پیش‌بینی PCOS از طریق مدلینگ هوش مصنوعی با نشانگرهای زیستی: یک مطالعه اولیه برای تجویز تمرینات ورزشی در زنان ایرانی

زهرا عرب طاهری‌زاده[۱]، دکتر ولی‌الله دبیدی روشن[۲،۳*]، دکتر مرضیه میرزایی[۴]

۱. دانشجوی دکتری فیزیولوژی ورزشی، دانشکده علوم ورزشی، دانشگاه مازندران، بابلسر، ایران.
۲. استاد فیزیولوژی ورزشی، دانشکده علوم ورزشی، دانشگاه مازندران، بابلسر، ایران.
۳. استاد فیزیولوژی ورزشی، مرکز تحقیقات سلامت و عملکرد ورزشی، دانشکده علوم ورزشی، دانشگاه مازندران، بابلسر، ایران.
۴. استادیار گروه زنان و مامایی، دانشکده پزشکی، دانشگاه آزاد اسلامی تهران، تهران، ایران.

تاریخ دریافت: ۱۴۰۴/۰۸/۰۱ **تاریخ پذیرش:** ۱۴۰۴/۱۱/۰۶

خلاصه

مقدمه: فرزندآوری و جوان‌سازی جامعه، از چالش‌های جدی دنیای امروزی است. سندرم تخمدان پلی‌کیستیک (PCOS) که با چاقی مرتبط می‌باشد، به یک نگرانی رایج برای باروری تبدیل شده است. تشخیص زودهنگام PCOS می‌تواند روند درمانی و کیفیت زندگی خانواده را بهبود بخشد. مطالعه حاضر با هدف اولیه آموزش مدل‌های یادگیری ماشین با داده‌های نشانگرهای هورمونی و سندرم متابولیک برای شناسایی PCOS و سپس ارائه راهکارهای تمرینات ورزشی در زنان نابارور انجام شد.

روش‌کار: داده‌های ۱۰۰۰ زن نابارور ۴۵-۲۰ ساله مبتلا و غیرمبتلا به PCOS تحت پردازش اولیه قرار گرفت و در نهایت داده‌های ۵۰۰ نفر، به روش‌های مختلف مانند جنگل تصادفی (RF)، حذف بازگشتی ویژگی (RFE) و اطلاعات متقابل (MI)، ویژگی‌های مهم انتخاب و وارد الگوریتم‌های مختلف شد.

یافته‌ها: آموزش مدل درخت تصمیم (DT) با انتخاب ویژگی به روش RF، بالاترین دقت را در تشخیص نسبت به سایرین داشت. همچنین با توجه به نتایج، انتخاب ویژگی به روش RFE، MI و RF نشان داد که علاوه بر هورمون‌های جنسی، قند ناشتا، کلسترول، لیپوپروتئین با چگالی بالا، ویتامین D3 و هورمون‌های مرتبط با تیروئید، عوامل برجسته‌ای هستند که در پیش‌بینی با دقت بالای PCOS نقش دارند.

نتیجه‌گیری: تشخیص PCOS و بررسی اهمیت بیومارکرهای مرتبط با آن با هوش مصنوعی نشان داد، شاخص‌های متابولیکی نیز در کنار هورمون‌های جنسی از اهمیت ویژه‌ای در تشخیص و درمان این بیماری برخوردار می‌باشند.

کلمات کلیدی: تمرینات ورزشی، سندرم پلی‌کیستیک تخمدان، سندرم متابولیک، ناباروری، هوش مصنوعی

* **نویسنده مسئول مکاتبات:** دکتر ولی‌الله دبیدی روشن؛ مرکز تحقیقات سلامت و عملکرد ورزشی، دانشکده علوم ورزشی، دانشگاه مازندران، بابلسر، ایران. **تلفن:** ۰۱۱-۳۵۳۰۲۰۰۰؛ **پست الکترونیک:** v.dabidi@umz.ac.ir

مقدمه

صنعتی شدن جوامع در طی چند دهه اخیر، موجب گسترش بی‌تحرکی و از این رو، بروز چاقی و مشکلات وابسته به آن شده است (۱). اگرچه بافت چربی با ترشح موادی از قبیل آدیپوکاین‌ها، در فرآیندهای مختلف نقش دارد (۲)، اما مطالعات نشان می‌دهد چربی اضافی در زنان از طریق مکانسیم‌هایی منجر به عدم تخمک‌گذاری، اختلال پروتئین‌های متصل شونده به هورمون‌های جنسی و سطوح بالای آندروژن شده که با کاهش تعداد فولیکول‌ها و سطح پروژسترون، منجر به ناباروری می‌شوند (۳). ناباروری به‌عنوان یکی از مهم‌ترین مسائل بهداشتی در جهان محسوب می‌شود (۴) و سندرم پلی‌کیستیک تخمدان (PCOS)[1] یا همان تنبلی تخمدان، به‌عنوان یکی از علل اصلی ناباروری در زنان شناخته شده است (۵). با توجه به برنامه‌های کلان ملی و بین‌المللی برای جوان‌سازی جامعه و فرزندآوری، هر رویکردی که موجب کنترل چاقی و مشکلات وابسته به آن شود، تأثیر بسزایی در کاهش هزینه‌های دولت مردان و خانواده‌ها خواهد داشت. در این راستا، در کنار هر تشخیص بالینی که ممکن است پرهزینه، سخت و پرخطا باشد، رویکردهای جدید مانند فناوری‌های مبتنی بر یادگیری ماشینی، برای تشخیص و تعیین شاخص‌های جدید مرتبط با بیماری‌ها به‌سرعت در حال رشد است تا هر بیماری به‌ویژه سندرم تخمدان پلی‌کیستیک را زودهنگام و با سرعت و دقت بالا تشخیص دهد.

یادگیری ماشینی، یکی از شاخه‌های هوش مصنوعی می‌باشد که در بسیاری از رشته‌ها از جمله پزشکی، جایگاه ویژه‌ای پیدا کرده است. این تکنولوژی با تکیه بر الگوریتم‌ها و آموزش دیدن با استفاده از انبوهی از داده‌ها، قادر به یادگیری بوده و می‌تواند در بسیاری از موارد، به‌خصوص تشخیص و پیش‌بینی بیماری تصمیم‌گیری کند (۶)، استفاده از الگوریتم‌های مرتبط با انتخاب ویژگی می‌تواند ارتباط شاخص‌های کمتر شناخته شده را با بیماری موردنظر نشان داده و ابعاد دیگری از عوامل تأثیرگذار بر بیماری را مشخص می‌کند. این به نوبه خود، زمینه را برای تحقیقات در مسیرهای جدید درمان و اصلاح سطوح شاخص‌های جدید مرتبط با بیماری‌ها به‌ویژه سندرم تخمدان پلی‌کیستیک مشخص می‌کند. بدین‌منظور مطالعاتی در حیطه تعیین ویژگی‌های مرتبط با بیماری‌های دیگر و آموزش الگوریتم برای تشخیص آن انجام شده است؛ از جمله در مطالعه طهماسبی و همکاران (۲۰۲۲) به تشخیص بیماری‌های کبدی توسط یادگیری ماشینی بر اساس ویژگی‌های تعیین شده توسط مدل پرداخته شد و الگوریتم پیشنهادی آموزش دیده توانست با دقت ۹۱٪ این بیماری را تشخیص دهد (۷). در مطالعه‌ای دیگر در سال ۲۰۱۹ به تشخیص بیماری‌های قلبی توسط الگوریتم‌های یادگیری ماشینی پرداخته شد و مدل‌های آموزش دیده با شاخص‌های تعیین شده توانستند این بیماری‌ها را با دقت بالا تشخیص بدهند (۸). همچنین در مطالعه ال وی و همکاران (۲۰۲۲) آموزش الگوریتم‌های هوش مصنوعی به‌وسیله تصاویر چشم توانست بیماری سندرم تخمدان پلی‌کیستیک را دقت بالایی تشخیص دهد (۹). در مطالعه تاکره و همکاران (۲۰۲۰)، پس از بررسی شاخص‌های مرتبط با این بیماری، ۳۰ ویژگی با بالاترین قدرت تشخیص توسط الگوریتم‌های هوش مصنوعی انتخاب شدند که پس از آموزش و ارزیابی مدل‌های مختلف یادگیری ماشین، جنگل تصادفی بالاترین و معتبرترین مدل برای تشخیص این بیماری بود (۱۰). دنی و همکاران (۲۰۱۹)، از ۲۳ ویژگی حاصل از نتایج آزمایشات بالینی و متابولیک، ۸ ویژگی بالقوه را برای آموزش دیدن الگوریتم یادگیری ماشینی انتخاب کردند و پس بررسی‌های لازم، جنگل تصادفی دقیق‌ترین روش برای پیش‌بینی سندرم تخمدان پلی‌کیستیک با دقت ۸۹/۰۲٪ انتخاب شد (۱۱). با این وجود، با توجه به ضرورت سالمند شدن بسیاری از کشورها و الزام دولت‌ها بر سیاست جوان‌سازی در جامعه، مطالعات محدودی به بررسی پیش‌بینی سندرم تخمدان پلی‌کیستیک به‌ویژه در زنان چاق نابارور ایرانی پرداختند.

