طراحی سیستم تشخیصی هوشمند به‌منظور پیش‌بینی چاقی در کودکان با استفاده از عوامل شیوه زندگی، وضعیت اجتماعی– اقتصادی و اطلاعات وراثتی

مرتضی فرحی [۱]، حمیدرضا مراتب [۲]، رویا کلیشادی [۳]، محمد اسماعیل مطلق [۴]

مقاله پژوهشی

چکیده

مقدمه: چاقی یکی از مهم‌ترین مشکلات تغذیه‌ای در سراسر دنیا است. اگرچه مطالعات به غیر از مشکلات فیزیکی و حرکتی، گزارشی مبنی بر بروز بیماری‌های مزمن ناشی از چاقی در دوران کودکی ارایه نداده‌اند؛ اما چاقی در کودکی می‌تواند منجر به طیف وسیعی از مشکلات در آینده شود. در این مطالعه با هدف تشخیص زود هنگام بر آن هستیم تا با طراحی یک سیستم هوشمند احتمال وقوع چاقی را بر اساس اطلاعات اولیه‌ای بر اساس نحوه شیوه زندگی و یا سایر متغیرهای عمومی اولیه، پیش‌بینی کنیم.

روش‌ها: در این مطالعه ۹۷۹۵ نفر (۴۹/۱۷٪ پسر) درگروه‌های سنی ۶ تا ۱۸ سال براساس چهارمین فاز مطالعه ملی گسترده کاسپین، مورد بررسی قرارگرفته‌اند. متغیرهای ورودی سیستم براساس عادات تغذیه‌ای، فعالیت ورزشی، اطلاعات وراثتی، وضعیت اجتماعی و اقتصادی و سابقه چاقی و دیابت مشخص گردیده‌اند. سپس بر اساس روش‌های داده‌کاوی و هوش مصنوعی، مشکل چاقی شناسایی و بررسی شده است. روش‌های انتخاب ویژگی نیز برای بهینه‌سازی سیستم طراحی شده در نظر گرفته شده است.

یافته‌ها: عملکرد روش‌های دسته‌بندی مورد مطالعه توسط روش ارزیابی متقابل دولایه‌ای مورد بررسی قرار گرفت. با توجه به نتایج حاصل از ارزیابی، بهترین نتیجه توسط روش دسته‌بندی ماشین‌بردار پشتیبان به دست آمده است. دقت و صحت این روش شناسایی پس از انتخاب ویژگی به ترتیب ۶۳/۳ و ۸۳/۷ درصد بوده است. ویژگی‌هایی ازقبیل سن، فعالیت ورزشی، نوع تغذیه در دوران نوزادی و سابقه دیابت در خانواده به‌عنوان تأثیرگذارترین ویژگی‌ها در ایجاد روند چاقی در دو گروه دختر و پسر شناسایی شده‌اند.

نتیجه‌گیری: طراحی سیستم‌های تشخیصی هوشمند با استفاده از پارامترهایی از قبیل شیوه زندگی، وضعیت اجتماعی اقتصادی و اطلاعات وراثتی می‌تواند در پیشگیری چاقی کودکان در آینده کمک نموده و به اصلاح شیوه زندگی آنان، بیانجامد. طراحی این سیستم‌های هوشمند تحت وب، می‌تواند خدمات پیشگیری را برای خانواده‌ها به راحتی ممکن سازد..

واژه‌های کلیدی: داده‌کاوی؛چاقی؛روش‌های دسته‌بندی؛سیستم تشخیصی؛ هوش مصنوعی

ارجاع: فرحی مرتضی، مراتب حمیدرضا، کلیشادی رویا، مطلق محمد اسماعیل. **طراحی سیستم تشخیصی هوشمند به‌منظور پیش‌بینی چاقی در کودکان با استفاده از عوامل شیوه زندگی، وضعیت اجتماعی- اقتصادی و اطلاعات وراثتی.** مجله تحقیقات نظام سلامت ۱۳۹۳؛ ۱۰(۳): ۵۸۶-۵۷۱

