردار (الف)، تفاسیر مربوطه (ب) و احتمال پیش بینی (ج)

بحث

بالاترین دقت پیش‌بینی مربوط به الگوریتم XGBoost بود. سپس کاربرد روش توضیح دهنده LIME روی مدل جعبه سیاه شبکه عصبی مصنوعی با انتخاب یک خانم باردار نشان داده شد. بدین وسیله سعی در شفاف‌سازی چرایی پیش‌بینی مورد نظر گردید. با پیش‌آگهی در مورد خطرات تهدید کننده برای خانم-های باردار می‌توان در جهت کاهش میزان وقوع زایمان زودرس قدم برداشت. به‌عنوان نمونه، برای خانم باردار انتخاب شده، با کاستن از عوامل خطر زایمان و یا اقدام به شرکت در کلاس‌های آمادگی قبل از زایمان، می‌توان از وقوع زایمان زودرس جلوگیری کرد. توضیحاتی که این روش تفسیرپذیری ارائه می‌کند، بسیار شهودی است. گفتنی آن‌که، برخلاف رگرسیون لجستیک که تفسیر عمومی را مبتنی‌بر پارامترهای براوردشده دراختیار می‌گذارد، استفاده از الگوریتم LIME تحت مدل یادگیری ماشینی آموزش داده شده، تفسیر شخصی‌سازی شده را ارائه می‌کند.

پیش‌بینی دقیق وقوع زایمان زودرس، موضوعی چالش‌برانگیز است. علاوه‌بر متغیرهایی که در پژوهش حاضر وارد مطالعه شده‌اند، متغیرهای دیگری مانند عوامل ژنتیکی و محیطی نیز وجود دارند که نقش پیش‌بین را برای زایمان زودرس بازی کند. مطالعه Raja و همکاران با بهره‌گیری از سه مدل یادگیری ماشینی رگرسیون لجستیک، درخت تصمیم و ماشین بردار پشتیبان[1] وقوع زایمان زودرس را پیش‌بینی کردند. ماشین بردار پشتیبان با دقت ۹۰/۹ درصد، بالاترین دقت را بین سه مدل

[1] Support Vector Machine (SVM)

[Downloaded from irje.tums.ac.ir on 2026-03-24]
[DOI: 10.18502/ijre.v20i1.17622]

آموزش داده شده داشته است (۶). در مطالعه‌ی دیگری که Sun و همکاران انجام دادند شش مدل یادگیری ماشینی جهت پیش‌بینی وقوع زایمان زودرس توسعه داد. نتایج نشان داد مدل درخت تصمیم با در نظر گرفتن معیار مساحت زیر منحنی راک و دقت بهتر از سایر مدل‌ها عمل کرده است (۵). Koivu و همکاران با استفاده از شبکه عصبی مصنوعی، رگرسیون لجستیک و درخت تصمیم تقویت گرادیان، مدل پیش‌بینی‌کننده را توسعه دادند. مدل‌های ترکیبی این الگوریتم‌ها نیز مورد آزمون قرار گرفت. بهترین مدل از نظر عملکرد، مساحت زیر منحنی راک به میزان ۰/۶۴ را به خود اختصاص داده است (۷). Rawashdeh و همکاران برای پیش‌بینی زایمان زودرس در خانم‌های دارای سرکلاژ دهانه رحم، از چهار مدل درخت تصمیم، جنگل تصادفی، کا-نزدیکترین همسایه و شبکه‌ی عصبی مصنوعی استفاده کردند و در آن برای رفع عدم توازن توزیع کلاس‌ها، از روش بیش‌نمونه‌گیری اقلیت مصنوعی استفاده کردند یافته‌ها نشان داد که طبقه‌بندی جنگل تصادفی از نظر میانگین و حساسیت بهترین نتایج را داشتند. در پایان چنین گزارش کردند که استفاده از مدل‌های محاسباتی، برای پیش‌بینی نیاز به سرکلاژ و بارداری قابل توسعه بوده و این مدل‌ها دارای حساسیت متوسط/بالا برای کاربرد بالینی هستند (۱۲). Damaso و همکاران، از تحلیل رگرسیون لجستیک چندمتغیره برای پیش‌بینی زایمان زودرس خودبه‌خودی در سه ماهه اول، استفاده کردند. نتایج نشان داد سیگار کشیدن و سابقه زایمان زودرس بدون زایمان ترم، به‌طور معناداری با زایمان زودرس ناگهانی در ارتباط است. در پایان چنین نتیجه‌گیری شد که ویژگی‌ها و سابقه مادر در سه ماهه اول می‌تواند به‌طور معنی‌داری وقوع زایمان خودبه‌خود را قبل از هفته ۳۴ بارداری پیش‌بینی کند (۱۳). Pushpalatha و همکاران از دو روش رگرسیون لجستیک و ماشین بردار پشتیبان در پیش‌بینی وقوع زایمان زودرس استفاده نمودند. نتایج این مطالعه نشان داد سوابق پزشکی برای پیش‌بینی وقوع زایمان زودرس قابل استفاده بود (۱۵). Grigorescu و همکاران با استفاده از شبکه‌های عصبی پیچشی[1] و همچنین بهره‌گیری از روش تفسیرسازی موسوم به انتشار ارتباط لایه‌ای[2] که تفسیری بصری از تصمیمات تحت شبکه عصبی را ارائه می‌دهد، مطالعه را برای طبقه‌بندی زایمان زودرس انجام داده‌اند. در این مطالعه، مدل یاد شده روی تصاویر weighted MRI-T2[3] آموزش داده شد. از طریق الگوریتم تفسیرسازی مذکور، نقشه‌های حرارتی[4] تولید شده، تصمیمات اتخاذ شده توسط مدل آموزش دیده (یک مدل طبقه‌بندی‌کننده) را به تصویر می‌کشد (۲۴).