[1] Polycystic ovary syndrome

اگرچه مصرف داروهای شیمیایی به‌عنوان یکی از روش‌های درمان بسیاری از بیماری‌ها مؤثر می‌باشد، اما افزایش نگرانی بابت عوارض جانبی داروهای شیمیایی، تمایل به مصرف این داروها و پایبندی در مصرف را کاهش داده و در مقابل بهره‌گیری از راهکارهای غیردارویی نظیر اجرای تمرینات منظم ورزشی در کنترل و درمان این بیماری را در طی سال‌های اخیر افزایش داده است (۱۲، ۱۳)؛ به‌عنوان مثال در مطالعه ریبرو و همکاران (۲۰۲۰)، که تأثیر دو نوع تمرین هوازی متناوب و مداوم را به‌مدت ۱۶ هفته بر پارامترهای هورمونی، متابولیک، ترکیب بدن و شاخص‌های آنتروپومتریک این بیماران مورد ارزیابی قرار دادند، کاهش شاخص‌های آنتروپومتریکی و هیپرآندروژنیسم در هر دو گروه تمرینی مشاهده شد. به همین ترتیب، اجرای تمرین هوازی متناوب باعث کاهش شاخص آندروژن و بهبود پروفایل لیپیدی شد (۱۴). در مطالعه ال عیسی و همکاران (۲۰۱۷)، که بر روی سه گروه سنی همسان انجام شد، ۱۲ هفته تمرین هوازی با شدت متوسط از طریق تعدیل چربی، سطوح آدیپونکتین، هورمون آنتی‌مولرین و هورمون‌های باروری، بر عملکرد تولید مثل تأثیر مثبت و معنی‌داری داشت (۱۵). علی‌رغم مطالعات مذکور، از حیث تعیین مؤثرترین فاکتورها در تشخیص و بهبود این بیماری، اگرچه مطالعاتی به بررسی ویژگی‌های مرتبط با این بیماری و آموزش مدل‌هایی برای تشخیص آن پرداخته‌اند (۱۶-۱۸)، اما تاکنون مطالعه‌ای به بررسی داده‌های مرتبط با شاخص‌های متابولیکی و هورمون‌های مرتبط با تولید مثل به‌منظور آموزش الگوریتم‌های هوش مصنوعی و ایجاد مدل هوش مصنوعی به‌ویژه در زنان ایرانی نپرداخته است.

با توجه به اهمیت تشخیص سندرم تخمدان پلی‌کیستیک به‌عنوان یکی از اصلی‌ترین علل ناباروری و همچنین ارتباط چاقی با ناباروری از یک‌سو و از سوی دیگر لزوم استفاده از روش‌های نوین از قبیل هوش مصنوعی با هدف تشخیص سریع‌تر و دقیق‌تر برای اثرگذاری مطلوب و انتخاب درمان‌های جایگزین داروهای شیمیایی و یا رویکردهای غیردارویی، مطالعه حاضر درصدد اجرای پژوهشی با هدف جمع‌آوری و تدوین داده‌های اولیه مربوط به سندرم تخمدان پلی‌کیستیک زنان نابارور ایرانی به جهت آموزش الگوریتم‌های یادگیری ماشینی می‌باشد تا بستری برای محققان حوزه‌های پزشکی و ورزشی جهت ارائه درمان‌های دارویی و غیردارویی و بهترین استراتژی تمرینی (تمرین هوازی، تمرینات تناوبی با شدت بالا، تمرینات مقاومتی و یا تمرینات ترکیبی) و سایر رویکردهای سبک سالم زندگی برای محققان آتی فراهم آورد. بنابراین هدف از اجرای مطالعه حاضر آن بود که آیا با استفاده از نشانگرهای زیستی تهاجمی می‌توان وقوع سندرم تخمدان پلی‌کیستیک را در زنان ایرانی پیش‌بینی نمود؟، در صورت پاسخ مثبت، کدام یک از شاخص‌های زیستی با دقت بالاتری می‌تواند این سندرم را در زنان ایرانی پیش‌بینی کند. به علاوه با توجه به فقدان منابع مرتبط به‌ویژه در ارتباط با تأثیر فعالیت منظم بدنی بر تغییرات شاخص‌های زیستی، هدف دوم مطالعه حاضر، ارائه توصیه‌های سبک زندگی مرتبط با فعالیت بدنی در این دسته از زنان است.

روش‌کار

۱- طرح تحقیق و ملاحظات اخلاقی

طرح تحقیق حاضر از نوع مطالعات کاربردی بود که به پیش‌بینی وقوع سندرم تخمدان پلی‌کیستیک بر اساس داده‌های هورمونی بیوشیمیایی و متابولیکی می‌پردازد. پروتکل این طرح تحقیق توسط گروه فیزیولوژی دانشگاه تأیید و در کمیته اخلاق زیستی بررسی و پس از تأیید با کد اخلاق IR.UMZ.REC.1402.037 اجرا شد. بر اساس طرح تحقیق حاضر، محققان متعهد شدند که تمام ملاحظات اخلاقی از جمله حفظ اطلاعات فردی آزمودنی‌ها را در تمام مراحل اجرای تحقیق مورد توجه قرار دهند.

۲- نحوه جمع‌آوری داده‌ها و معیار روتردام

پس از هماهنگی اولیه و انعقاد تفاهم‌نامه بین دانشگاه و مراکز آموزشی- پژوهشی ناباروری، داده‌های ۵۰۰ بیمار نابارور مبتلا به سندرم تخمدان پلی‌کیستیک و غیرمبتلا جمع‌آوری و در اختیار محققان قرار گرفت. از

این تعداد، ۲۵۰ داده مربوط به بیماران سندرم تخمدان پلی‌کیستیک و ۲۵۰ داده مربوط به افراد غیرمبتلا به سندرم تخمدان پلی‌کیستیک بود. تعداد داده‌ها بر اساس مطالعات قبلی محققان و با توجه به قوانین موجود در هوش مصنوعی انتخاب شد (۱۹). برای تشخیص سندرم تخمدان پلی‌کیستیک از معیار روتردام استفاده شد. این معیار، یکی از معتبرترین معیارهای تشخیص این بیماری محسوب می‌شود و بر اساس آن در صورت داشتن ۲ مورد از ۳ گزینه: ۱. هایپرآندروژنیسم ۲. تخمدان پلی‌کیستیک ۳. اختلالات تخمک‌گذاری، احتمال ابتلاء به سندرم پلی‌کیستیک تخمدان مشخص می‌شود (۲۰) و به‌منظور تأیید نهایی این بیماری، آزمایشات تکمیلی مانند تیروئید، هایپرپرولاکتمیا و غیره نیاز بود. مطابق با این معیار، ۴ فنوتایپ برای این بیماری در نظر گرفته می‌شود. A: هایپرآندروژنیسم بالینی/ بیوشیمیایی + تخمدان پلی‌کیستیک؛ B: هایپرآندروژنیسم بالینی/ بیوشیمیایی + اختلالات تخمک‌گذاری؛ C: اختلالات تخمک‌گذاری + تخمدان پلی‌کیستیک؛ D: هایپرآندروژنیسم بالینی/ بیوشیمیایی + اختلالات تخمک‌گذاری + تخمدان پلی‌کیستیک (۲۱).