تاریخ دریافت: ۱۳۹۳/۰۶/۲۳ **تاریخ پذیرش: ۱۳۹۳/۰۷/۱۹**

۱. دانشجو ، گروه مهندسی پزشکی، دانشکده فنی مهندسی، دانشگاه اصفهان، اصفهان، ایران
۲. استادیار گروه مهندسی پزشکی، دانشکده فنی مهندسی، دانشگاه اصفهان، اصفهان، ایران
۳. استاد، دانشگاه علوم پزشکی اصفهان، مرکز تحقیقات رشد و نمو کودکان، اصفهان، ایران (نویسنده مسؤول)
Email: roya_kelishadi@gmail.com
۴. متخصص اطفال-اداره کل سلامت نوجوانان،جوانان ومدارس- وزارت بهداشت،درمان وآموزش پزشکی،تهران

[Downloaded from hsr.mui.ac.ir on 2025-07-09]

مقدمه

مشکلات مرتبط با شاخص‌های آنتروپومتریک، اعم از افزایش و یا کاهش غیرطبیعی آن از مهم‌ترین مسایل عصر حاضر به‌شمار می‌آیند. در قرن اخیر و خصوصاً چند دهه گذشته تغییرات شیوه زندگی به شکلی بوده است که شیوع چاقی بر اساس شاخص‌های متناسب آنتروپومتریک، روند رو به رشدی داشته است. مطالعات نشان می‌دهند سالیانه ۲/۸ میلیون نفر بر اثر عوارض چاقی جان خود را از دست می‌دهند (۱). تحقیقات نشان می‌دهند که به‌طور میانگین ۴۳ میلیون کودک زیر پنج سال در جهان دچار اضافه‌وزن هستند که از این میزان ۳۵ میلیون نفر در کشورهای در حال توسعه و ۸ میلیون نفر در کشورهای توسعه‌یافته زندگی می‌کنند (۲). چاقی در کودکی زمینه را برای چاقی در بزرگ‌سالی فراهم می‌کند که منجر به افزایش مشکلات سلامتی و کاهش کیفیت زندگی می‌شود (۲). متأسفانه تعریف واحدی برای بیان چاقی در دوران کودکی وجود ندارد و سازمان‌های مختلف تعاریف متعددی از قبیل: میزان وزن، تاخوردگی عضله پشت بازو، صدک وزن به قد و شاخص توده بدنی (BMI یا Body Mass Index) ارایه کرده‌اند (۳). شاخص توده بدنی معیاری بسیار مناسب و قابل‌قبول در جهت تخمین وضعیت وزنی بدن به دست می‌دهد که از تقسیم وزن بر مجذور قد به دست می‌آید (۴). وضعیت چاقی برحسب شاخص توده بدنی به‌طور مشخص بر اساس مطالعات انجام‌شده به عادت و کیفیت زندگی مرتبط است (۵).اجرای برنامه‌های پیش‌آگهی همانند پیش‌بینی بهنگام چاقی در دوران کودکی و نوجوانی همواره مورد توجه متصدیان سیستم‌های بهداشتی بوده است.

با توجه به اطلاعات منتشرشده از سازمان جهانی سلامت (WHO یا World Health Organization) چهار گروه بر اساس طبقه‌بندی شاخص توده بدنی تعریف‌شده است: افراد لاغر، افراد دارای وزن نرمال، افراد دارای اضافه‌وزن و افراد چاق. شاخص توده بدنی می‌تواند توسط عوامل مختلفی که شیوه زندگی را شکل می‌دهند، تحت‌تأثیر قرار گیرد. میزان فعالیت بدنی، نوع غذای مصرفی و وضعیت اقتصادی–اجتماعی از جمله این عوامل می‌باشند. شیوه زندگی به‌کلیه رفتارها و فعالیت‌های معمول و عرفی روزانه تلقی می‌شود؛ که توسط افراد در طول زندگی‌شان انجام می‌شود. این فعالیت‌ها می‌تواند بر سلامت افراد تأثیرگذار باشد. با انتخاب یک شیوه زندگی خاص در راستای مجموعه‌ای از اقدامات در جهت حفظ سلامت شخصی و جلوگیری از ابتلا به امراض مختلف که با انتخاب تغذیه مناسب، استراحت، ورزش کردن منظم، کنترل وزن بدن و استعمال نکردن دخانیات و الکل ایجاد می‌شود، می‌توان شیوع بسیاری از بیماری‌های مزمن را کاهش داد. این مجموعه اقدامات و تلاش‌ها شیوه زندگی فرد را شکل می‌دهد (۷–۶).

عوامل شکل‌دهنده شیوه زندگی و وضعیت اجتماعی–اقتصادی از مهم‌ترین عوامل ابتلا به چاقی هستند که بررسی آن‌ها می‌تواند سهم عمده‌ای در کاهش بروز آن داشته باشد.

