

مقایسه شبکه‌های عصبی مصنوعی، درخت تصمیم، تحلیل تشخیصی و رگرسیون لوجستیک در پیش‌بینی بارداری ناخواسته در مادران مولتی‌پار شهر خرم‌آباد

نویسندگان: فرزاد ابراهیم‌زاده^۱، فرید زایری^۳، نسیم وهابی^۲، علی آذربو^۴،
کتایون بختیار^۵ و آغافاطمه حسینی^{۶*}

۱. مربی آمار زیستی گروه بهداشت عمومی، دانشکده بهداشت و تغذیه دانشگاه علوم پزشکی لرستان، خرم‌آباد، ایران
۲. دانشجوی دکتری آمار زیستی گروه آمار زیستی، دانشکده علوم پزشکی دانشگاه تربیت مدرس، تهران، ایران
۳. دانشیار آمار زیستی، عضو مرکز تحقیقات پروتئومیکس، دانشکده پیراپزشکی دانشگاه علوم پزشکی شهید بهشتی، تهران، ایران
۴. دانشجوی دکتری آمار گروه آمار، دانشکده علوم ریاضی و کامپیوتر دانشگاه صنعتی امیرکبیر، تهران، ایران
۵. مربی مامایی گروه بهداشت عمومی، دانشکده بهداشت و تغذیه دانشگاه علوم پزشکی لرستان، خرم‌آباد، ایران
۶. مربی آمار زیستی گروه آمار زیستی، دانشکده بهداشت دانشگاه علوم پزشکی ایران، تهران، ایران

E-mail: hosseini_af@yahoo.com

* نویسنده مسئول: آغافاطمه حسینی

چکیده

مقدمه و هدف: بارداری ناخواسته، نوعی از بارداری است که دست‌کم از نظر یکی از زوجین، ناخواسته باشد و پیامدهایی نامطلوب را برای خانواده و اجتماع به‌همراه دارد. در این مطالعه با استفاده از چهار مدل طبقه‌بندی، وقوع بارداری ناخواسته در جمعیت شهری خرم‌آباد پیش‌بینی شد و مدل‌ها مورد مقایسه قرار گرفتند.

مواد و روش‌ها: در این مطالعه مقطعی، ۴۶۷ نفر از مادران باردار مولتی‌پار مراجعه‌کننده به مراکز بهداشتی-درمانی خرم‌آباد در سال ۱۳۹۰ با روش نمونه‌گیری طبقه‌ای و خوشه‌ای انتخاب شده، متغیرهای مرتبط اندازه‌گیری شدند. در مدل‌بندی داده‌ها از مدل رگرسیون لوجستیک، تحلیل تشخیصی، درخت تصمیم‌گیری و شبکه‌های عصبی پرسپترون سه‌لایه و نرم‌افزار SPSS و MATLAB استفاده شد؛ برای مقایسه این مدل‌ها، شاخص‌های حساسیت، ویژگی، سطح زیر منحنی مشخصه و میزان پیش‌بینی صحیح به‌کار گرفته شدند.

نتایج: شیوع بارداری ناخواسته، ۳/۲۲ درصد بود. عملکرد مدل‌ها براساس شاخص سطح زیر منحنی مشخصه به ترتیب عبارت بود: از شبکه‌های عصبی مصنوعی (۰/۷۴۱)، درخت تصمیم‌گیری (۰/۷۳۱)، رگرسیون لوجستیک (۰/۷۱۲) و تحلیل تشخیصی (۰/۷۱۱). بیشترین درصد حساسیت به مدل درخت تصمیم‌گیری (۷۳/۵ درصد) و بیشترین میزان ویژگی به شبکه عصبی مصنوعی (۶۲/۳ درصد)، مربوط بوده است.

نتیجه‌گیری: با توجه به شیوع به نسبت بالای بارداری ناخواسته در شهر خرم‌آباد، لزوم بازنگری در برنامه‌های تنظیم خانواده، محسوس است. به‌رغم تشابه نسبی روش‌های یادشده، اگر توان پیشگویی بالاتر مدل، مدنظر محقق باشد، مدل شبکه‌های عصبی و اگر تفسیرپذیری بهتر نتایج، مدنظر باشد، استفاده از درخت تصمیم‌گیری و رگرسیون لوجستیک توصیه می‌شوند.

واژگان کلیدی: شبکه عصبی مصنوعی، درخت تصمیم‌گیری، رگرسیون لوجستیک، تحلیل تشخیصی، مادر باردار مولتی‌پار.

دانشور

پزشکی

دوماهنامه علمی-پژوهشی

دانشگاه شاهد

سال بیست‌ودوم-شماره ۱۱۶

اردیبهشت ۱۳۹۴

دریافت: ۱۳۹۳/۱۲/۱۶

آخرین اصلاح‌ها: ۱۳۹۴/۰۱/۲۲

پذیرش: ۱۳۹۴/۰۱/۲۶

مقدمه

بدیهی است که بارداری ناخواسته، به‌خودی‌خود، پیامدهایی نامطلوب را برای مادر، کودک، والدین و اجتماع به‌همراه دارد و شایع‌ترین پیامد آن، اقدام به سقط‌های غیرقانونی و غیرایمن است که از علل اصلی مرگ‌ومیر و معلولیت مادران به‌شمار می‌روند؛ همچنین به دلیل احساسات و عواطف ناخوشایند مادر، طی دوران بارداری، فشار روانی و عصبی وی افزایش یافته، مراجعات و مراقبت‌های دوران بارداری، کاهش چشمگیر می‌یابند و حتی ممکن است آثاری نظیر افسردگی، خودکشی و کاهش کیفیت زندگی مادر را در پی داشته باشند (۳، ۱۰، ۱۳)؛ در ضمن، مهم‌ترین تأثیرهای بارداری ناخواسته روی کودک، اغلب در قالب «تولد نوزادان نارس و کم‌وزن، کم‌اشتهایی، بیش‌فعالی، آزار و اذیت و بی‌توجهی از سوی والدین و سرانجام افزایش مرگ‌ومیر کودکان»، خود را نشان می‌دهند (۳، ۵، ۱۲، ۱۶).

حال با توجه به اهمیت موضوع و با علم به اینکه وقوع بارداری ناخواسته در هر جامعه، از عوامل فردی و اجتماعی متنوع و منحصر به فردی، متأثر می‌شود، لازم است در هر جامعه‌ای به‌طور جداگانه عوامل تعیین‌کننده آن، مورد بررسی جامع قرار گیرند. استفاده از روش‌هایی موسوم به مدل‌های رده‌بندی، این امکان را فراهم می‌سازند تا براساس مجموعه‌ای از ویژگی‌ها و متغیرهای مرتبط با زوجین، خطر وقوع بارداری ناخواسته در این افراد پیش‌بینی شود؛ رایج‌ترین روش‌های رده‌بندی در این راستا عبارت‌اند از: رگرسیون لجستیک، تحلیل تشخیصی، درخت تصمیم‌گیری و شبکه‌های عصبی مصنوعی که البته دو روش اخیر، از روش‌های ناپارامتری موسوم به داده‌کاوی محسوب می‌شوند؛ این روش‌ها، صرف‌نظر از تفاوت در روش‌های برآوردی و الگوریتم‌های محاسباتی، از نظر دقت رده‌بندی، تفسیرپذیری نتایج، زمان محاسباتی و در دسترس بودن نرم‌افزارهای آماری با یکدیگر متفاوت‌اند که در بسیاری از مطالعات به این تفاوت‌ها

بارداری ناخواسته یا بدون برنامه‌ریزی، نوعی از بارداری است که دست‌کم از نظر یکی از زوجین ناخواسته باشد (۱، ۲). همه ساله در جهان، حدود ۲۰۰ میلیون بارداری رخ می‌دهند که به‌طور تقریبی یک‌سوم، یعنی حدود ۷۵ میلیون مورد آنها ناخواسته‌اند و از سویی، سالانه ۵۰ میلیون بارداری ناخواسته به سقط منجر می‌شوند که ۲۰ میلیون آنها در شرایط غیرایمن صورت می‌گیرند (۳-۵). براساس مطالعات انجام‌شده در ایالات متحده، ژاپن و تانزانیا، شیوع بارداری ناخواسته به‌ترتیب: ۴۸ درصد، ۶۷/۲ درصد و ۲۳/۷ درصد اعلام شده‌است (۶-۸). مطالعات انجام‌گرفته در ایران نیز از آن حکایت‌دارند که به‌رغم دسترسی آسان زوجین به وسایل پیشگیری، سالیانه ۴۰۰ تا ۵۰۰ هزار مورد بارداری ناخواسته اتفاق می‌افتند که ۱۹ درصد این بارداری‌ها (حدود ۸۰ هزار مورد) به سقط جنین منجر می‌شوند (۹-۱۱).