۳- پیش‌پردازش داده‌ها

شکل ۱، مراحل اجرای فرآیندهای پیش‌پردازش و یادگیری ماشین را در این مطالعه نشان می‌دهد. با توجه به اینکه برخی الگوریتم‌ها به نرمالیته و داده‌های پرت حساس می‌باشند، به‌منظور افزایش دقت و بهبود نتایج الگوریتم‌های یادگیری ماشینی، داده‌ها در مرحله اول پردازش شد و مقادیر گمشده، اشتباه و پرت با استفاده از نمودار جعبه‌ای شناسایی و بسته به نوع ویژگی، با استفاده از روش‌های آماری با میانگین، میانه یا مد جایگزین شدند. بر اساس مطالعات قبلی، در صورتی که شاخص موردنظر بیشتر از ۳۰٪ داده گمشده داشته باشد، حذف و در غیر این‌صورت برای داده‌های پیوسته از میانگین و در صورتی که شاخص به‌صورت طبقه‌بندی باشد، از مد استفاده می‌شود. همچنین نظر به اینکه برخی ویژگی‌ها پزشکی می‌باشند، چنانچه داده‌ای در درون یک کلاس، پرت بود، اما در محدوده پزشکی قرار داشت، از تغییر آن صرف‌نظر کرده، زیرا انتظار می‌رود در تعیین عوامل خطرزای بیماری کمک و به ساخت مدلی دقیق‌تر منجر شود (۸، ۱۹).

شکل ۱- فلوچارت تحقیق

۴- انتخاب متغیرهای اصلی مطالعه

انتخاب ویژگی بهینه و مناسب در یادگیری ماشینی می‌تواند به کاهش پیچیدگی و افزایش دقت مدل منجر شود (۱۹). همچنین با توجه به اینکه شناسایی و انتخاب ویژگی‌های مرتبط با بیماری سندرم تخمدان پلی‌کیستیک، از اهداف اصلی این مطالعه می‌باشد، از سه روش انتخاب ویژگی در یادگیری ماشینی شامل فیلتر (Filter methods)، دسته‌بندی (Wrapper methods) و تعبیه (Embedded method) استفاده شد تا مدل‌ها بر اساس ویژگی‌های مرتبط، آموزش دیده و ارزیابی شوند. بدین‌منظور در ابتدا ۲۱ متغیر شامل اندازه‌گیری‌های انجام شده توسط آزمایشگاه و مرکز رادیولوژی، با توجه به مشورت پزشکان و متخصصان مربوطه شامل سن (Age)، قد (Height)، وزن (Weight)، قاعدگی منظم (Mense Regular)، کیست تخمدان‌ها (Ovary.CYST)، هایپرآندروژنیسم (Hyperandrogenism)، آکنه (Acne)، هیرسوتیسم (Hirsutism)، هورمون آنتی‌مولرین (AMH)، محرک فولیکولی (FSH)، لوتئیزه کننده (LH)، محرک تیروئید (TSH)، پرولاکتین (Prolactin)، آنتی‌بادی ضد تیروپراکسیداز (Anti-TPO)، شاخص قندخون ناشتا (FBS)، هموگلوبین گلیکوزیله (HbA1)، کلسترول (Cholesterol)، تری گلیسیرید (TG)، لیپوپروتئین با چگالی پایین (LDL)، لیپوپروتئین با چگالی بالا (HDL) و ویتامین د۳ (VITD3) جمع‌آوری شد. اطلاعات مربوط به متغیرهای جمع‌آوری شده به‌منظور مدلینگ هوش مصنوعی در جدول ۱ نمایش داده شده است.

جدول ۱- اطلاعات مربوط به آمار توصیفی برخی ویژگی‌های استفاده شده در مجموعه داده‌ها

ویژگی	میانگین	انحراف معیار	حداقل	٪۲۵	٪۵۰	٪۷۵	حداکثر
سندرم تخمدان پلی‌کیستیک (دارد/ ندارد)	--	--	۰	--	--	--	۱
سن (سال)	۳۴/۷۷	۶/۳۶	۲۰	۳۰	۳۵	۴۰	۴۵
وزن (کیلوگرم)	۶۸/۹۲	۱۳/۷۵	۴۲	۶۰	۶۷	۷۶	۱۶۰
قد (سانتی‌متر)	۱۶۰	۵/۱۶	۱۴۱	۱۵۸	۱۶۱	۱۶۳	۱۷۵
کیست تخمدان (دارد/ ندارد)	--	--	۰	--	--	--	۱
قاعدگی منظم (دارد/ ندارد)	--	--	۰	--	--	--	۱
هایپرآندروژنیسم (دارد/ ندارد)	--	--	۰	--	--	--	۱
آکنه (دارد/ ندارد)	--	--	۰	--	--	--	۱
هیرسوتیسم (دارد/ ندارد)	--	--	۰	--	--	--	۱
آنتی‌مولرین (نانوگرم/ میلی‌لیتر)	۴/۹۶	۴/۱	۰/۰۱	۲	۳/۹	۷/۶	۱۹/۶
هورمون محرک فولیکول (میلی‌واحد بین‌المللی در هر میلی‌لیتر)	۸/۹	۱۴/۲۸	۱	۴/۷۹	۵/۷	۷/۲	۹۶
هورمون لوتئینی (میلی‌واحد بین‌المللی در هر میلی‌لیتر)	۹/۳۲	۸/۱۲	۰/۱	۴/۹	۸/۴	۱۰/۳	۹۶
پرولاکتین (نانوگرم/ میلی‌لیتر)	۲۳/۳	۴۲/۴	۰/۶	۱۰/۰۷	۱۵/۶۸	۲۲/۸۵	۴۶۲
هورمون محرک تیروئید (میلی‌واحد بین‌المللی در هر لیتر)	۲/۳۶	۱/۷۵	۰/۰۱	۱/۴	۲/۱	۲/۸	۲۳
آنتی‌بادی ضد تیروپراکسیداز (واحد بین‌المللی در هر میلی‌لیتر)	۵۵/۸۴	۳۰۰/۷۴	۰/۰۸	۵/۰۰	۱۸/۵	۴۴	۷۳۰۱
هموگلوبین A1c (درصد)	۵/۴۳	۰/۵۸	۴/۳	۵/۱	۵/۴	۶	۸
قندخون ناشتا (میلی‌گرم/ دسی‌لیتر)	۹۵/۱۶	۱۰/۶	۶۸	۹۰	۹۴	۹۸	۱۸۴
تری‌گلیسیرید (میلی‌گرم/ دسی‌لیتر)	۱۲۹/۸۶	۵۲/۱۶	۱۶	۹۷/۷۵	۱۲۷	۱۵۱/۲۵	۳۶۶
کلسترول (میلی‌گرم/ دسی‌لیتر)	۱۷۹/۲۹	۳۱/۵	۳۰	۱۶۵	۱۷۵	۱۹۲	۳۰۷
لیپوپروتئین با چگالی کم (میلی‌گرم/ دسی‌لیتر)	۱۰۴/۰۱	۲۷/۱۴	۱۳	۹۱	۱۰۵	۱۱۷	۲۱۹
لیپوپروتئین با چگالی بالا (میلی‌گرم/ دسی‌لیتر)	۴۷/۲۷	۹/۹۳	۱۷	۴۲	۴۶	۵۲	۱۰۷
ویتامین D3 (نانوگرم/ میلی‌لیتر)	۵۲/۹۸	۱۲/۴۳	۴	۱۹	۲۴	۲۹	۱۲۴

۴،۱ روش فیلتر (Filter)

در این روش بر اساس روش‌های آماری، به ویژگی‌ها امتیاز داده شده و ویژگی‌ها مستقل از الگوریتم‌ها انتخاب می‌شوند (۲۲). از جمله الگوریتم‌های موجود در این روش می‌توان اطلاعات متقابل (MI)، مربع کای دو و ضریب همبستگی را نام برد.

۴،۲ روش دسته‌بندی (Wrapper)

رویکرد دسته‌بندی از روش فیلتر دقیق‌تر می‌باشد، اما از لحاظ محاسباتی، سنگین‌تر است (۲۳). در این روش زیرمجموعه‌های مختلف از ویژگی‌ها انتخاب و مدل به‌وسیله آنها آموزش داده می‌شود. ویژگی‌ها تا زمانی که مدل به یک عملکرد خوب برسد، حذف و یا اضافه می‌شود. در نهایت بهترین زیرمجموعه از ویژگی‌ها برای آموزش مدل اصلی انتخاب می‌شود (۱۹). از جمله الگوریتم‌های موجود در این روش حذف بازگشتی ویژگی (RFE)[1] و انتخاب ویژگی با روش جستجوی رو به جلو (SFS)[2] می‌باشد.