این پژوهش دارای چندین نقطه قوت می‌باشد. نخست، تعداد بالا نمونه‌های مطالعه می‌باشد که به افزایش هرچه بیشتر قدرت پیش‌بینی مدل کمک می‌کند. دوم، وجود تعداد زیادی از متغیرهای پیشگو است. نه تنها متغیرهای جمعیت‌شناختی و وضعیت سلامت و سوابق بیماری مادر بلکه متغیرهایی در مورد شرایط بارداری و زایمان مثل شرکت در کلاس‌های آمادگی برای زایمان در این مطالعه گنجانده شده است. به‌عنوان سومین نقطه قوت می‌توان به استفاده از رویکرد مدل-آگنوستیک و به‌طور خاص روش LIME در تفسیر مدل‌های غیر شفاف اشاره کرد. این مورد به هرچه بیشتر ملموس کردن نتایج مطالعه کمک می‌کند. با این وجود، به‌دلیل نبودِ متغیر شاخص توده‌ی بدنی[5] مادر؛ این مطالعه نتوانست از آن برای پیش‌بینی وقوع زایمان زودرس، به‌عنوان یک پیشگوی مهم، استفاده کند.

نتیجه‌گیری

این مطالعه، قدم‌های نخست را در جهت استفاده از مدل‌های یادگیری ماشینی در پیش‌بینی دقیق وقوع زایمان زودرس و همچنین تفسیر شخصی‌سازی شده این مدل‌ها برداشته است. نتایج مطالعه، امکان‌پذیر بودن پیش‌بینی زایمان زودرس را با استفاده از مدل‌های یادگیری ماشینی و همچنین تفسیر آن‌ها به صورت محلی و عمومی نشان می‌دهد. بااین‌حال، افزودن متغیرهای پیشگوی مهم دیگر، ازجمله شاخص توده بدنی مادر و همچنین استفاده از ترکیب‌های مختلف آبرپارامترها می‌تواند بر افزایش دقت پیش‌بینی برای مطالعات آینده تاثیرگذار باشد. با بهره‌گیری از روش LIME، ارائه تفسیری شهودی از مدل‌های جعبه سیاه ممکن گردید. براین اساس و با این روش، می‌توان

[1]Convolutional Neural Networks (CNNs)
[2]Layer-wise Relevance Propagation (LRP)
[3]T 2weighted Magnetic Resonance Imaging (MRI)-
[4]Heat Map
[5]BMI

[Downloaded from irje.tums.ac.ir on 2026-03-24]
[DOI: 10.18502/ijre.v20i1.17622]

برای یک خانم باردار، توصیه‌های پیشگیرانه مختص آن فرد را در جلوگیری از زایمان زودرس ارائه نمود. تفسیرهای حاصل، می‌تواند به متخصصین حوزه‌ی سلامت در درک تأثیر هر یک از عوامل خطر، کمک کرده و در نتیجه ارائه مراقبت‌های مناسب برای کاهش عوارض و مرگ و میر مربوط به زایمان زودرس فراهم شود.