براساس آمار وزارت بهداشت ایران از هر چهار بارداری، یک بارداری از نظر زوجین، ناخواسته است (۵، ۱۲). براساس مرور نظام‌مند (سیستماتیک) صورت‌گرفته در ایران، شیوع بارداری ناخواسته، حدود ۲۹/۷ درصد بوده‌است و همچنین، مطالعه مروری دیگری، این شیوع را حدود ۳۰/۶ درصد برآورد کرده‌است (۱، ۲).

مهم‌ترین علت بارداری ناخواسته، عدم استفاده از روش‌های پیشگیری از بارداری به‌رغم عدم تمایل به بارداری و نیز استفاده ناصحیح از روش و در نتیجه، شکست است (۴، ۹، ۱۰، ۱۲)؛ در واقع، تحریم‌های اجتماعی، نابرابری حقوق اجتماعی مرد و زن، مخالفت همسران، عدم دسترسی به روش‌های مدرن پیشگیری، برنامه‌ریزی ناکافی، کم‌رنگ‌بودن نقش مشاوره و نداشتن مهارت کافی کارکنان بهداشتی، از جمله عواملی هستند که سبب می‌شوند، روش‌های توصیه‌شده با شرایط فرد سازگاری نداشته، موجب عدم استفاده یا شکست روش‌ها شوند (۴).

اشاره شده است (۱۹-۱۷).

مروری بر مطالعات انجام شده در زمینه بارداری ناخواسته نشان می‌دهد که در بیشتر موارد، فقط به بررسی شیوع این عارضه پرداخته شده و در نهایت با استفاده از آزمون‌های آماری t مستقل و مجذور کای به بررسی ارتباط تک‌تک متغیرهای مستقل با وقوع بارداری ناخواسته پرداخته شده است (۲۶-۲۰، ۱۳، ۵، ۳)؛ در تعدادی معدود از مطالعات نیز سعی شده تا با استفاده از مدل‌های رده‌بندی، اثر هم‌زمان مجموعه‌ای از متغیرها بر بارداری ناخواسته بررسی شود که البته به دلیل محدودیت‌های موجود در انتخاب جامعه آماری، نتایج این مطالعات به‌طور عمده، قابل‌تعمیم به سایر جوامع نیستند (۲۷، ۱۱، ۹، ۴، ۳). مدل‌های آماری مورد استفاده در پژوهش‌های پیشین، شامل رگرسیون لجستیک ساده (۲۹-۲۹، ۱۱، ۶، ۳، ۲۷)، رگرسیون لجستیک چندجمله‌ای (۳۰، ۹)، شبکه‌های عصبی مصنوعی (۳۱، ۳۰، ۲۷) و مدل‌های لگ خطی (۳۲) بوده‌اند. لازم به یادآوری است که با وجود استفاده گسترده از روش درخت تصمیم‌گیری در پیش‌بینی بسیاری از پیامدهای پزشکی (۳۳-۳۹)، در مطالعاتی محدود در حیطه بارداری ناخواسته از این روش استفاده شده است.

حال با توجه به اهمیت موضوع بارداری ناخواسته و با علم به اینکه تاکنون مطالعه مقایسه‌ای جامعی به‌منظور پیش‌بینی بارداری ناخواسته در غرب کشور صورت نگرفته است، در این مطالعه سعی می‌شود تا با استفاده از چهار روش آماری «رگرسیون لجستیک، تحلیل تشخیصی، درخت تصمیم‌گیری و شبکه‌های عصبی مصنوعی»، بهترین مدل طبقه‌بندی در پیش‌بینی بارداری ناخواسته در جمعیت شهری خرم‌آباد معرفی شود.

مواد و روش‌ها

داده‌های پژوهش

جامعه مورد بررسی در این مطالعه مقطعی، تمامی مادران باردار مولتی‌پار مراجعه‌کننده به مراکز بهداشتی-

درمانی شهری خرم‌آباد در سال ۱۳۹۰ برای دریافت مراقبت‌های دوران بارداری هستند. نمونه‌ای متشکل از ۶۷ زن باردار با استفاده از روش نمونه‌گیری تصادفی طبقه‌ای انتخاب شدند سپس، درون هر طبقه از نمونه‌گیری خوشه‌ای استفاده شد به این صورت که ابتدا مراکز بهداشتی-درمانی شهری خرم‌آباد به سه طبقه شمال شهر، مرکز شهر و جنوب شهر تقسیم شده، دو مرکز (سرخوشه) از شمال، دو مرکز از جنوب و سه مرکز از محدوده مرکز شهر به‌طور تصادفی انتخاب شدند؛ سپس مادرانی که به‌منظور مراقبت‌های دوران بارداری به مراکز بهداشتی منتخب مراجعه کردند، به ترتیب مراجعه به نمونه وارد شدند تا حجم نمونه نهایی مورد نیاز از هر مرکز انتخاب شود.

ابزار اندازه‌گیری و جمع‌آوری اطلاعات در مطالعه حاضر، برگه جمع‌آوری اطلاعات محقق‌ساخته است که در اختیار مادران باردار قرار گرفت تا به‌صورت خودایفا به آن پاسخ دهند؛ البته برای مادران بی‌سواد یا کم‌سواد با روش مصاحبه تکمیل شد؛ این برگه، «مشخصات دموگرافیک مادران (نام مرکز بهداشتی-درمانی، سن زوجین، اختلاف سنی زوجین، مرتبه بارداری، فاصله بارداری فعلی از بارداری پیشین، تعداد فرزندان زنده پسر، نوع روش پیشگیری، سطح تحصیلات زوجین، سطح درآمد خانوار، شغل زوجین، زیربنای واحد مسکونی و مالکیت خودروی شخصی) و نیز پرسشی را در زمینه خواسته یا ناخواسته بودن بارداری فعلی از نظر دست‌کم یکی از زوجین به‌عنوان پیامد اصلی مورد بررسی» دربرمی‌گرفت.

روش‌های آماری

رگرسیون لجستیک^۱: یک مدل رگرسیون برای تحلیل متغیرهای پاسخ دوحالتی (موفقیت و شکست) است که در واقع، عضوی از خانواده مدل‌های خطی تعمیم‌یافته بوده، از تابع لجیت به‌عنوان تابع ربط استفاده می‌کند؛ قالب (فرم) کلی این مدل به‌صورت زیر است:

^۱- Logistic Regression

هستند.

درخت تصمیم^۴ این درخت، از سه جزء اصلی شامل ریشه، گره داخلی و برگ تشکیل شده و روند به این صورت است که ابتدا یک متغیر مستقل به عنوان ریشه انتخاب و به چندین گره داخلی تقسیم می‌شود؛ هر گره داخلی نیز مانند ریشه به گره‌هایی دیگر تقسیم می‌شود تا در نهایت به هر گره، یک رده از متغیر وابسته نسبت داده شود؛ این گره‌ها برگ نامیده می‌شوند. اغلب، برای انتخاب متغیرهای مهم در الگوی طبقه‌بندی درختی، از تابعی با عنوان ناخالصی^۵ و شاخصی به نام جینی^۶ استفاده می‌شود. ابتدا براساس تابع ناخالصی و شاخص جینی، مقدار تابع ناخالصی در حالت کلی برای متغیر وابسته محاسبه می‌شود سپس برای تمامی متغیرهای مستقل، با توجه به بهترین تقسیم‌های دوتایی برای متغیر وابسته، مقدار تابع ناخالصی در هر یک از دو زیرمجموعه ایجاد شده، محاسبه و میانگین وزنی آنها از مقدار تابع ناخالصی کل کسرمی‌شود. از میان متغیرهای مستقل، متغیری که دارای بیشترین مقدار برای این رابطه باشد در گام اول برای طبقه‌بندی درختی انتخاب می‌شود. در مواجهه با متغیرهای مستقل کمی از بهترین نقطه برش برای تقسیم دوتایی استفاده می‌شود و در برخورد با متغیرهای کیفی نیز هر سطح متغیر به عنوان یک زیرشاخه درخت طبقه‌بندی در نظر گرفته می‌شود.