۴،۳ روش تعبیه (Embedded)

در این روش، مدل انتخابی برای آموزش با استفاده از ویژگی‌های موجود آموزش دیده و بهترین ویژگی به جهت دستیابی به بهترین عملکرد مدل در حین آموزش انتخاب می‌شود. این روش، سریع‌تر از روش دسته‌بندی می‌باشد و نسبت به رویکرد فیلتر، از دقت بالاتری برخوردار می‌باشد (۱۹). از جمله مدل‌های قابل استفاده در این روش، درخت تصمیم می‌باشد.

در این مطالعه به‌منظور استفاده از هر سه رویکرد، از روش‌های اطلاعات متقابل (MI)، حذف ویژگی‌ها به‌صورت بازگشتی (RFE) و جنگل تصادفی (RF) استفاده شد.

۵- پیاده‌سازی و اجرای مدل

در این مطالعه از نرم‌افزار Navigator 2.6.3 Anaconda و Jupyter 7.0.8 استفاده شد. دیتاهای بیماران برای آموزش مدل‌ها به دو دسته آموزش و تست به نسبت ۸۰ به ۲۰ تقسیم شد. سپس دیتاهای آموزش برای آموزش مدل‌ها استفاده و دیتاهای تست برای ارزیابی مدل‌های آموزش دیده به‌کار گرفته شد. همچنین از مدل‌های مختلف یادگیری ماشینی مانند بیز ساده[3]، K- نزدیک‌ترین همسایه[4] و درخت تصمیم[5] استفاده شد. همچنین مدل‌های ترکیبی در بهبود عملکرد مدل‌های یادگیری ماشینی تأثیر به‌سزایی دارد که از جمله این روش‌ها می‌توان به بگینگ[6]، استکینگ[7]، بوستینگ[8] اشاره کرد که از تکنیک‌های شناخته شده در مدل‌های ترکیبی می‌باشد. در این مطالعه از یک ترکیبی به روش استکینگ شامل سه الگوریتم ماشین‌بردار پشتیبانی (SVM)[9]، جنگل تصادفی[10] و آدابوست (AdaBoost) نیز استفاده شد.

۶- ارزیابی مدل

به‌منظور بررسی کارایی مدل‌ها از معیارهای ارزیابی مانند حساسیت[11]، دقت[12]، ویژگی[13] و صحت[14] استفاده شد. معیار حساسیت، نسبت افراد درست بیمار تشخیص داده شده به کل افراد بیمار، معیار اختصاصیت، نسبت افراد درست سالم تشخیص داده شده به کل افراد سالم و دقت، بیانگر میزان پیش‌بینی درست این بیماری توسط مدل می‌باشد که توسط ماتریکس تداخل[15] حساب می‌شود.

یافته‌ها

۱- نتایج روش‌های انتخاب ویژگی

در این بخش نتایج سه رویکرد انتخاب ویژگی بر روی ۲۱ ویژگی جمع‌آوری شده برای بیماران سندرم تخمدان پلی‌کیستیک ارائه می‌شود.

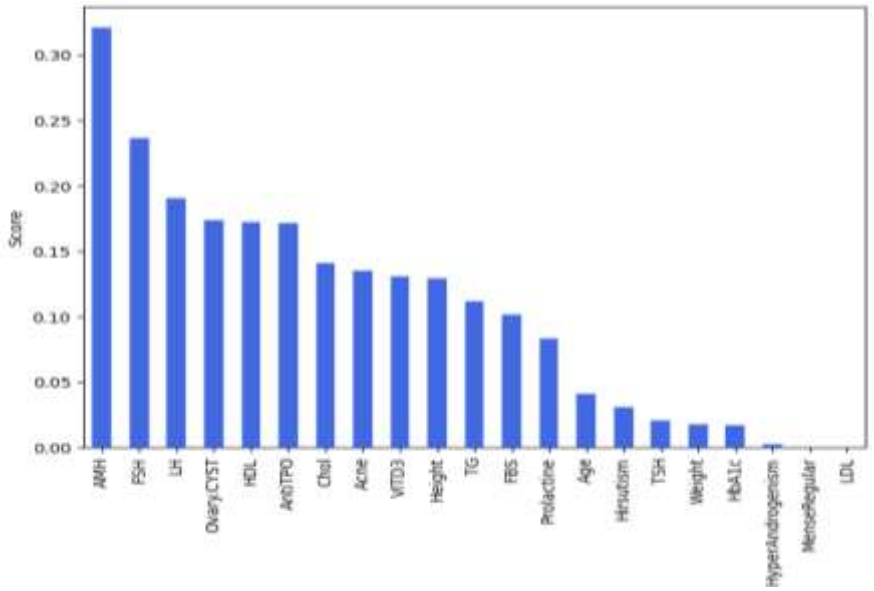
۱،۱ امتیازدهی ویژگی‌ها به روش MI

بعد از اجرای MI بر روی داده‌های مربوط به بیماران سندرم تخمدان پلی‌کیستیک، نتایج امتیازات مرتبط با

[1] Recursive feature elimination
[2] Sequential feature selection

[3] Naive Bayes-NB
[4] K-Nearest Neighbors-KNN
[5] Decision tree-DT
[6] Bagging
[7] Stacking
[8] Boosting
[9] Support vector machines
[10] Random forest-RF
[11] Sensitivity
[12] Accuracy
[13] Specificity
[14] Precision
[15] Confuse matrix

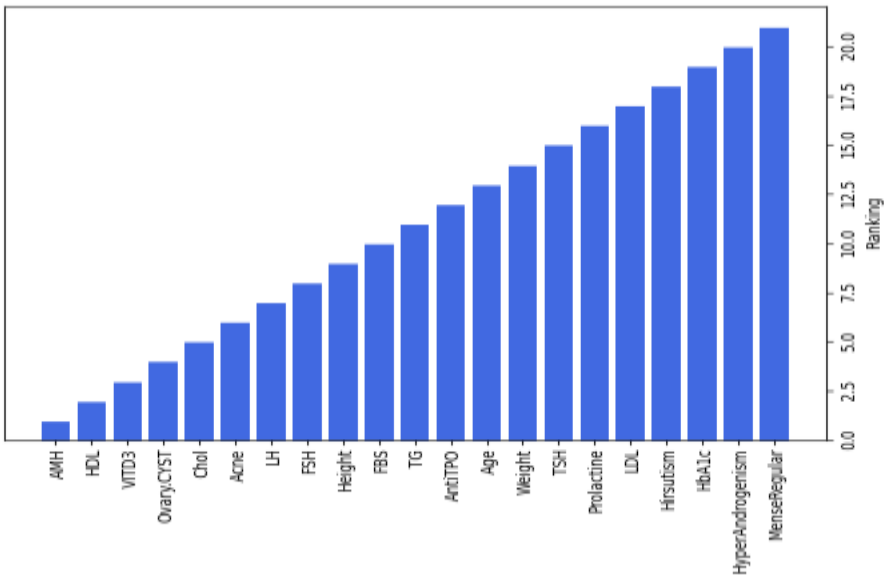
هر متغیر در شکل ۲ نشان داده شد. مطابق با این نمودار، هورمون AMH با امتیاز ۰/۳۴ بالاترین و ویژگی‌های FSH، LH، Ovary.Cyst، HDL، Anti TPO، Chol، Acne، VITD3 و Height در مرتبه ۲ تا ۱۰ قرار گرفته و جزء ۱۰ متغیر با بالاترین امتیاز رده‌بندی شدند. پایین‌ترین امتیازات به LDL و Mense Regular داده شد.



شکل ۲- نتایج مربوط به امتیازدهی ویژگی‌های مربوط به بیماران سندرم تخمدان پلی‌کیستیک به روش MI

۱,۲ رتبه بندی ویژگی‌ها به روش RFE

شکل ۳ رتبه‌بندی ویژگی‌های مرتبط با مجموعه داده جمع‌آوری شده بیماران سندرم تخمدان پلی‌کیستیک با رویکردWrapper و با استفاده از روش RFE را نشان می‌دهد. در این روش، رتبه کمتر نشان‌دهنده اهمیت بیشتر ویژگی می‌باشد، لذا AMHدارای رتبه اول و ویژگی‌های HDL، VITD3، Ovary.CYST، Chol، Acne، LH، FSH، Height و FBS در رتبه‌های بعدی قرار گرفتند. Mense Regularرتبه آخر را شامل شد.