میزان فعالیت بدنی ازمهم‌ترین بخش‌های شیوه زندگی است که موجب کنترل مصرف انرژی در بدن می‌شود؛ بنابراین فعالیت بدنی ازجمله مقوله‌هایی است که منجر به مصرف انرژی کنترل‌شده در جهت بهینه‌سازی عملکرد بدن می‌شود. در حقیقت چاقی نتیجه نامتعادل بودن انرژی دریافتی و مصرفی است. فعالیت بدنی موجب تقویت عضلات و تشکیل سیستم استخوان‌بندی مستحکم در بدن شده و در واقع چهارچوب کلی بدن در دوران رشد را شکل می‌دهد. مطالعات نشان داده‌اند که عدم فعالیت بدنی منجر به بروز بیشتر چاقی می‌شود. در واقع فعالیت بدنی می‌تواند یک راهکار تایید شده و مؤثر در جهت پیشگیری و درمان چاقی باشد (۹–۷).

عملکرد تغذیه‌ای به‌عنوان جزء مهم دیگری از شیوه زندگی است که می‌تواند نقش مؤثری را در احتمال ابتلا به چاقی ایفا کند (۱۰). عملکرد تغذیه‌ای را می‌توان به دو بخش نوع غذای مصرفی و نوع عادات مرتبط با خوردن تقسیم کرد. مصرف شیرینی و فست فودها به‌عنوان مثالی از نوع غذای مصرفی تأثیرگذار بر عامل چاقی و افزایش شاخص توده بدنی معرفی می‌شوند (۱۲–۱۱). پرخوری و تندخوری نیز به‌عنوان مثال‌هایی از نحوه عادات خوردن منجر به افزایش احتمال

[Downloaded from hsr.mui.ac.ir on 2025-07-09]

ابتلا به چاقی می‌شوند (۸). مطالعات بیشتر نشان داده‌اند استعمال دخانیات و مصرف الکل نیز به‌عنوان بخش‌های دیگری از شیوه زندگی، تأثیر مستقیمی در افزایش روند چاقی خواهند گذاشت (۱۱).

مرور متون در این حوزه نشان می‌دهد، برخی مشکلات اقتصادی و اجتماعی نیز می‌تواند با چاقی در ارتباط باشد. وضعیت اجتماعی– اقتصادی از طریق تأثیرگذاری بر عواملی مختلفی از قبیل درآمد تحصیلات، شغل، سطح محیط زندگی و عوامل وابسته فرهنگی می‌تواند بر بروز چاقی تأثیرگذار باشد. به‌عنوان مثال محل نامناسب زندگی می‌تواند منجر به کاهش فعالیت‌های بدنی گردد. از سویی دیگر وجود مشکلات اقتصادی و یا فقر فرهنگی می‌تواند منجر به بروز عادات نادرست تغذیه‌ای و یا مصرف مواردی گردد که بروز چاقی را افزایش دهد (۱۴–۱۳). علاوه بر آن زندگی در کنار افراد با سطح فرهنگی متفاوت، می‌تواند منجر به تأثیرپذیری نحوه شیوه زندگی از آن‌ها گردد. به‌عنوان مثال گزارش‌ها حاکی از آن است، هنگامی‌که فردی از کشوری توسعه‌یافته و دارای وضعیت مناسب اجتماعی- اقتصادی به کشورهای درحال‌توسعه مهاجرت می‌کند دچار عارضه افزایش وزن می‌شود (۱۳).

سامانه‌های هوشمند تشخیصی، فنّاوری جدیدی هستند که با توجه به ظهور دستگاه‌های کامپیوتری و دیجیتال در جهت جلوگیری از خطاها و مشکلات تشخیصی که ناشی از محدودیت‌های انسانی است، به‌صورت روزافزون مورد توجه محققین قرار گرفته‌اند (۱۵). تشخیص و پیش‌بینی در حوزه پزشکی فرایند پیچیده‌ای است که می‌تواند بخش‌های مختلف را تحت‌تأثیر قرار دهد. با توجه به پیچیدگی سامانه‌های زنده و بیولوژیکی، اطلاعات زیادی باید برای یک تصمیم‌گیری و یا پیش‌بینی درست پزشکی مدنظر قرار داده شوند و این موضوع موجب پیچیدگی این امر حتی برای افراد متخصص بالینی که به‌صورت روزانه با تعداد زیادی بیمار سروکار دارند، می‌شود. توسعه سامانه‌های هوشمند در جهت مقاصد پزشکی و بالینی، از این موضوع نشأت می‌گیرد که سیستم‌های هوشمند دارای ویژگی‌های خاصی هستند که آن‌ها را برای این کار مناسب می‌نماید. ازجمله این ویژگی‌ها می‌توان به قابلیت ذخیره و بازیابی مجدد حجم بسیار زیادی از اطلاعات در زمان بسیار کم و انجام پردازش‌های منطقی و ریاضیاتی مختلف با سرعت بسیار بالا نام برد؛ بنابراین طراحی یک سیستم هوشمند می‌تواند در زمینه دسته‌بندی اختلالات و تصمیم‌گیری بر پایه اطلاعات ورودی بسیار کارآمد باشد (۱۷–۱۶).