در مدلی خاص از درخت تصمیم‌گیری موسوم به CART با استفاده از روشی با عنوان هزینه-پیچیدگی، اندازه (عمق) مناسب درخت تعیین می‌شود؛ این روش می‌تواند میان دقت درخت تصمیم‌گیری و اندازه آن، تعادل برقرار کند؛ با استفاده از این روش، براساس اندازه خطای طبقه‌بندی و تعداد گره‌های درخت، تصمیم‌گرفته می‌شود که کدام گره از درخت باید هرس شود (۱۸،۳۹).

$$P = P(y_i = 1) = \frac{e^{\alpha + \sum_{i=1}^n \beta_i x_i}}{1 + e^{\alpha + \sum_{i=1}^n \beta_i x_i}}$$

که در آن α ، عرض از مبدأ مدل؛ x_i متغیر کمکی یا مستقل و β_i ضریب متغیر کمکی نام است. رگرسیون لجستیک قابلیت دارد که علاوه بر انجام پیش‌بینی و برآورد ضرایب و اثر هر متغیر کمکی، برای اهداف طبقه‌بندی و تشخیص نیز به کار برده شود.

شبکه عصبی مصنوعی^۱ در واقع، شبیه‌سازی مغز انسان از طریق مدل‌بندی سلول‌های عصبی است که هر سلول عصبی به عنوان یک واحد پردازشگر عمل می‌کند. شبکه عصبی پرسپترون چندلایه^۲ (MLP)، یکی از پرکاربردترین انواع شبکه است و ساختار آن، شامل چند لایه (لایه ورودی، لایه(های) میانی و لایه خروجی) است که در هر لایه، تعدادی گره و تابع فعالیت تعریف می‌شود؛ خروجی هر لایه به وسیله مجموع ضرایب وزنی آن لایه محاسبه شده، از طریق تابع فعالیت به لایه بعدی ارسال می‌شود.

روش‌ها و الگوریتم‌هایی متفاوت برای پیدا کردن وزن‌ها وجود دارند که در شبکه MLP از الگوریتم پس‌انتشار خطا^۳ (BP) استفاده می‌شود؛ به علاوه، تابع فعالیت در مدل شبکه عصبی همانند توابع ربط در مدل‌های خطی تعمیم یافته است که به عنوان نمونه می‌توان به توابع سیگموئید و تانژانت هایپربولیک اشاره کرد. در یک حالت کلی و ساده شبکه عصبی MLP با یک لایه میانی، مقدار خروجی واحد نام به صورت زیر محاسبه می‌شود:

$$\hat{y}_i = f_2 \left(\beta_0 + \sum_{k=1}^p w_k f_1 \left(\beta_k + \sum_{j=1}^m x_{ij} w_{kj} \right) \right)$$

که در آن، m تعداد مشاهدات؛ p تعداد گره‌های لایه میانی؛ n تعداد گره‌های لایه ورودی (تعداد متغیرهای کمکی)؛ w_{kj} وزن مربوط به ورودی x_{ij} در گره k ام؛ w_k وزن مربوط به گره k ام؛ β_0 و β_k به ترتیب، مقادیر اریبی لایه‌های خروجی و میانی و f_1 و f_2 به ترتیب، توابع فعالیت لایه‌های میانی و خروجی شبکه

⁴ - Decision Tree
⁵ - ImpurityFunction
⁶ - Gini

¹ - Artificial Neural Network
² - Multi Layer Perceptron
³ - Back Propagation

برای رسم درخت تصمیم‌گیری از درخت رده‌بندی CART و به‌منظور انتخاب متغیرهای مهم و نقطه برش آنها به‌ترتیب از تابع ناخالصی و ضریب جینی استفاده شد؛ درنهایت، برای مقایسه مدل‌های یادشده از شاخص‌های حساسیت، ویژگی، سطح زیر منحنی مشخصه و میزان پیش‌بینی صحیح استفاده‌شده و همچنین، وجود تفاوت آماری معنادار میان شاخص سطح زیر منحنی مشخصه عملکرد در چهار روش یادشده نیز آزمون شد (با استفاده از نرم‌افزار STATA).

نتایج

میانگین (\pm انحراف معیار) سن ۴۶۷ مادر مولتی‌پار مورد بررسی در این مطالعه مقطعی ۲۹/۷ ($\pm ۴/۹۴$) سال (محدوده سنی ۵ تا ۱۶ سال) بود؛ ۷۴/۲ درصد (۳۴۷ نفر) از مادران در گروه سنی پایین‌تر از ۳۵ سال قرارداشته، به‌طور تقریبی ۲۹ درصد (۱۳۶ نفر) از آنها دارای فرزند زنده پسر بودند. از میان مادران باردار، ۴۷/۷ درصد (۲۲۳ نفر)، دارای مرتبه بارداری دو؛ ۳۷/۱ (۱۷۳ نفر)، دارای مرتبه بارداری سه و ۱۵/۲ درصد (۷۱ نفر)، دارای مرتبه بارداری بیشتر از چهار بودند. فاصله بارداری فعلی از بارداری پیشین در ۲۱/۲ درصد (۹۹ نفر) از مادران، کمتر از دو سال؛ در ۲۳/۸ درصد (۱۱۱ نفر)، میان دو تا چهار سال و در ۵۵/۰ درصد (۲۵۷ نفر) بیشتر از چهار سال بود. ۸۰/۲ درصد (۳۷۴ نفر) از مادران باردار، تحصیلات دانشگاهی داشتند و ۷۴/۴ درصد (۳۴۷ نفر) از آنها درآمد زیر ۵۰۰ هزار تومان داشتند. از میان تمامی بارداری‌ها، ۳۲/۳ درصد (۱۵۱) از آنها ناخواسته بوده و از میان افرادی که بارداری ناخواسته داشته‌اند، ۹۲/۵ درصد (۴۳۲ نفر)، پیش از بارداری از روش‌های پیشگیری استفاده‌می‌کرده‌اند. به‌منظور بررسی ارتباط متغیرها با بارداری ناخواسته، نتایج تحلیل تک‌متغیره در جدول ۱ ارائه شده‌است.

تحلیل تشخیصی^۱: این روش برای دسته‌بندی و تخصیص مشاهدات جدید به گروه‌های از پیش تعیین‌شده است. تخصیص مشاهدات به گروه‌ها با استفاده از تابع تشخیص انجام می‌شود و تابع تشخیص فیشر، یکی از پرکاربردترین توابع مورد استفاده در تحلیل تشخیصی است. مقدار تابع تشخیص به‌طور کلی به‌صورت زیر محاسبه می‌شود:

$$f = \sum_{i=1}^n \alpha_i x_i$$

که در آن α_i ضریب تابع تشخیص مربوط به متغیر کمکی x_i بوده و f مقدار تابع تشخیص است که با مقدار به‌دست‌آمده از رابطه $m = \frac{1}{2} (\bar{x}_1 - \bar{x}_2)' s_p^{-1} (\bar{x}_1 + \bar{x}_2)$ مقایسه‌شده، ملاک تصمیم‌گیری در گروه‌بندی مشاهده جدید به یک گروه خاص قرار می‌گیرد.

برای توصیف و مدل‌بندی داده‌ها و مقایسه مدل رگرسیون لوجستیک، تحلیل تشخیصی، درخت تصمیم‌گیری و شبکه عصبی مصنوعی از نرم‌افزارهای SPSS و MATLAB نسخه R2008a استفاده شد. به‌منظور انتخاب متغیرهای مورد بررسی از رگرسیون لوجستیک به روش انتخاب پیشرو استفاده شد که سطح معناداری برای ورود یک متغیر به مدل $P = 0/15$ و خروج یک متغیر از مدل $P = 0/10$ بود و مقادیر احتمال کمتر از ۰/۰۵ ($P < 0/05$) به‌عنوان معنادار آماری در نظر گرفته شدند. برای برازش شبکه عصبی مصنوعی، داده‌ها به دو بخش آموزش و آزمایش شبکه تقسیم شدند که داده‌های آموزش شبکه در حدود ۷۰ درصد (۳۲۷ مشاهده) از داده‌های مطالعه را تشکیل دادند؛ درنهایت، شبکه عصبی مصنوعی با ده‌گانه ورودی، هفت‌گانه میانی، دو‌گانه خروجی و الگوریتم یادگیری پس‌انتشار خطا انتخاب شد. در تحلیل تشخیصی از تابع تشخیص فیشر برای تخصیص مشاهدات جدید به گروه‌های از پیش تعیین‌شده استفاده شد.