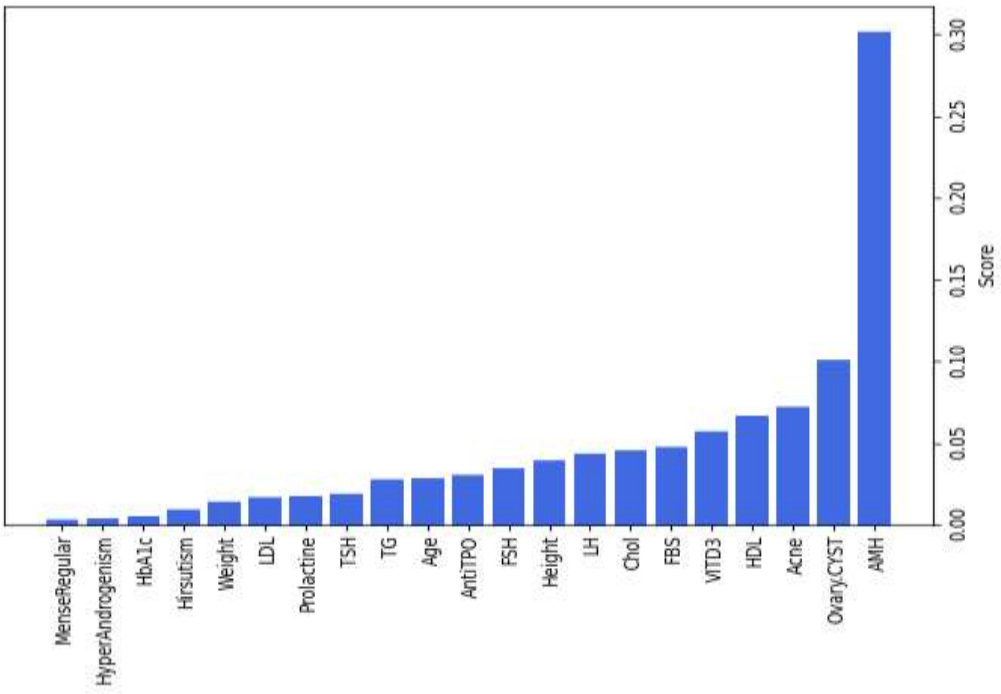
شکل ۳- نتایج مربوط به رتبه‌بندی ویژگی‌های مربوط به بیماران سندرم تخمدان پلی‌کیستیک به روش RFE

۱,۳ امتیازدهی ویژگی‌ها با استفاده از RF

شکل ۴ ویژگی‌های انتخاب شده بر پایه RF را نشان می‌دهد. مهم‌ترین ویژگی هورمون AMH با امتیاز ۰/۲۸۲ بود و سپس Ovary.CYST، Acne، HDL، VITD3، FBS، Chol، LH، Height و FSH جزء ۱۰ ویژگی با بالاترین امتیاز قرار گرفتند. همچنین Mense Regular کمترین امتیاز را به‌خود اختصاص داد.

۲- ارزیابی مدل‌های آموزش‌دیده با ویژگی‌های منتخب

نتایج معیارهای ارزیابی پس از آموزش مدل‌های یادگیری ماشینی با استفاده از ویژگی‌های انتخاب شده به‌وسیله هر سه رویکرد موجود در جدول ۳ نشان داده شده است.



شکل ۴- نتایج مربوط به رتبه‌بندی ویژگی‌های مربوط به بیماران سندرم تخمدان پلی‌کیستیک به روش RF

جدول ۲- ویژگی‌های انتخاب شده با سه رویکرد انتخاب ویژگی

شماره	الگوریتم انتخاب ویژگی	تعداد	اسامی ویژگی‌های انتخاب شده
۱	MI	۱۰	AMH, LH, FSH, Ovary.Cyst, HDL, Anti TPO, Chol, Acne, VITD3, Height
۲	RFE	۱۰	AMH, HDL, VITD3, Ovary.CYST, Chol, Acne, LH, FSH, Height, FBS
۳	RF	۱۰	AMH, Ovary.CYST, Acne, HDL, VITD3, FBS, Chol, LH, Height, FSH

MI: اطلاعات متقابل، RFE: حذف ویژگی‌ها به‌صورت بازگشتی، RF: جنگل تصادفی

جدول ۳- مقایسه نتایج عملکرد مدل‌های یادگیری ماشینی توسط رویکردهای انتخاب ویژگی

انتخاب ویژگی	مدل	دقت (درصد)	صحت (درصد)	حساسیت (درصد)	ویژگی (درصد)
MI	KNN	۸۸	۹۴	۹۴	۸۲
	NB	۹۰	۹۱	۸۹	۹۰
	DT	۸۸	۸۹	۸۶	۸۹
	Stacking	۹۰	۹۰	۹۲	۸۹
RFE	KNN	۹۰	۸۶	۹۷	۸۴
	NB	۸۸	۸۴	۹۵	۸۱
	DT	۹۱	۸۷	۹۵	۸۷
	Stacking	۹۰	۹۱	۹۱	۹۰
RF	KNN	۹۰	۸۷	۹۵	۸۵
	NB	۸۷	۸۳	۹۴	۷۹
	DT	۹۲	۹۲	۹۳	۹۱
	Stacking	۹۱	۸۷	۹۸	۸۴

MI: اطلاعات متقابل، RFE: حذف ویژگی‌ها به‌صورت بازگشتی، RF: جنگل تصادفی

در مجموع نتایج نشان داد در انتخاب ویژگی به روش MI، مدل‌های NB و Stacking بالاترین دقت و در رویکرد انتخاب ویژگی به روش RFE، مدل‌های KNN، DT و Stacking دقت بالاتری را از خود نشان دادند. همچنین در انتخاب ویژگی به روش RF مدل DT، Stacking و KNN بالاترین دقت را در تشخیص داشتند.

بحث

مطالعه حاضر در زمره نخستین پژوهش‌هایی است که در آن مدلینگ یادگیری ماشینی (هوش مصنوعی) جهت پیش‌بینی سندرم تخمدان پلی‌کیستیک با استفاده از نشانگرهای زیستی در زنان چاق نابارور ایرانی و با هدف فراهم‌سازی زمینه‌های پژوهشی برای محققان حوزه علوم ورزشی برای تجویز رویکردهای غیردارویی از قبیل تمرینات ورزشی به این دسته زنان پرداخت. مهم‌ترین یافته مطالعه حاضر این بود که چاقی از طریق تغییرات در ویژگی‌های هورمونی و متابولیکی، در بروز ناباروری زنان ایرانی اثرگذار است و از این‌رو به‌عنوان یک عامل مداخله‌گر عمل می‌کند. در این راستا، یادگیری ماشین بر اساس دیتاست‌های هورمونی و متابولیکی نشان داد که علاوه بر هورمون‌های جنسی، قند خون ناشتا، هورمون‌های مربوط به تیروئید و همچنین شاخص‌های ویتامین D3، کلسترول، لیپوپروتئین با چگالی بالا و آنتی‌بادی ضد تیروپراکسیداز نیز از اهمیت زیادی در پیش‌بینی سندرم تخمدان پلی‌کیستیک برخوردار هستند و بهترین دسته از ویژگی‌ها در تشخیص این بیماری، ویژگی‌های انتخاب شده به روش RF و مدل آموزش دیده DT می‌باشد. با توجه به اثربخشی تأیید شده تمرینات منظم هوازی و بی‌هوازی بر هورمون‌های تولید مثلی (۲۴-۲۸) و شاخص‌های سندرم متابولیکی (۲۴، ۲۹-۳۴) می‌توان فرض کرد که به‌کارگیری اینگونه رویکرد غیردارویی، دارای ظرفیتی برای بهبود وضعیت باروری در زنان سندرم تخمدان پلی‌کیستیک ایرانی می‌باشد که مستلزم مطالعات بعدی خواهد بود. نتایج مطالعه حاضر بینش جدیدی در مورد شناسایی بالقوه زنان مبتلا به سندرم تخمدان پلی‌کیستیک با استفاده از هوش مصنوعی و مدلینگ یادگیری ماشین ارائه می‌دهد و ممکن است اهمیت پیش‌بینی زودهنگام و نظارت دقیق بر زنان درگیر را آشکار کند و دریچه‌ای را برای بکارگیری راهکارهای غیردارویی از قبیل فعالیت ورزشی جهت تغییر سودمند در هورمون‌های جنسی (به‌ویژه AMH، LH، FSH) و سندرم متابولیک

(Chol، HDL، FBS) در زنان جوان دارای اضافه وزن و چاق بگشاید.