در این مطالعه بر آنیم تا عارضه چاقی را بر اساس ویژگی‌های تغذیه‌ای، رفتاری و وراثتی پیش‌بینی کنیم. با توجه به اینکه ویژگی‌های مذکور در هر فرد ساختار متفاوتی دارند، سیستم اقدام به تخمین وضعیت فرد از نظر دارا بودن عارضه چاقی می‌کند. بر این اساس و با توجه به نتیجه به‌دست‌آمده می‌توان با پردازش ویژگی‌های تأثیرگذار، برنامه‌ریزی مناسبی برای هر فرد تدوین نمود. این سیستم به‌نوعی امکان پردازش و تجزیه و تحلیل ویژگی‌های فردی را با توجه به نتیجه امکان‌پذیر می‌سازد، به‌عبارتی‌دیگر می‌توان با تغییر در ساختار ویژگی‌های فردی تغییر در خروجی را مشاهده کرد و به مجموعه‌ای مناسب از ویژگی‌های فردی برای دستیابی به وضعیت مناسب وزنی دست یافت. هم‌چنین از این روش هوشمند به‌منظور شناسایی حداقل پارامترهای اثرگذار عارضه موردنظر استفاده خواهد شد که می‌تواند کاربرد مؤثری در سیستم‌های پیشگیری بالینی ارایه دهد.

روش‌ها

داده‌های مورد مطالعه

مجموعه داده مورد استفاده در این مطالعه مقطعی به‌عنوان بخشی از پایش ملی دانش‌آموزان با رفتار پرخطر در چهارمین فاز از سری مطالعات دوره‌ای پایش مدرسه‌ای با نام «پایش کودکان و نوجوانان در راستای جلوگیری از بیماری‌های مزمن در افراد بالغ ایرانی» یا همان مطالعه کاسپین (Childhood and Adolescence Surveillance and Prevention of Adult Non-Communicable Disease in Iran) است. مطالعه کاسپین در فاز چهارم در سال ۱۳۹۱–۱۳۹۰ در مناطق شهری و روستایی در ۳۱ استان

[Downloaded from hsr.mui.ac.ir on 2025-07-09]

کشور ایران انجام گرفته است. روش نمونه‌برداری به‌صورت چندمرحله‌ای Cluster در ۳۰ استان ایران (۴۸ Cluster ۱۰ نفره در هر استان) بوده است. جزئیات مربوط به روش نمونه‌برداری، تخمین نمونه و همسان‌سازی در مرجع (۱۸) قابل دسترسی است.

پروتکل‌های مطالعه با توجه به قوانین موجود در بیانیه هلسینکی تنظیم شده است و توسط بورد تخصصی اخلاق پزشکی در سطوح دانشگاه و وزارت بهداشت تایید شده است (۲۰-۱۹، ۱۰). افراد پس از توضیح کامل اهداف و پروتکل‌های آزمایش به والدینشان وارد مطالعه شده‌اند. در این مطالعه، به افراد اطمینان داده شده است که پاسخ‌های آن‌ها بدون در نظر گرفتن هویتشان، بررسی می‌شود. دانش‌آموزان می‌توانستند هر زمان که بخواهند از مطالعه خارج شده بدون آن‌که جریمه‌ای پرداخت کنند. تعهدنامه کتبی و موافقت ضمنی به ترتیب از اولیا و خود دانش‌آموزان، اخذ گردید.