^۱- Discriminant Analysis

جدول ۱. توزیع فراوانی کل بارداری‌ها و بارداری‌های ناخواسته زنان باردار مراجعه‌کننده به مراکز بهداشتی-درمانی شهری خرم‌آباد در سال ۱۳۹۰ به تفکیک متغیرهای مختلف

مقدار احتمال	بارداری ناخواسته		طبقه	متغیر
	درصد (تعداد)	کل درصد (تعداد)		
<۰/۰۰۱	۷۴/۲ (۱۱۲)	۸۲/۲ (۳۸۴)	کمتر از ۳۵ سال	سن مادر
	۲۵/۸ (۳۹)	۱۷/۸ (۸۳)	بیشتر از ۳۵ سال	
<۰/۰۰۱	۷۰/۹ (۱۰۷)	۴۱/۸ (۱۹۵)	دارد	داشتن فرزند زنده پسر
	۲۹/۱ (۴۴)	۵۸/۲ (۲۷۲)	ندارد	
<۰/۰۰۱	۴۷/۷ (۷۲)	۶۷/۲ (۳۱۴)	۲	مرتب‌بندی بارداری
	۳۷/۱ (۵۶)	۲۲/۹ (۱۰۷)	۳	
	۱۵/۲ (۲۳)	۹/۹ (۴۶)	۴ و بیشتر	
<۰/۰۰۱	۲۱/۳ (۳۲)	۱۳/۳ (۶۲)	کمتر از ۲ سال	فاصله بارداری فعلی از بارداری پیشین
	۲۳/۸ (۳۶)	۲۲/۳ (۱۰۴)	۲ تا ۴ سال	
	۵۵/۰ (۸۳)	۶۴/۵ (۳۰۱)	بیشتر از ۴ سال	
۰/۴۴۹	۱۲/۸ (۱۹)	۱۵/۰ (۷۰)	دانشگاهی	سطح تحصیلات مادر
	۸۷/۲ (۱۳۲)	۸۵/۰ (۳۹۷)	غیر دانشگاهی	
۰/۲۲۸	۶۸/۹ (۱۰۴)	۷۰/۴ (۳۲۹)	کمتر از ۵۰۰ هزار تومان	درآمد
	۳۱/۱ (۴۷)	۲۹/۶ (۱۳۸)	بیشتر یا مساوی ۵۰۰ هزار تومان	

رگرسیون لجستیک در جدول ۲ ارائه شده‌اند؛ از میان متغیرهای مورد بررسی، متغیرهای «مرتب‌بندی بارداری، فاصله بارداری فعلی از بارداری پیشین و داشتن فرزند زنده پسر»، رابطه آماری معناداری با بروز بارداری ناخواسته داشته و تنها متغیر سن، رابطه آماری معناداری نداشت ($P = 0/102$)؛ به‌علاوه با توجه به مقادیر نسبت بخت‌ها، متغیر مرتبه بارداری ($OR=3/39$)، بیشترین تأثیر را در بروز بارداری ناخواسته داشته‌است؛ به‌علاوه، سطح زیر منحنی مشخصه عملکرد برای این مدل، ۰/۷۱۲ و میزان پیش‌بینی‌های صحیح مدل، ۰/۶۳۸ بود.

با توجه به نتایج تحلیل یک‌متغیره، سطح تحصیلات مادر ($P = 0/449$) و سطح درآمد، رابطه‌ای معنادار با بارداری ناخواسته نشان‌ندادند ($P = 0/228$) اما متغیرهای سن مادر، داشتن فرزند زنده پسر، مرتبه بارداری و فاصله بارداری فعلی از بارداری پیشین، رابطه‌ای معنادار داشتند؛ در نتیجه به‌عنوان متغیرهای کمکی در مدل رگرسیون لجستیک، شبکه عصبی مصنوعی، درخت تصمیم‌گیری و تحلیل تشخیصی وارد شدند. یافته‌های رگرسیون لجستیک: نتایج برازش مدل

جدول ۲. تأثیر متغیرهای سن مادر، داشتن فرزند زنده پسر، مرتبه بارداری و فاصله بارداری فعلی از بارداری پیشین زنان باردار مراجعه‌کننده به مراکز بهداشتی-درمانی شهری خرم‌آباد در سال ۱۳۹۰، روی بارداری ناخواسته با استفاده از

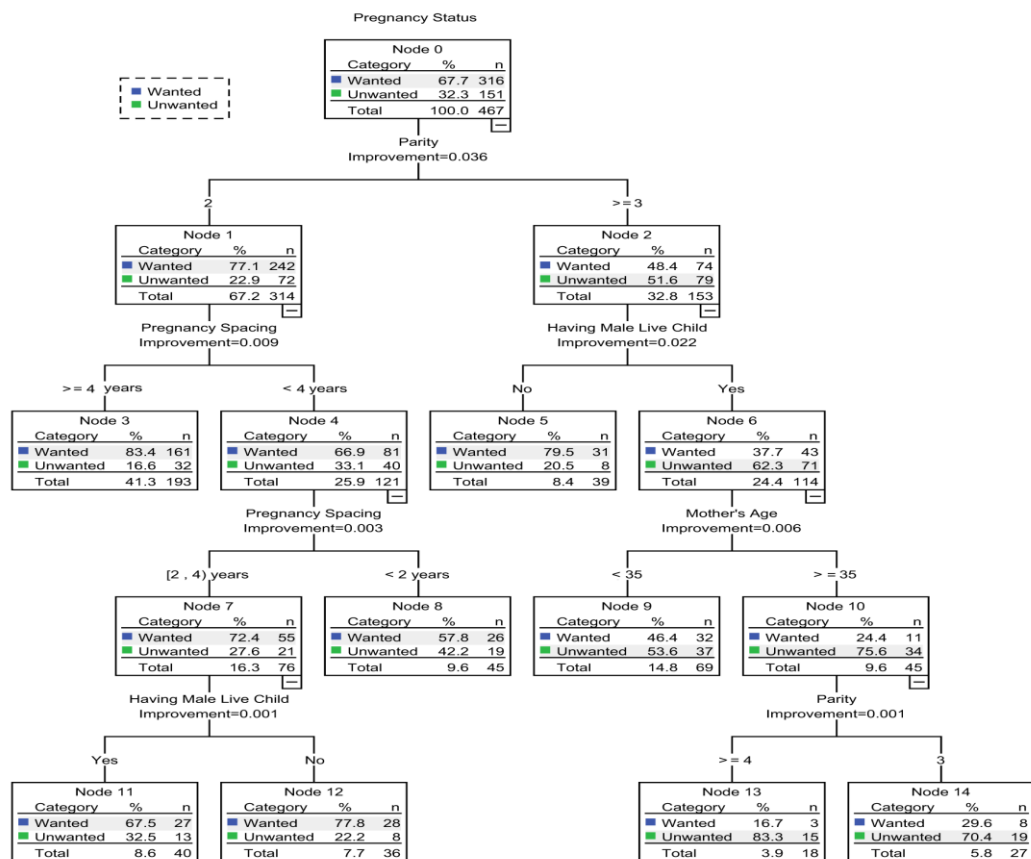
مدل رگرسیون لجستیک

مقدار احتمال	فاصله اطمینان ۹۵ درصد برای نسبت بخت‌ها		نسبت بخت‌ها (OR)	گروه	متغیر
	حد بالا	حد پایین			
۰/۱۰۲	-	-	مرجع	کمتر از ۳۵ سال	سن مادر
-	-	-	۱/۶۰۱	بیشتر از ۳۵ سال	
۰/۱۰۲	۲/۸۱۷	۰/۹۱۰	مرجع	دارد	داشتن فرزند زنده پسر
۰/۱۰	-	-	۱/۷۹۱	ندارد	
<۰/۰۰۱	-	-	مرجع	۲	مرتب‌بندی بارداری
<۰/۰۰۱	۵/۶۳۵	۲/۰۴۲	۳/۳۹۲	۳	
۰/۰۰۴	۵/۶۲۶	۱/۴۰۳	۲/۸۰۹	۴ و بیشتر	
<۰/۰۰۱	-	-	مرجع	کمتر از ۲ سال	فاصله بارداری فعلی از بارداری پیشین
۰/۰۲۶	۰/۹۰۹	۰/۲۳۳	۰/۴۶۰	۲ تا ۴ سال	
<۰/۰۰۱	۰/۴۷۳	۰/۱۴۰	۰/۲۵۷	بیشتر از ۴ سال	

۰/۶۵۰ به دست آمد.

یافته‌های درخت تصمیم: با توجه به تابع ناخالصی و شاخص جینی، مرتبه بارداری به عنوان متغیر ورودی (مهم‌ترین متغیر پیشگو) انتخاب شد و درخت تصمیم‌گیری (CART) با عمق ۴ به صورت شکل ۱ به دست آمد. سطح زیر منحنی مشخصه عملکرد این مدل، ۰/۷۳۱ و میزان پیش‌بینی صحیح آن، ۰/۶۴۸ به دست آمد؛ با توجه به این درخت تصمیم، متغیر فاصله بارداری فعلی از بارداری پیشین در مادران با مرتبه بارداری دو و همچنین داشتن فرزند زنده پسر در مادران با مرتبه بارداری بالاتر از دو، دارای اهمیت بالاست؛ به علاوه، متغیر سن مادر نیز در مادران با مرتبه بارداری بزرگ‌تر از دو و دارای فرزند زنده پسر، بسیار اهمیت دارد.