همسو با یافته‌های مطالعه حاضر، شواهد متعدد قبلی مؤید آن است که چاقی مرتبط با سندرم تخمدان پلی‌کیستیک اغلب بر اساس اندازه‌گیری شاخص توده بدنی و یا شاخص‌های چاقی مرکزی تشخیص داده می‌شود؛ به‌گونه‌ای که برخی محققان، ضرب‌المثل پیدایش با الویت مرغ یا تخم‌مرغ را برای روابط سندرم تخمدان پلی‌کیستیک با چاقی به‌کار می‌برند (۳۵). افزایش چاقی، با افزایش غلظت آندروژن و بدتر شدن اختلال قاعدگی در زنان و نوجوانان مبتلا به سندرم تخمدان پلی‌کیستیک مرتبط است؛ به همین ترتیب، سندرم تخمدان پلی‌کیستیک و چاقی هر دو با مقاومت به انسولین و التهاب مزمن درجه پایین مرتبط هستند (۳۲). به‌علاوه، چاقی یک عنصر اصلی در ایجاد مقاومت به انسولین است که با عدم تخمک‌گذاری، به‌عنوان یک عامل با سندرم تخمدان پلی‌کیستیک مرتبط است. چاقی بیشتر با غلظت آندروژن بیشتر در افرادی که سندرم تخمدان پلی‌کیستیک ندارند نیز مرتبط است (۳۴). زنان مبتلا به سندرم تخمدان پلی‌کیستیک در بارداری و حفظ بارداری مشکل دارند، زیرا افزایش سطح آندروژن در بدن منجر به عدم تخمک‌گذاری و آمنوره ثانویه می‌شود. گفته شده است که زنان مبتلا به سندرم تخمدان پلی‌کیستیک سیکل قاعدگی بیش از ۴۰ روز دارند و اغلب به‌جای خونریزی قاعدگی مناسب، خونریزی ناگهانی را تجربه می‌کنند. این ارتباط تا حدی با تأثیر مضر چاقی بر غلظت گلوبولین متصل به هورمون جنسی (SHBG[1]) توضیح داده می‌شود، زیرا سطوح پایین‌تر SHBG، فراهمی زیستی تستوسترون را افزایش می‌دهد. از سوی دیگر، چاقی با سطوح بالاتر تستوسترون کل مرتبط است (۳۶) که به‌معنای سنتز آندروژن بیشتر است.

مطالعات دیگری نیز به تعیین شاخص‌های جدید مرتبط با این بیماری به‌منظور تشخیص و درمان بیماری سندرم تخمدان پلی‌کیستیک پرداخته‌اند. به‌عنوان نمونه، در سال ۲۰۱۹ در پژوهشی به تعیین شاخص‌ها برای شناسایی سندرم تخمدان پلی‌کیستیک با استفاده از فنون یادگیری ماشین پرداختند. این محققان به غیر از شاخص‌های مربوطه، ویژگی‌هایی نظیر فشار خون، انجام تمرینات ورزشی و مصرف فست‌فود را نیز در آموزش مدل و تشخیص این بیماری مورد استفاده قرار دادند (۱۱). مطالعه دیگری نیز ویژگی‌های مربوط به نوع گروه خون، میزان ویتامین D3 و قند خون تصادفی را در کنار سایر شاخص‌های شناخته شده به‌منظور آموزش مدل‌ها استفاده کرد (۱۰). در مطالعه حاضر دقت بالا در الگوریتم‌های آموزش داده شده توسط ویژگی‌های انتخاب شده نظیر FBS، HDL و Chol نشان می‌دهد هرگونه بهبود در این دسته ویژگی‌ها در این بیماران می‌تواند در فرآیند کنترل و درمان این بیماری مؤثر باشد، لذا از آنجایی که روش‌های غیردارویی در سال‌های اخیر مورد توجه قرار گرفته است، انتخاب روش‌های تأثیرگذار بر این شاخص‌ها می‌تواند از جمله رویکردهای درمانی محسوب شود.

موضوع دیگری که ارزش بحث دارد، تأثیر فعالیت بدنی در هموئوستاز پارامترهای مختلف است. تاکنون ارتباط سندرم تخمدان پلی‌کیستیک و چاقی با سندرم متابولیک و التهاب قبلاً توسط محققان مختلف گزارش شده است (۳۲). در همین راستا، مطالعات انجام شده مؤید آن است که اجرای انواع تمرینات ورزشی از طریق کاهش وزن و چربی شکمی می‌تواند اثرات مثبتی بر ویژگی‌های متابولیکی بیماران مبتلا به سندرم تخمدان پلی‌کیستیک داشته باشد (۲۶، ۲۷). به‌علاوه، اجرای تمرینات منظم ورزشی از طریق تأثیر بر هورمون‌های استرسی از قبیل ACTH، CRF و کورتیزول باعث بهبود هورمون‌های تولید مثل و وضعیت تخمک‌گذاری در مبتلایان به سندرم تخمدان پلی‌کیستیک شود (۳۷، ۳۸). در مطالعه جوهام و همکاران (۲۰۲۴) تمرینات هوازی متوسط ۸ هفته‌ای، موجب بهبود حساسیت به انسولین و آمادگی قلبی متابولیک در ۱۱ زن مبتلا به سندرم تخمدان پلی‌کیستیک در مقابل ۱۰ گروه کنترل با سن و وزن مشابه شد (۳۲). یو و همکاران (۲۰۲۳)

[1] sex hormone-binding globulin

یک مطالعه مقطعی بر روی ۹۴۷ زن دارای اضافه وزن مبتلا به سندرم تخمدان پلی‌کیستیک انجام دادند که بر اساس شاخص توده بدنی (BMI) به‌عنوان دارای اضافه وزن (شاخص توده بدنی بیشتر یا مساوی ۲۴ کیلوگرم بر متر مربع) یا غیر اضافه وزن (شاخص توده بدنی کمتر یا مساوی ۲۳/۹ کیلوگرم بر متر مربع) طبقه‌بندی شدند. افزایش شاخص‌های مقاومت به انسولین و شیوع بیشتر آکانتوزیس نیگریکانس هر دو در گروه دارای اضافه وزن مشاهده شد. بین دو گروه (اضافه وزن و غیر اضافه وزن) از نظر هیپرآندروژنیسم بالینی و بیوشیمیایی تفاوت معنی‌داری وجود نداشت (۳۹). با توجه به اینکه از یک سو، ویژگی‌هایی از قبیل ترکیب بدنی، چربی خون و شاخص‌های هورمونی و سونوگرافی به‌عنوان عوامل مؤثر در این بیماری توسط الگوریتم‌های هوش مصنوعی تشخیص داده شده است و از سوی دیگر، فعالیت منظم بدنی، اثرات غیرقابل انکاری بر پارامترهای فوق دارد، لذا بر اساس پیشنهاد سازمان جهانی بهداشت برای کاهش چربی خون، بهبود مقاومت به انسولین و وضعیت تخمک‌گذاری در زنان مبتلا به سندرم تخمدان پلی‌کیستیک توصیه می‌شود تمرینات هوازی ۵-۳ جلسه در هفته و به‌مدت ۶۰-۳۰ دقیقه در هر جلسه با شدت متوسط تا شدید (۷۵-۶۵٪ ضربان قلب بیشینه) را برای یک دوره حداقل ۱۲ هفته اجرا نمایند (۴۰). در صورتی که فرد آسیب‌پذیر یا مبتدی بوده و همچنین با توجه به اینکه افراد دارای اضافه وزن عموماً بی‌تحرک هستند، لازم است تمرینات از شدت‌های کم (۶۰-۳۵٪ ضربان قلب بیشینه) شروع و پس از سازگاری به مرور شدت افزایش یابد (۴۱).

نقطه قوت و محدودیت‌های پژوهش

برخلاف سایر پژوهش‌های قبلی، در این پژوهش تلاش شد آموزش مدلینگ یادگیری ماشین ابتدا با جمع‌آوری و ایجاد یک دیتاست واقعی از ۵۰۰ زن ایرانی مبتلا و غیرمبتلا به سندرم تخمدان پلی‌کیستیک انجام شود. از دیگر نقاط قوت این مطالعه، جمع‌آوری داده‌های متعدد از قبیل پارامترهای سندرم متابولیکی و هورمونی و تولید مثلی زنان با سندرم تخمدان پلی‌کیستیک در برابر زنان غیر سندرم تخمدان پلی‌کیستیک بود. از آنجایی که چاقی و افزایش چربی به‌عنوان یکی از عوامل مؤثر در این بیماری می‌باشد، می‌توان از مهم‌ترین مزیت این دیتاها، وجود آزمایشات مربوط به چربی خون و شاخص‌های مربوط به مقاومت به انسولین در کنار شاخص‌های آنتروپومتریکی در این بیماران را نام برد که با تداوم مطالعات آتی در این حوزه می‌تواند رویکردهای درمانی همه‌جانبه و کم‌هزینه را پیشنهاد دهد.