تشکر و قدردانی

نویسندگان مقاله بر خود لازم می‌دانند که از دفتر سلامت خانواده و جمعیت وزارت بهداشت درمان و آموزش پزشکی به‌خاطر همکاری و دراختیارگذاری داده‌های این مطالعه، قدردانی نمایند.

References

1. Beck S, Wojdyla D, Say L, Betran AP, Merialdi M, Requejo JH, et al. The worldwide incidence of preterm birth: a systematic review of maternal mortality and morbidity. Bull World Health Organ. 2010;88(1):31-8.
2. Chawanpaiboon S, Vogel JP, Moller AB, et al. Global, regional, and national estimates of levels of preterm birth in 2014: a systematic review and modelling analysis. Lancet Glob Health. 2019; 7(1): :e37-e46.
3. Blencowe H, Cousens S, Chou D, Oestergaard M, Say L, Moller A-B, et al. Born Too Soon: The global epidemiology of 15 million preterm births. Reproductive Health. 2013;10(1):S2.
4. Liu L, Oza S, Hogan D, Perin J, Rudan I, Lawn JE, et al. Global, regional, and national causes of child mortality in 2000–13, with projections to inform post-2015 priorities: an updated systematic analysis. The Lancet. 2015;385(9966):430-40.
5. Sun Q, Zou X, Yan Y, Zhang H, Wang S, Gao Y, et al. Machine Learning-Based Prediction Model of Preterm Birth Using Electronic Health Record. J Healthc Eng. 2022;2022:9635526.
6. Raja R, Mukherjee I, Sarkar BK. A Machine Learning-Based Prediction Model for Preterm Birth in Rural India. Journal of Healthcare Engineering. 2021;2021(1):6665573.
7. Koivu A, Sairanen M. Predicting risk of stillbirth and preterm pregnancies with machine learning. Health Inf Sci Syst. 2020;8(1):14.
8. Gunning D, Aha D. DARPA's Explainable Artificial Intelligence (XAI) Program. AI Magazine. 2019;40(2):44-58.
9. Ribeiro M, Singh S, Guestrin C. "Why Should I Trust You?": Explaining the Predictions of Any Classifier2016. 97-101 p.
10. Gabbay F, Bar-Lev S, Montano O, Hadad N. A LIME-Based Explainable Machine Learning Model for Predicting the Severity Level of COVID-19 Diagnosed Patients. Applied Sciences. 2021;11(21):10417.
11. Duell J, Fan X, Burnett B, Aarts G, Zhou SM, editors. A Comparison of Explanations Given by Explainable Artificial Intelligence Methods on Analysing Electronic Health Records. 2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI); 2021 27-30 July 2021.
12. Rawashdeh H, Awawdeh S, Shannag F, Henawi E, Faris H, Obeid N, et al. Intelligent system based on data mining techniques for prediction of preterm birth for women with cervical cerclage. Computational Biology and Chemistry. 2020;85:107233.
13. Damaso EL, Rolnik DL, Cavalli RC, Quintana SM, Duarte G, da Silva Costa F, et al. Prediction of Preterm Birth by Maternal Characteristics and Medical History in the Brazilian Population. Journal of pregnancy. 2019;2019(1):4395217.
14. Moreira MWL, Rodrigues JJPC, Marcondes GAB, Neto AJV, Kumar N, Diez IDLT, editors. A Preterm Birth Risk Prediction System for Mobile Health Applications Based on the Support Vector Machine Algorithm. 2018 IEEE International Conference on Communications (ICC); 2018 20-24 May 2018.
15. Prema NS, Pushpalatha MP. Machine learning approach for preterm birth prediction based on maternal chronic conditions. InEmerging Research in Electronics, Computer Science and Technology: Proceedings of International Conference, ICERECT 2018 2019 (pp. 581-588). Springer Singapore.
16. Khalili N, Moradi-Lakeh M, Heidarzadeh M. Low birth weight in Iran based on Iranian Maternal and Neonatal Network (IMaN). Med J Islam Repub Iran. 2019;33:30.
17. Pourshirazi M, Heidarzadeh M, Taheri M, Esmaily H, Babaey F, Talkhi N, et al. Cesarean delivery in Iran: a population-based analysis using the Robson classification system. BMC Pregnancy Childbirth. 2022;22(1):185.
18. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. J Artif Int Res. 2002;16(1):321–57.
19. Mustaqim AZ, Adi S, Pristyanto Y, Astuti Y, editors. The Effect of Recursive Feature Elimination with Cross-Validation (RFECV) Feature Selection Algorithm toward Classifier Performance on Credit Card Fraud Detection. 2021 International Conference on Artificial Intelligence and Computer Science Technology (ICAICST); 2021 29-30 June 2021.
20. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; San Francisco, California, USA: Association for Computing Machinery; 2016. p. 785–94.
21. Do HJ, Moon KM, Jin H-S. Machine Learning Models for Predicting Mortality in 7472 Very Low Birth Weight Infants Using Data from a Nationwide Neonatal Network. Diagnostics (Basel). 2022: 12(3):625.
22. Doshi-Velez F, Kim B. Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608. 2017 Feb 28.
23. Lundberg SM, Nair B, Vavilala MS, Horibe M, Eisses MJ, Adams T, et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. Nature Biomedical Engineering. 2018;2(10):749-60.
24. Grigorescu I, Cordero-Grande L, Edwards AD, Hajnal J, Modat M, Deprez M. Interpretable convolutional neural networks for preterm birth classification. arXiv preprint arXiv:191000071. 2019.