یافته‌های شبکه عصبی مصنوعی: برای مدل‌سازی شبکه عصبی مصنوعی پرسپترون سه‌لایه، با توجه به مجموعه داده‌های گروه آموزشی و متغیرهای مستقل ورودی تمامی شبکه‌های مبنی بر پنج تا بیست گره در لایه میانی، با اندازه حرکت ۰/۸۰ تا ۰/۹۵، نرخ یادگیری ۰/۰۱ تا ۰/۴۰ و توابع پیوند سیگموئید و تانژانت هایپربولیک ارزیابی شدند؛ پس از بررسی تمامی مدل‌های ممکن برای ساختار شبکه عصبی سه‌لایه در نهایت، شبکه با ده گره ورودی، هفت گره میانی، دو گره خروجی، نرخ یادگیری ۰/۰۵، توابع فعالیت تانژانت هایپربولیک و سیگموئید و الگوریتم پخش‌خطا به عنوان بهترین شبکه عصبی انتخاب شد. سطح زیر منحنی مشخصه عملکرد ۰/۷۴۱ و میزان پیش‌بینی‌های صحیح مدل،



شکل ۱. درخت تصمیم‌گیری بارداری خواسته و ناخواسته زنان باردار مراجعه‌کننده به مراکز بهداشتی-درمانی شهری خرم‌آباد در سال ۱۳۹۰ با توجه به متغیرهای سن مادر، داشتن فرزند زنده پسر، مرتبه بارداری و فاصله بارداری فعلی از بارداری پیشین

در نهایت برای مقایسه چهار روش یادشده، مقادیر شاخص‌های حساسیت، ویژگی، سطح زیر منحنی مشخصه و پیش‌بینی صحیح هریک از مدل‌ها در جدول ۳ ارائه شده است. با توجه به مقادیر گزارش شده به طور کلی، شبکه عصبی مصنوعی و درخت تصمیم‌گیری بهتر و دقیق‌تر از سایر روش‌ها توانسته‌اند بارداری خواسته و ناخواسته را رده‌بندی و پیش‌بینی کنند؛ البته لازم به یادآوری است که تفاوت آماری معناداری میان سطوح زیر منحنی مشخصه عملکرد روش‌های نامبرده مشاهده نشد ($P > 0.05$).

یافته‌های تحلیل تشخیصی: با توجه به برآورد ضرایب تابع تشخیص، قالب (فرم) کلی مدل به صورت زیر حاصل شد که در آن، X_1 سن مادر؛ X_2 داشتن فرزند زنده پسر؛ X_3 مرتبه بارداری و X_4 فاصله بارداری فعلی از بارداری پیشین است. سطح زیر منحنی مشخصه عملکرد و میزان پیش‌بینی صحیح این مدل به ترتیب: ۰/۷۱۱ و ۰/۶۳۳ شدند.

$$f = 0.265x_1 + 0.370x_2 + 0.627x_3 - 0.603x_4$$

جدول ۳. مقادیر شاخص‌های حساسیت، ویژگی، سطح زیر منحنی مشخصه و پیش‌بینی صحیح مدل رگرسیون لجستیک، شبکه عصبی مصنوعی، درخت تصمیم‌گیری و تحلیل تشخیصی در بررسی بارداری خواسته و ناخواسته زنان باردار مراجعه‌کننده به مراکز بهداشتی-درمانی شهری خرم‌آباد در سال ۱۳۹۰

روش	سطح زیر منحنی مشخصه	حساسیت (%)	ویژگی (%)	پیش‌بینی صحیح (%)
رگرسیون لجستیک	۰/۷۱۲	۷۰/۲	۶۰/۱*	۰/۶۳۸
تحلیل تشخیصی	۰/۷۱۱	۶۹/۵	۶۱/۱	۰/۶۳۳
درخت تصمیم‌گیری	۰/۷۳۱	۷۳/۵	۶۰/۸	۰/۶۴۸
شبکه عصبی مصنوعی	۰/۷۴۱	۷۰/۹	۶۲/۳	۰/۶۵۰

بحث و نتیجه‌گیری

تمام ارتباطات پیچیده غیرخطی و اثرهای متقابل میان متغیرهای مستقل فراگرفته می‌شوند لذا توان پیشگویی این مدل به طور قابل ملاحظه‌ای بالا خواهد بود (۴۰٪)؛ با وجود این، در صورت عدم انتخاب معماری مناسبی برای شبکه و به خصوص انتخاب تعداد زیادی از لایه‌های پنهان و گره‌ها، ممکن است شبکه، خطای تصادفی موجود در داده‌های تحت یادگیری را نیز فراگیرد (بیش‌یادگیری) و در نتیجه، به رغم دقت بالای شبکه در طبقه‌بندی داده‌های تحت یادگیری، در خصوص داده‌های تحت آزمون به خوبی عمل نکند. مهم‌ترین ضعف شبکه‌های عصبی مصنوعی به عنوان یک روش جعبه سیاه، عدم تفسیرپذیری وزن‌ها، پیچیدگی و زمان‌بر بودن روش و الگوریتم محاسبات است و بنابراین برای اجرای مناسب آن باید از نرم‌افزارهای آماری مرتبط با داده‌کاوی استفاده کرد (۴۴-۱۹، ۴۰-۱۷).

در مطالعه حاضر، براساس متغیرهای سن مادر، مرتبه بارداری، فاصله بارداری فعلی از بارداری پیشین و داشتن فرزند زنده پسر و با بهره‌گیری از مدل رگرسیون لجستیک، تحلیل تشخیصی، درخت تصمیم‌گیری و شبکه‌های عصبی مصنوعی، به پیشگویی بارداری ناخواسته در شهر خرم‌آباد مبادرت شد. براساس شاخص سطح زیر منحنی مشخصه عملکرد، بهترین روش‌ها به ترتیب عبارت بودند از: شبکه‌های عصبی مصنوعی، درخت تصمیم‌گیری، رگرسیون لجستیک و سرانجام، تحلیل تشخیصی؛ البته همان‌طور که در نتایج بیان شد، تفاوت آماری معناداری میان سطوح زیر منحنی مشخصه عملکرد روش‌های نامبرده مشاهده نشد.

دلیل برتری نسبی شبکه‌های عصبی مصنوعی در پیش‌بینی بارداری خواسته و ناخواسته را باید در ماهیت فرایند یادگیری دانست چراکه براساس فرایند یادگیری،

محسوب نمی‌شود. نتایج مطالعات شبیه‌سازی شده نشان می‌دهند، زمانی که بیشتر متغیرها کیفی و دارای تعداد سطوحی اندک باشند، تحلیل تشخیصی، ضعیف‌تر از رگرسیون لوجستیک عمل می‌کند (۴۵)؛ یکی دیگر از مشکلات مربوط به تحلیل تشخیصی همانند شبکه‌های عصبی، عدم قابلیت تفسیرپذیری ضرایب تابع تشخیص است (۱۹، ۴۵-۱۷).

در مطالعه محمدپور اصل و همکاران که روی ۱۵۷۶ نفر از زنان شهر تبریز در سال ۱۳۸۳ صورت گرفت، شیوع بارداری ناخواسته، ۲۶/۷ درصد بوده است که با استفاده از روش رگرسیون لوجستیک، عوامل تأثیرگذار بر آن، سن بالای زن و تعداد فرزندان زنده بیشتر گزارش شد (۳)؛ البته در مطالعه حاضر، متغیر سن مادر با استفاده از مدل رگرسیون لوجستیک، معنادار نشد که این به دلیل انتخاب مادران مولتی‌پار در این پژوهش است که باعث محدود شدن بازه سنی مادران مورد مطالعه شده است؛ گرچه نتایج درخت تصمیم‌گیری از اهمیت بالای سن مادر در پیش‌بینی بارداری ناخواسته به‌ویژه در مادران دارای مرتبه بارداری سوم یا بالاتر و نیز دارای فرزند زنده پسر حکایت داشته است.

در مطالعه پورحیدری و همکاران که روی زنان شهر شاهرود در سال‌های ۱۳۸۳ و ۱۳۸۴ صورت گرفت، شیوع بارداری ناخواسته، ۳۱ درصد بوده است که با استفاده از رگرسیون لوجستیک، تعداد فرزند زنده بیشتر را به‌عنوان عامل خطرزا در بارداری ناخواسته معرفی کرده‌اند که البته در مطالعه، تعداد فرزند زنده پسر به‌عنوان متغیر مستقل استفاده شده است (۴).