علی‌رغم موارد مذکور، همانند سایر مطالعات، محدودیت‌هایی در این مسیر وجود داشت که می‌تواند کانون توجه محققان آتی قرار گیرد. برای نمونه، هزینه‌بر بودن آزمایشات و تست‌ها و همچنین عدم همکاری مراجعه‌کنندگان سبب می‌شد تا اندازه‌گیری دیگر شاخص‌های مؤثر بر این بیماری با مشکل مواجه شود. از سوی دیگر، نبودِ سیستم رجیستری کامل و بی‌نقص با برچسب‌گذاری بر داده‌های مربوط به این بیماری به‌منظور ثبت مشخصات مربوط به این دسته از بیماران، منجر به زمان‌بر بودن جمع‌آوری داده‌ها می‌شد. با توجه به اینکه بیماری سندرم تخمدان پلی‌کیستیک در زمره بیماری‌های چندعاملی و پیچیده می‌باشد، لذا پیشنهاد می‌شود داده‌های مربوط به ویژگی‌هایی که هزینه‌بر نبوده، مانند شاخص‌های آنتروپومتریکی و پرسشنامه‌های روحی - روانی در کنار شاخص‌های هورمونی و متابولیکی جمع‌آوری و توسط هوش مصنوعی برای تعیین و تشخیص عوامل با همبستگی بالا با این بیماری بررسی شود تا مسیرهای کم‌هزینه‌تر و درمان سریع‌تر برای این بیماری مشخص شود.

نتیجه‌گیری

یافته‌های حاصل از مدلینگ یادگیری ماشین در پیش‌بینی سندرم تخمدان پلی‌کیستیک در زنان چاق نابارور ایرانی نشان داد که شاخص‌های AMH، FSH، LH و Anti-TPO در این دسته از زنان چاق با سندرم تخمدان پلی‌کیستیک نسبت به زنان غیرسندرم تخمدان پلی‌کیستیک متفاوت بوده و DT آموزش دیده شده با انتخاب ویژگی به روش RF، بالاترین دقت را در

تشخیص این بیماری دارد. مطالعه تحقیقاتی مذکور به جامعه پزشکی جهت تشخیص و درمان زودهنگام سندرم تخمدان پلی‌کیستیک کمک می‌کند. به‌علاوه، ظرفیتی را برای محققین علوم ورزشی جهت نرمال‌سازی بیومارکرهای هورمونی و متابولیک با رویکردهای نوین تمرینی فراهم می‌کند.

تشکر و قدردانی

این مقاله برگرفته از رساله دکتری گرایش فیزیولوژی ورزشی می‌باشد. بدین‌وسیله از مساعدت دکتر نسیم طبیب‌نژاد در مرکز ناباروری پژوهشکده علوم تولید مثل یزد و مرکز خدمات تندرستی دکتر روشن به‌خاطر همکاری در اجرای این طرح، تشکر و قدردانی می‌شود.

ملاحظات اخلاقی

این پژوهش در کمیته اخلاق زیستی با کد اخلاق IR.UMZ.REC.1402.037 تأیید شد. تمام ملاحظات اخلاقی مطابق با مقرارت و دستورالعمل‌های مربوطه انجام شد. از آنجایی که داده‌های بیماران ناشناس و با کد دریافت شد، نیاز به رضایت‌نامه آگاهانه نبود.

تضاد منافع

بنابر اظهار نویسندگان این مقاله، تعارض منافعی وجود نداشت.

حمایت مالی

این مطالعه تحت حمایت مادی بنیاد ملی علم ایران (INSF) برگرفته شده از طرح شماره ۴۰۳۷۴۱۱ انجام شده است.

مشارکت نویسندگان

دکتر دبیدی روشن به‌عنوان نویسنده مسئول، طراحی و نظارت بر طرح را بر عهده داشتند. پیش‌نویس‌ها را مرور و نسخه نهایی را تأیید کردند. دکتر میرزایی و خانم عرب طاهری در تهیه و جمع‌آوری داده‌ها مشارکت داشتند. خانم عرب طاهری در اجرای طرح، تحلیل داده‌ها با مشاوره دکتر میرازیی نقش داشتند. دکتر دبیدی روشن و خانم عرب طاهری در نگارش مقاله مشارکت داشتند. تمام نویسندگان نسخه نهایی را مطالعه و تأیید کردند.

منابع

1. Dabidi Roshan V, Noorbakhsh S, Ziatabar Ahmadi SF, Nikseresht M. The physiological impacts of the low- and moderate-volume-HIIT with or without nano-curcumin supplement in women in menopause with obesity. Advances in Public Health 2024; 2024(1):5243088.
2. Chen W, Pang Y. Metabolic syndrome and PCOS: pathogenesis and the role of metabolites. Metabolites 2021; 11(12):869.
3. Gautam D, Purandare N, Maxwell CV, Rosser ML, O'Brien P, Mocanu E, et al. The challenges of obesity for fertility: A FIGO literature review. International Journal of Gynecology & Obstetrics 2023; 160:50-5.
4. Obeagu EI, Njar VE, Obeagu GU. Infertility: Prevalence and consequences. Int J Curr Res Chem Pharm Sci 2023; 10(7):43-50.
5. Gnoth C, Godehardt E, Frank-Herrmann P, Friol K, Tigges J, Freundl G. Definition and prevalence of subfertility and infertility. Human reproduction 2005; 20(5):1144-7.
6. Sajda P. Machine learning for detection and diagnosis of disease. Annu. Rev. Biomed. Eng 2006; 8(1):537-65.
7. Tahmasbi H, Besharati R, Alishahi M. A Model for Diagnosing Liver Diseases using “Machine Learning” Techniques. Studies in Medical Sciences 2022; 33(11):814-22.
8. Veisi H, Ghaedsharaf HR, Ebrahimi M. Improving the performance of machine learning algorithms for heart disease diagnosis by optimizing data and features. Soft computing journal 2021; 8(1):70-85.
9. Lv W, Song Y, Fu R, Lin X, Su Y, Jin X, et al. Deep learning algorithm for automated detection of polycystic ovary syndrome using scleral images. Frontiers in Endocrinology 2022; 12:789878.
10. Thakre V, Vedpathak S, Thakre K, Sonawani S. PCOcare: PCOS detection and prediction using machine learning algorithms. Biosci Biotechnol Res Commun 2020; 13(14):240-4.
11. Denny A, Raj A, Ashok A, Ram CM, George R. i-hope: Detection and prediction system for polycystic ovary syndrome (pcos) using machine learning techniques. InTENCON 2019-2019 IEEE Region 10 Conference (TENCON) 2019: 673-678.
12. Calcaterra V, Magenes VC, Massini G, De Sanctis L, Fabiano V, Zuccotti G. High fat diet and polycystic ovary syndrome (PCOS) in adolescence: an overview of nutritional strategies. Nutrients 2024; 16(7):938.