[Downloaded from irje.tums.ac.ir on 2026-03-24]

[DOI: 10.18502/ijre.v20i1.17622]

Tehran University of Medical Sciences

Iranian Journal of Epidemiology

https://irje.tums.ac.ir/

IrEA Iranian Epidemiological Association

Original Article

Predicting the Occurrence of Preterm Birth and Determining its Risk Factors Individually Using an Interpretable Machine Learning Model

Ramin Farrokhi[1], Samaneh Hosseinzadeh[2], Abbas Habibelahi[3], Akbar Biglarian[4]

1- MSc. Student in Biostatistics, Department of Biostatistics and Epidemiology, University of Social Welfare and Rehabilitation Sciences, Tehran, Iran

2- Assistant Professor of Biostatistics, Department of Biostatistics and Epidemiology, University of Social Welfare and Rehabilitation Sciences, Tehran, Iran

3- Assistant Professor of Pediatrics, Vice Chancellery for Health, Iran Ministry of Health and Medical Education, Tehran, Iran

4- Professor of Biostatistics, Department of Biostatistics and Epidemiology, Social Determinants of Health Research Center, Social Health Research Institute, University of Social Welfare and Rehabilitation Sciences, Tehran, Iran

Article Information

Received
27 December 2023

Accepted
24 April 2024

Corresponding author
Akbar Biglarian

Corresponding author E-mail
abiglarian@gmail.com

Keywords:
Pregnancy, Premature birth, Machine learning, Interpretability, Model-agnostic

Abstract

Background and Objectives: Identifying pregnant women who are at risk of premature birth and determining its risk factors is essential because it affects their health. This study aimed to use an interpretable machine-learning model to predict premature birth.

Methods: In this study, data from 149,350 births in Tehran in 2019 were utilized from the Iranian Mothers and Babies Network (IMaN) dataset. Various factors related to the mother and the fetus, such as the mother's demographic variables and health status, medical history, pregnancy conditions, childbirth, and associated risks, were considered. The machine learning models, including multilayer neural networks, random forest, and XGBoost, were employed to predict the occurrence of preterm birth after data preprocessing. The models were evaluated based on accuracy, sensitivity, specificity, and area under the ROC curve. The Python programming language version 3.10.0 was applied to analyze the data.

Results: About 8.67% of births were premature. The XGBoost algorithm achieved the highest prediction accuracy (90%). According to the model output, multiple births, which account for 46% of pregnant women's births, had the highest importance score. Delivery risk factors had a score of 41%, and other variables, including neurological and mental illness, preeclampsia, and cardiovascular disease, were subsequently ranked in order of importance for this particular individual.

Conclusion: Using an interpretable machine learning method could predict the occurrence of premature birth. Based on risk factors, the interpretable machine learning method can provide personalized preventive recommendations for every pregnant woman, aiming to reduce the risk of preterm birth.

Copyright © 2024 The Authors. Published by Tehran University of Medical Sciences.

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International license (https://creativecommons.org/licenses/by-nc/4.0/). Non-commercial uses of the work are permitted, provided the original work is properly cited.

[Downloaded from irje.tums.ac.ir on 2026-03-24]

[DOI: 10.18502/ijre.v20i1.17622]