در مطالعه امانی و همکاران که روی ۳۲۸ نفر از زنان شهر اردبیل صورت گرفت، شیوع بارداری ناخواسته، ۶۰/۷ درصد بوده است و با استفاده از روش رگرسیون لوجستیک، بارداری ناخواسته با مرتبه بارداری، ارتباطی معنادار نشان داد ولی با سن بالای مادر، ارتباطی را نشان‌داد که با نتایج رگرسیون لوجستیک مطالعه حاضر همخوانی دارد؛ گرچه متغیر سن بالای مادر در مطالعه حاضر، توسط درخت تصمیم‌گیری به‌عنوان متغیری

درخت تصمیم‌گیری نیز به‌عنوان یک روش ناپارامتری و ابزاری قدرتمند در داده‌کاوی، دقتی بالا را در پیشگویی بارداری ناخواسته نشان داده است. نتایج مطالعات شبیه‌سازی شده نشان می‌دهند که این روش، زمانی که داده‌ها کشیدگی و چولگی بالایی داشته باشند یا زمانی که درصد بالایی از متغیرها کیفی باشند (همانند مطالعه حاضر)، روشی مناسب برای داده‌کاوی است (۱۸).

مهم‌ترین مزیت این روش، قابلیت تفسیرپذیری بسیار بالای آن به‌واسطه ساختار درختی است لذا زمانی که به‌جای تمرکز بر دقت بالای روش پیشگویی، به تفسیرپذیری نتایج، علاقه‌مند باشیم، یکی از بهترین گزینه‌ها خواهد بود (۳۹-۳۴، ۱۹)؛ باوجود این، یکی از مشکلات روش درخت تصمیم‌گیری در این است که در هر مرحله از اجرای الگوریتم، تنها براساس یک متغیر تصمیم گرفته، فرایند تفکیک را انجام می‌دهد (۳۹).

مهم‌ترین مزیت روش رگرسیون لوجستیک، این است که به‌عنوان یک روش مبتنی بر مدل، قادر است میزان تأثیر هریک از متغیرهای مستقل را بر وقوع بارداری ناخواسته، مورد آزمون قرارداد، مقایسه کند؛ در نتیجه از طریق مدلی که در قالب یک فرم بسته ارائه می‌دهد، تفسیرپذیری مطلوبی دارد؛ برای نمونه، در این مطالعه می‌توان بیان کرد که فاصله بارداری زیر دو سال، مرتبه زایمان سوم یا بالاتر و نیز دارا بودن فرزند زنده پسر، شانس نسبی بارداری ناخواسته را افزایش می‌دهند (۴۵-۱۹، ۴۰-۱۷)؛ به‌رغم مزیت‌های نامبرده، مدل رگرسیون لوجستیک، به‌شدت، تحت تأثیر همبستگی بالای میان متغیرهای مستقل و وجود ارتباطات غیرخطی قرار می‌گیرد (۱۹).

در این مطالعه با وجود عدم تفاوت معنادار آماری میان سطوح زیر منحنی مشخصه عملکرد، ضعیف‌ترین نتایج پیشگویی به تحلیل تشخیصی، مربوط بوده است چراکه روشی، پارامتری است که به پذیره نرمال بودن چندمتغیره وابسته باشد و در نتیجه، مادامی که این پذیره، برقرار نباشد، روشی مناسب برای مقاصد طبقه‌بندی

در مطالعه *سادات هاشمی* و همکاران که در سال ۱۳۸۲، روی زنان شهر تهران صورت گرفت، با استفاده از روش‌های آماری، نظیر رگرسیون لجستیک، پرابیت، تحلیل تشخیصی خطی و شبکه‌های عصبی مصنوعی، تأثیر عواملی مانند سن زن، تعداد فرزند زنده دختر و تعداد فرزند زنده پسر بر بارداری ناخواسته تعیین شد؛ در نهایت، شبکه‌های عصبی مصنوعی و به‌ویژه پرسپترون سه‌لایه به‌عنوان بهترین مدل و تحلیل تشخیصی به‌عنوان بدترین روش گزارش شدند (۲۷)؛ در مطالعه حاضر نیز، قدرت پیشگویی شبکه‌های عصبی مصنوعی، بالاتر از سایر روش‌ها و روش تحلیل تشخیصی، پایین‌تر از سایر روش‌ها بوده است.

در مطالعه دیگر *سادات هاشمی* و همکاران که در سال ۲۰۰۵، روی زنان تهرانی انجام دادند، شیوع بارداری ناخواسته، ۳۱/۱ درصد تعیین شد و با به‌کارگیری شبکه‌های عصبی مصنوعی و رگرسیون لجستیک چندجمله‌ای و با در نظر گرفتن متغیرهای سن زن، تعداد فرزندان زنده، تحصیلات زن و مرد، آگاهی اولیه نسبت به مبانی تنظیم خانواده و نوع روش پیشگیری پیش از بارداری، توان پیشگویی دو روش مقایسه و روش شبکه‌های عصبی مصنوعی به‌عنوان روش بهتر پیشنهاد شد؛ البته صرف نظر از ماهیت چهارسطحی متغیر وابسته و نیز ترکیب به‌نسبت متفاوت متغیرهای مستقل، از نظر برتری شبکه‌های عصبی مصنوعی با مطالعه حاضر همخوانی دارد (۳۰).

در مطالعه *خلج‌آبادی فراهانی* و همکاران که روی ۴۱۴۱ نفر از زنان شهر تهران در سال ۱۳۷۵ انجام گرفت، شیوع بارداری ناخواسته، ۳۱/۱ درصد بیان شد که با استفاده از روش رگرسیون لجستیک چندجمله‌ای، عوامل سن بالای زن و تعداد فرزندان زنده پسر را بر وقوع بارداری ناخواسته، مؤثر دانستند (۹)؛ نتایج این مطالعه با نتایج درخت تصمیم‌گیری مطالعه حاضر، منطبق است ولی با نتایج رگرسیون لجستیک مطالعه حاضر در خصوص اثر سن بالای زن مغایرت دارد که علت این تناقض را می‌توان ناشی از گروه‌بندی متفاوت

بااهمیت تلقی شد که علت این امر را می‌توان ناشی از ماهیت متفاوت درخت تصمیم‌گیری در بررسی تأثیرگذاری متغیر سن دانست؛ به این معنا که اثرگذاری متغیر سن، تنها درخصوص مادرانی تأیید شده که در مرتبه سوم و بالاتر بارداری قرار داشته، درعین حال، دارای فرزند زنده پسر نیز بوده‌اند، نه برای تمامی مادران مورد مطالعه (۱۱).

در مطالعه‌ای که *وکیلی* و همکاران در سال ۱۳۸۹ روی زنان ساکن یزد انجام دادند، شیوع بارداری ناخواسته، حدود ۲۴/۵ درصد بود که براساس رگرسیون لجستیک، تعداد فرزندان زنده بیشتر بر وقوع بارداری ناخواسته تأثیرگذار بوده است (۲۸)؛ البته براساس نتایج حاصل از رگرسیون لجستیک و نیز درخت تصمیم‌گیری مطالعه حاضر نیز، نتیجه‌ای مشابه درخصوص تعداد فرزندان زنده پسر به‌دست آمده است.

در مطالعه *گوتو* و همکاران که در سال ۲۰۰۲، روی زنان ژاپنی انجام دادند، ۴۶/۲ درصد زنان، دارای سابقه بارداری ناخواسته بودند که با استفاده از رگرسیون لجستیک، متغیرهای مرتبه بارداری بالاتر و تعداد فرزند زنده بیشتر، رابطه معناداری با وقوع بارداری ناخواسته داشته‌اند (۶)؛ این نتیجه با یافته‌های مطالعه حاضر، به‌طور کامل همخوانی دارد، هرچند که شیوع بارداری ناخواسته در زنان ژاپنی به‌وضوح بالاتر بوده است.

در مطالعه *کلاورت* و همکاران که در سال ۲۰۱۳، روی مادران تانزانایی و با استفاده از روش رگرسیون لجستیک انجام دادند، ارتباط سن بالای مادر با وقوع بارداری ناخواسته، معنادار به‌دست آمد که البته با نتایج حاصل از رگرسیون لجستیک مطالعه حاضر همخوانی ندارد (۸)؛ علت این یافته را می‌توان در ترکیب به‌طور کامل، متفاوت متغیرهای مستقل موجود در دو مطالعه و نیز استفاده مطالعه حاضر از مادران مولتی‌پار مطرح کرد؛ باوجوداین، نتیجه مطالعه یادشده با نتیجه حاصل از روش درخت تصمیم‌گیری مبنی بر ارتباط سن بالا و وقوع بارداری ناخواسته همخوانی دارد.