13. Ray NM. From Nature To Nurture: A Review of Herbal Approaches to Polycystic Ovary Syndrome (PCOS). InObstetrics & Gynaecology Forum 2024; 34(3).
14. Ribeiro VB, Kogure GS, Lopes IP, Silva RC, Pedroso DC, de Melo AS, et al. Effects of continuous and intermittent aerobic physical training on hormonal and metabolic profile, and body composition in women with polycystic ovary syndrome: A randomized controlled trial. Clinical endocrinology 2020; 93(2):173-86.
15. Al-Eisa E, Gabr SA, Alghadir AH. Effects of supervised aerobic training on the levels of anti-Mullerian hormone and adiposity measures in women with normo-ovulatory and polycystic ovary syndrome. JPMA. The Journal of the Pakistan Medical Association 2017; 67(4):499-507.
16. Khanna VV, Chadaga K, Sampathila N, Prabhu S, Bhandage V, Hegde GK. A distinctive explainable machine learning framework for detection of polycystic ovary syndrome. Applied System Innovation 2023; 6(2):32.
17. Nandipati SC, Ying CX. Polycystic Ovarian Syndrome (PCOS) classification and feature selection by machine learning techniques. Applied Mathematics and Computational Intelligence (AMCI) 2020; 9:65-74.
18. Kumari R, Singh J, Gosain A. SmS: SMOTE-stacked hybrid model for diagnosis of polycystic ovary syndrome using feature selection method. Expert Systems with Applications 2023; 225:120102.
19. Elmannai H, El-Rashidy N, Mashal I, Alohali MA, Farag S, El-Sappagh S, et al. Polycystic ovary syndrome detection machine learning model based on optimized feature selection and explainable artificial intelligence. Diagnostics 2023; 13(8):1506.
20. Moore AM, Campbell RE. Polycystic ovary syndrome: understanding the role of the brain. Frontiers in neuroendocrinology 2017; 46:1-14.
21. Shroff R, Syrop CH, Davis W, Van Voorhis BJ, Dokras A. Risk of metabolic complications in the new PCOS phenotypes based on the Rotterdam criteria. Fertility and sterility 2007; 88(5):1389-95.
22. Parlak B, Uysal AK. A novel filter feature selection method for text classification: Extensive Feature Selector. Journal of Information Science 2023; 49(1):59-78.
23. Liu Z, Yang J, Wang L, Chang Y. A novel relation aware wrapper method for feature selection. Pattern Recognition 2023; 140:109566.
24. Almenning I, Rieber-Mohn A, Lundgren KM, Shetelig Løvvik T, Garnaes KK, Moholdt T. Effects of high intensity interval training and strength training on metabolic, cardiovascular and hormonal outcomes in women with polycystic ovary syndrome: a pilot study. Plos one 2015; 10(9):e0138793.
25. Rao M, Khan AA, Adnan QU. Effects of high-intensity interval training and strength training on levels of testosterone and physical activity among women with polycystic ovary syndrome. Obstetrics & gynecology science 2022; 65(4):368-75.
26. Ribeiro VB, Lopes IP, Dos Reis RM, Silva RC, Mendes MC, Melo AS, et al. Continuous versus intermittent aerobic exercise in the improvement of quality of life for women with polycystic ovary syndrome: A randomized controlled trial. Journal of health psychology 2021; 26(9):1307-17.
27. Kogure GS, Lopes IP, Ribeiro VB, Mendes MC, Kodato S, Furtado CL, et al. The effects of aerobic physical exercises on body image among women with polycystic ovary syndrome. Journal of affective disorders 2020; 262:350-8.
28. Lopes IP, Ribeiro VB, Reis RM, Silva RC, Dutra de Souza HC, Kogure GS, et al. Comparison of the effect of intermittent and continuous aerobic physical training on sexual function of women with polycystic ovary syndrome: randomized controlled trial. The journal of sexual medicine 2018; 15(11):1609-19.
29. Aktaş HŞ, Uzun YE, Kutlu O, Pençe HH, Özçelik F, Çil EÖ, et al. The effects of high intensity-interval training on vaspin, adiponectin and leptin levels in women with polycystic ovary syndrome. Archives of physiology and biochemistry 2022; 128(1):37-42.
30. Giallauria F, Palomba S, Maresca L, Vuolo L, Tafuri D, Lombardi G, et al. Exercise training improves autonomic function and inflammatory pattern in women with polycystic ovary syndrome (PCOS). Clinical endocrinology 2008; 69(5):792-8.
31. Vigorito C, Giallauria F, Palomba S, Cascella T, Manguso F, Lucci R, et al. Beneficial effects of a three-month structured exercise training program on cardiopulmonary functional capacity in young women with polycystic ovary syndrome. The Journal of Clinical Endocrinology & Metabolism 2007; 92(4):1379-84.
32. Joham AE, Naderpoor N, Celik O. Mechanisms involved in the development of obesity with PCOS. Frontiers in Endocrinology 2024; 15:1444299.
33. Ghafoor S. Impact of Obesity and Weight Management on Women with Polycystic Ovary Syndrome and Coexisting Obesity–A Brief Narrative Review. Actual Gynecology & Obstetrics/Aktuální Gynekologie a Porodnictví 2023; 15.
34. Aggarwal M, Chakole S. Prevalence of polycystic ovarian syndrome and its link to obesity in adolescent girls. Cureus 2023; 15(9).
35. Rosenberg SL. The relationship between PCOS and obesity: which comes first?. The Science Journal of the Lander College of Arts and Sciences 2019; 13(1):5.
36. Krakowsky Y, Conners W, Morgentaler A. Serum concentrations of sex hormone–binding globulin vary widely in younger and older men: Clinical data from a men's health practice. European Urology Focus 2019; 5(2):273-9.
37. Ruiz-González D, Cavero-Redondo I, Hernández-Martínez A, Baena-Raya A, Martínez-Forte S, Altmäe S, et al. Comparative efficacy of exercise, diet and/or pharmacological interventions on BMI, ovulation, and

hormonal profile in reproductive-aged women with overweight or obesity: a systematic review and network meta-analysis. Human reproduction update 2024; 30(4):472-87.
38. Borer KT. Advanced exercise endocrinology. Human Kinetics; 2013.
39. Yu J, Zhou Y, Ding J, Zhang D, Yu C, Huang H. Characteristics and possible mechanisms of metabolic disorder in overweight women with polycystic ovary syndrome. Frontiers in Endocrinology 2023; 13:970733.
40. O'Donoghue G, Blake C, Cunningham C, Lennon O, Perrotta C. What exercise prescription is optimal to improve body composition and cardiorespiratory fitness in adults living with obesity? A network meta-analysis. Obesity Reviews 2021; 22(2):e13137.
41. Higgins JP, Higgins CL. Prescribing exercise to help your patients lose weight. Cleve Clin J Med 2016; 83(2):141-50.

Predicting PCOS through Artificial Intelligence Modeling with Biomarkers: A Preliminary Study for Prescribing Exercise in Iranian Women

Zahra Arab Taherizadeh[1], Valiollah Dabidi Roshan[2,3*], Marzieh Mirzaei[4]

1. PhD Student in Exercise Physiology, School of Sports Sciences, University of Mazandaran, Babolsar, Iran.
2. Professor, Department of Exercise Physiology, School of Sports Sciences, University of Mazandaran, Babolsar, Iran.
3. Professor, Department of Exercise Physiology, Athletic Performance and Health Research Centre, School of Sports Sciences, University of Mazandaran, Babolsar, Iran.
4. Assistant Professor, Department of Obstetrics and Gynecology, Faculty of Medicine, Islamic Azad University of Tehran, Tehran, Iran.

Received: Oct 23, 2025 **Accepted:** Jan 26, 2026

Abstract

Introduction: Childbearing and rejuvenating society are serious challenges in today's world. Polycystic ovary syndrome (PCOS), which is associated with obesity, has become a common concern for fertility. Early diagnosis of PCOS can improve the treatment process and the quality of life of the family. The present study was conducted with aim to train machine learning models with hormonal marker data and metabolic syndrome to identify PCOS and then present exercise strategies for infertile women.

Methods: Data from 1000 infertile women aged 20-45 years with and without PCOS were initially processed, and subsequently, 500 individuals were selected using various methods, including random forests (RF), recurrent feature elimination (RFE), and interaction (MI), to identify important features and incorporate them into different algorithms.

Results: Training Decision Tree (DT) model with feature selection by the RF method had the highest accuracy in diagnosis compared to others. Also, according to the results of feature selection by RFE, MI, and RF methods, it was shown that, in addition to sex hormones, fasting glucose, cholesterol, high-density lipoprotein, vitamin D3, and thyroid-related hormones are prominent factors that play a role in predicting PCOS with high accuracy.

Conclusion: Diagnosis of PCOS and investigation of its related biomarkers with artificial intelligence showed that metabolic indicators, along with sex hormones, are of particular importance in the diagnosis and treatment of this disease.

Keywords: Artificial Intelligence, Exercise, Infertility, Metabolic Syndrome, Polycystic Ovary Syndrome

► **Please cite this article as:**
Arab Taherizadeh Z, Dabidi Roshan V, Mirzaei M. Predicting PCOS through Artificial Intelligence Modeling with Biomarkers: A Preliminary Study for Prescribing Exercise in Iranian Women. **Iran J Obstet Gynecol Infertil** 2026; 28(11):31-45. **DOI:** 10.22038/ijogi.2026.27605