حال با توجه به نقش مدل‌بندی‌های آماری در تعیین مهم‌ترین عوامل تعیین‌کننده بارداری‌های ناخواسته و نیز تمرکز مطالعه حاضر، روی مادران مولتی‌پار و همچنین با توجه به این موضوع که ماهیت تعیین‌کننده‌های بارداری ناخواسته در مادران مولتی‌پار و نولی‌پار، متفاوت است، پیشنهاد می‌شود، در مطالعه‌ای جداگانه به بررسی مقایسه‌ای نقش عوامل خطرزای مختلف در زنان دو گروه یادشده پرداخته شود.

از دیدگاه آماری نیز پیشنهاد می‌شود تا در مطالعه‌ای شبیه‌سازی‌شده، روش‌های مختلف رده‌بندی براساس معیارهایی مانند نوع و تعداد متغیرهای مستقل، توزیع آنها، حجم نمونه و نیز نوع معماری شبکه‌های عصبی مصنوعی (نظیر تعداد لایه‌ها، گره‌ها و نیز نوع توابع فعالیت) مقایسه شوند.

سپاس و قدردانی

این مقاله، حاصل بخشی از طرح تحقیقاتی با عنوان: «مقایسه شبکه‌های عصبی مصنوعی، درخت تصمیم‌گیری، تحلیل تشخیصی و رگرسیون لجستیک در پیش‌بینی بارداری ناخواسته در مادران مولتی‌پار شهر خرم‌آباد» مصوب دانشگاه علوم پزشکی ایران در سال ۱۳۹۳ است که با حمایت دانشگاه علوم پزشکی ایران اجرا شده است.

متغیر بارداری ناخواسته در دو مطالعه دانست.

در مطالعه *فقیه‌زاده* و همکاران که در سال ۲۰۰۳، روی زنان تهرانی انجام دادند، شیوع بارداری ناخواسته، ۳۸/۲ درصد بود که با استفاده از مدل‌های لگ‌خطی، رابطه آماری معناداری میان متغیرهای سن بالای مادر و فاصله بارداری پایین با بارداری ناخواسته گزارش شد (۳۲)؛ این یافته با نتایج مدل رگرسیون لجستیک مطالعه حاضر درخصوص فاصله بارداری پایین همخوانی داشته، درباره اثر سن بالای مادر همخوانی ندارد که این عدم همخوانی را می‌توان ناشی از تفاوت نمونه‌های انتخابی در دو مطالعه از نظر ترکیب متغیرهای زمینه‌ای و جمعیت‌شناختی و نیز ترکیب متغیرهای مستقل مورد استفاده در مدل‌بندی دانست.

در مطالعه *فاینر و زولنا* که (۲۰۱۱)، روی زنان ایالات متحده آمریکا انجام دادند، شیوع بارداری ناخواسته، حدود ۴۸ درصد گزارش شد که البته به علت حجم بسیار بالای نمونه، هیچ‌گونه آزمون و مدل‌بندی آماری روی داده‌ها انجام نشد؛ در این مطالعه در زنان پایین‌تر از سن ۲۵ سال و نیز دارای مرتبه بارداری بالاتر، شیوع بارداری ناخواسته به‌وضوح بالاتر بود (۷). ارتباط سن پایین با افزایش شیوع بارداری ناخواسته با مطالعه فعلی تناقض دارد که دلیل این امر را می‌توان به تفاوت‌های فرهنگی مادران ایالات متحده و مادران مطالعه حاضر نسبت داد.

با توجه به شیوع به‌نسبت بالای بارداری ناخواسته در مادران مولتی‌پار شهر خرم‌آباد، مطالعات بیشتر در این زمینه و نیز بازنگری برنامه‌های تنظیم خانواده و آموزش زنان در معرض خطر، اجتناب‌ناپذیر خواهد بود. عمده محتوای آموزشی بهتر است بر انتخاب تعداد مطلوب (ایده‌آل) فرزندان و ترکیب جنسیتی آنها از دیدگاه زوجین، فاصله‌گذاری مناسب میان موالید و نیز سن مناسب بارداری برای مادران، معطوف شود.

منابع

1. Moosazadeh M, Nekoei-moghadam M, Emrani Z, Amiresmaili M. Prevalence of unwanted pregnancy in iran: a systematic review and meta analysis. *International Journal of Health Planning and Management*. 2013. Available from: <http://onlinelibrary.wiley.com/doi/10.1002/hpm.2184/full>.
2. Najafi F, Iran-far SH, Rezayi M, Iran-far KH. Systematic review and meta analysis of prevalence of unwanted pregnancy in iran, 1995-2006. *Journal of Kermanshah University of Medical Sciences*. 2011; 16(57):280-287.
3. Mohammadpoorasl A, Rostami F, Ivanbagha R, Torabi SH. Prevalence of unwanted pregnancy and multivariate analysis of its correlates in tabriz, 2004. *Medical Sciences Journal of Islamic Azad University*. 2005; 15(4):201-206.
4. Pourtaheri M, Souzaia, Shamailan N. Prevalence of unwanted pregnancies and their correlate in pregnant woman in shahroud, Iran. *Payesh*. 2007; 6(1):63-70.
5. Khoushah-Mehri G, EbrahimTaheri G, Hatami Z, Safaari M. Prevalence of unwanted pregnancy and its related factors in women referring to health centers in south of tehran. *Journal of Nursing and Midwifery Faculty, Shahid Beheshti University of Medical Sciences*. 2008; 16(59):26-32.
6. Goto A, Yasumura S, Reich MR, Fukao A. Factors associated with unintended pregnancy in Yamagata, Japan. *Social Science and Medicine*. 2002;54(7):1065-79.
7. Lawrence BF, Mia RZ. Unintended Pregnancy in the United States: Incidence and Disparities, 2006. *Contraception*. 2011; 84(5):478-485.
8. Calvert C, Baisley K, Doyle A, Maganja K, Chagalucha J, Watson-Jones D, Hayes R, Ross D. Risk factors for unplanned pregnancy among young women in tanzania. *Journal of Family Planning and Reproductive Health Care*. 2013;39:1-2.
9. Khalaj-Abadi-Faraahani F, Sadat-Hashemi SM. Factors influencing unwanted pregnancies in Tehran. *Hakim*. 2002; 5(3):201-206.
10. Kazemi SA, Kousha A, Taddayon P, Mousavi-nasab N. The causes of unwanted pregnancy in 500 pregnant women which referring to Zanjan hospitals. *Journal of Zanjan University of Medical Sciences*. 2000; 37:39-45.
11. Amani F, Bashiri J, Nahan-Moghadam N, Tabraei Y. Application of logistic regression model in surveying effective causes of unwanted pregnancy. *Qom University OF Medical Sciences Journal*. 2010; 4(1):32-36.
12. Shahbazi A, Ghorbani R, Akbari-far M. A survey on the prevalence of unwanted pregnancy and some related factors in pregnant women who referred to the medical laboratories in Semnan. *Koomesh*. 2006; 7(3-4 (20)):133-137.
13. Mansouri A, Hosseini Sh, Dadgar S. Unexpected pregnancy and relative factors in pregnant women referring to mashhad maternity wards in 2004. *journal of birjand university of medical sciences*. 2009;16(38):65-71.
14. Jessica D. Gipson, Michael A. Koenig, Michelle J. Hindin. The effects of unintended pregnancy on infant, child, and parental health: a review of the literature. *Studies in Family Planning*. 2008; 39(1):18-38.
15. Enayati M, Abd-Alrahimi F. A Comparison of Mental Health and Marital Satisfaction Between 'Wanted' and 'Unwanted' Pregnancy Women in Ahwaz. *New Findings in Psychology*. 2008; 2(6):66-80.
16. Bayat H, Asefzadeh S. Prevalence of unintended pregnancies and its relationship with low birth weight. *Journal of Birjand University of Medical Sciences*. 2004; 11(1):28-31.
17. Dreiseitl S, Ohno-Machado L. Logistic regression and artificial neural network classification models: a methodology review. *Journal of Biomedical Informatics*. 2002;35:352-359.
18. Paul R. Harper A. A Review and comparison of classification algorithms for medical decision making. *health policy*. 2005;71:315-331.
19. Melody Y. A comparative assessment of classification methods. *Decision Support Systems*. 2003;35:441-454.

20. Yassae F. Prevalence of unwanted pregnancy for women which referring to Mahdieh hospital. *Journal of Medical Research, Shahid Beheshti University of Medical Sciences*. 2001; 26(2):133-136.
21. Abazari F, Arab M, Abbasszadeh A. Relationship of unwanted pregnancy and fertility behavior in pregnant women who visited maternity wards of Kerman hospitals. *Journal of Reproduction and Infertility*. 2003;4(13):39-46.
22. Robabi H, Asan-sarani H, Azarkish F, Dastfan Z, Dashipor A. The survey of factors associated with unwanted pregnancy among women referring to health care centers of Iranshahr in 2007. *Iranian Journal of Obstetrics, Gynecology and Infertility*. 2011;714(4):32-40.
23. Rakhshani F, Ansari Moghaddam AR, Tehrani H. Prevalence of unwanted pregnancy and associated factors in Zahedan, 1999. *Journal of Research in Medical Sciences*. 2003; 8(3):40-43.
24. Khalili M, Shooohani B, Soltani O, Pour-Najaf A. A study of factors effecting unwanted pregnancy among the women referring to Illam health centers, 2003. *Journal of Illam University of Medical Sciences*. 2004; 12(42-43):18-24.
25. Yazdani F. Comparison of the some factors in women with unwanted and planned pregnancy. *Family Health*. 2012;1(2):19-26.
26. Najafian M, Karami KB, Cheraghi M, Jafari RM. Prevalence of some factors relating with unwanted pregnancy, in Ahwaz, Iran, 2010. *International Scholarly Research Network*. 2011;4:30-34.
27. Sadat-Hashemi M, Kazemnejad A, Kavehei B. Use of various artificial neural networks for prediction of unwanted pregnancy and their comparison using traditional statistical methods. *Journal of Shahid Sadoughi University of Medical Sciences and Health Services*. 2003; 11(1):10-15.
28. Vakili M, Shahbazi H, Dehghani MH. The prevalence of unintended pregnancies and its related demographic factors in hospitals of Yazd city, 2008. *Toloo-e-Behdasht*. 2010;9(4):23-36.
29. Noroozi A, Khoram-Roodi Rozita, Sharifi Sh, Tahmasebi R. Prevalence of unwanted pregnancy and its related factors in the women covered by health centers in Bushehr province in 2003. *Iranian South Medical Journal*. 2005; 8(1):83-89.
30. Sadat-Hashemi M, Kazemnejad A, Lucas C, Badie K. Predicting the type of pregnancy using artificial neural networks and multinomial logistic regression: a comparison study. *Neural Computing and Applications*. 2005;14(3):198-202.
31. Nouhjah S, Niakan Kalhori SH. Artificial neural networks application to predict type of pregnancy in women equal or greater than 35 years of age. *International Journal of Computer and Information Technology*. 2014;3(1):177-182.
32. Faghihzadeh S, Babae G, Lmyian M, Mansourian F, Rezasoltani P. Factors associated with unwanted pregnancy. *Journal of Sex and Marital Therapy*. 2003;29(2):157-64.
33. Kitsantas P, Hollande Mr, Li L. Using classification trees to assess low birth weight outcomes. *Artificial Intelligence in Medicine*. 2006;38:275-289.
34. Edrisi M, Gharipour M, Faroughi A, Javeri F, Shahgholi B, Gharipour A. Decision support in prediction of metabolic syndrome with data mining methods. *Journal of Jahrom University of Medical Sciences*. 2011(9)2:48-58.
35. Toloiee-Ashlaghi A, Pourebrahimi A, Ebrahimi M, Ghasem-ahmad L. Using data mining techniques for prediction breast cancer recurrence. *Iranian Journal of Breast Disease*. 2013;5 (4):23-34.
36. Sepehri MM, Rahnama P, Shadpour P, Teimourpour B. A data mining based model for selecting type of treatment for kidney stone patients. *Tehran University Medical Journal*. 2009;67(6):421-7.
37. Keshtkar A, Majd-Zadeh SR, Mohammad K, Ramezan-Zadeh F, Borna S, Azemi-Khah A, Hosseini F. Determination of effective factors on preeclampsia severity: the application of classification and regression trees. *Journal of Gorgan University of Medical Sciences*. 2006; 8(2):47-54.
38. Saki M, Hajizade E, Fatemi R. Evaluation of prognostic variables for classifying the survival in colorectal patients using the decision tree. *Iranian Journal of Epidemiology* 2012;8(2):13-19.
39. Hosseini M, Tazhibi M, Amini M, Zareei AS, Jahani-Hashemi H. Using classification tree for prediction of diabetic retinopathy on type ii diabetes. *journal of isfahan medical school*. 2010;28(104):15-24.
40. Tu JV. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *Journal of Clinical Epidemiology*. 1996;49(11):1225-31.

41. Kim Y. Comparison of the decision tree, artificial neural network, and linear regression methods based on the number and types of independent variables and sample size. *Expert Systems with Applications*. 2008;34:1227-1234.
42. Parsaeian M, Mohammad K, Mahmoudi M, Zeraati H. Comparison of logistic regression and artificial neural network in low back pain prediction: second national health survey. *Iranian Journal of Public Health*. 2012;41(6):86-92.
43. Kazemnejad A, Sadat-Hashemi M, Faghihzadeh S, Kavehei B. Using artificial neural networks as statistical tools for analysis of medical data. *Journal of Research in Medical Sciences*. 2003;8(2):35-39.
44. Biglarian A, Babayi R, Azmi R. Application of artificial neural networks for determining the mortality risk factors in hospitals and comparing with logistic regression model. *Modares Medical Journal*. 2004;7(1):23-29.
45. Pohar M, Blas M, Turk S. Comparison of logistic regression and linear discriminant analysis: a simulation study. *Metodološki zvezki*. 2004;1(1):143-161.

Daneshvar

Medicine

*Scientific-Research
Journal of Shahed
University
22nd Year, No.116
April- May, 2015*

Received: 07/03/2015

Last revised: 11/04/2015

Accepted: 15/04/2015

Comparison of neural networks, decision trees, discriminant analysis and logistic regression for predicting unwanted pregnancy of multiparous women in Khorramabad

Farzad Ebrahimzadeh^{1,2}, Farid Zayeri³, Nasim Vahabi², Ali Azarbar⁴, Katayoun Bakhtiyar⁵, Agha-Fatemeh Hosseini^{6*}

1. Department of Public Health, Faculty of Health and Nutrition, Lorestan University of Medical Sciences, Khorramabad, Iran.
2. Department of Biostatistics, Faculty of Medicine, Tarbiat Modarres University, Tehran, Iran.
3. Associate Professor - Department of Biostatistics, Member of Proteomics Research Center, School of Paramedical Sciences, Shahid Beheshti, University of Medical Sciences, Tehran, Iran.
4. Department of Statistics, Faculty of Sciences, Amirkabir University of Technology, Tehran, Iran.
5. Department of Public Health, Faculty of Health and Nutrition, Lorestan University of Medical Sciences, Khorramabad, Iran.
6. Department of Biostatistics, Faculty of Health, Iran University of Medical Sciences, Tehran, Iran.

E-mail: hosseini_af@yahoo.com

Abstract

Background and Objective: Unwanted pregnancy is a pregnancy that is considered to be unwanted by at least one member of the couple, and has adverse consequences for the family and community. Using four classification models, this study predicted unwanted pregnancy in the urban population of Khorramabad and compared these classification models.

Materials and methods: In this cross-sectional study, 467 multiparous pregnant women attending the urban health care centers of Khorramabad in 2011 were selected using stratified and cluster sampling and risk factors were collected. The logistic regression model, discriminant analysis, decision trees, and CART artificial neural networks, along with the SPSS and MATLAB software were applied in data modeling. The indices of sensitivity, specificity, area under the ROC curve, and accuracy rate were applied to compare the models.

Results: The prevalence of unwanted pregnancy was 32.3%. Based on the index of area under the ROC curve, ROC the best models were artificial neural networks (0.741), decision tree (0.731), logistic regression (0.712) and discriminant analysis (0.711). The highest sensitivity was found for decision tree model (73.5%), and the highest specificity was for artificial neural network (62.3%).

Conclusion: Given the relatively high prevalence of unwanted pregnancy in Khorramabad, the revision of the family planning programs seems to be inevitable. Moreover, in selecting the best classification method, decision tree and logistic regression are recommended when the researcher is interested in better interpretability of the results, and the model of neural networks is recommended when a higher prediction power is intended.

Key words: Neural networks, Decision Tree, Logistic regression, Discriminant analysis, Multiparous women.