این مطالعه بر روی کودکان و نوجوانان دارای محدوده سنی ۶ تا ۱۸ سال انجام پذیرفته است که از بین ویژگی‌های ثبت‌شده در مجموعه‌داده ۲۳ ویژگی شامل: جنسیت، سن، وضعیت مصرف صبحانه، میزان مصرف شیرینی‌ها، سبزیجات، غذاهای آماده و فست‌فودها، نوع روغن مصرفی، میزان فعالیت بدنی، زمان خواب، میزان تماشای تلویزیون و رایانه، سابقه مصرف سیگار در خانواده و خود فرد، وضعیت دارا بودن مشکلات بی‌خوابی، افسردگی، اضطراب، وضعیت اجتماعی- اقتصادی، وزن فرد در هنگام تولد، نوع شیر مصرفی در دوران نوزادی، سابقه برخی بیماری‌ها در خانواده همانند عارضه چربی، چاقی و بیماری قند در این مطالعه مورد بررسی قرار گرفته‌اند.

در ادامه به توضیح روش‌های شناسایی مورد استفاده در این تحقیق می‌پردازیم. روش‌های مورد استفاده همگی از شناسایی‌کننده‌های پایه و پرکاربرد در مطالعات کاربردی می‌باشند. تمامی محاسبات و پردازش‌های آماری توسط نرم‌افزار SPSS نسخه ۲۲ انجام شده است. روش‌های شناسایی نیز به کمک نرم‌افزار برنامه‌نویسی MATLAB نسخه ۷ انجام شده است.

معرفی روش‌های هوشمند

اصطلاح طبقه‌بندی‌کننده را می‌توان به‌عنوان یک تابع در نظر گرفت که داده‌های ورودی را به چند گروه در خروجی تبدیل می‌کند (۲۱). در این مطالعه طبقه‌بندی‌کننده‌های مختلفی از قبیل تابع تفکیک‌کننده خطی (Linear Discriminant Function)، درجه دو (Quadratic)، بیز (Naïve Bayes)، ماشین بردار پشتیبان (Support Vector Machine)، شبکه عصبی احتمالاتی (Probabilistic Neural Network) و نرو- فازی (Neuro-Fuzzy) مورد بررسی قرار گرفته‌اند. این طبقه‌بندی‌کننده‌ها از نتایج کاربردی قابل قبولی در حوزه پزشکی برخوردار هستند (۲۴-۲۲). با توجه به این‌که شاخص توده بدنی وابسته به جنسیت است، نتایج برای هر دو گروه دختر و پسر به‌طور جداگانه ارایه گردیده است. نقاط برش برای تعیین طبقه‌های افراد چاق و افراد دارای اضافه وزن (برحسب شاخص توده بدنی) با استفاده از روش حداقل میانگین مربعات به‌دست آمده است (۲۵،۳). در این مطالعه زیرگروه افراد لاغر به دلیل کم بودن شیوع آن در بین جمعیت ایرانی و متمرکز بودن مقاله بر هدف شناسایی عارضه چاقی در بین افراد شرکت‌کننده از روند تحلیل داده‌ها حذف شده است. در ادامه به شرح خلاصه‌ای از انواع طبقه‌بندی‌کننده‌های مورد استفاده در این پژوهش می‌پردازیم.

تابع تفکیک‌کننده خطی و درجه دو اولین روش مورد استفاده در این مقاله است.

تابع تفکیک‌کننده، تابعی از الگوی داده‌های ورودی است که منجر به قوانینی برای جداسازی داده‌ها می‌شود. در تفکیک‌کننده خطی ناحیه بین ویژگی‌ها توسط خطوط جداساز از یکدیگر جدا می‌شوند تا بتوان کلاس‌های مختلف را جدا کرد. در تفکیک‌کننده درجه دو این نواحی توسط منحنی از یکدیگر جدا می‌شوند. در واقع این شیوه عضویت در یک کلاس را به‌صورت ترکیب خطی و یا غیرخطی از الگوهای ورودی مدل‌سازی می‌کند (۲۶).

طبقه‌بندی کننده بیز طبقه‌بندی‌کننده دیگراحتمالاتی است که بر اساس تئوری بیزی به وجود آمده است. این کلاسبند به

[Downloaded from hsr.mui.ac.ir on 2025-07-09]

دلیل سادگی در فهم و نیاز به تعداد داده یادگیری پایین برای تخمین پارامترهای کلاسبند از قبیل میانگین و واریانس، بسیار در مطالعات مورد استفاده قرار می‌گیرد (۲۶، ۲۱). روش سوم مورد استفاده ماشین بردار پشتیبان غیرخطی است که در این طبقه‌بندی‌کننده ویژگی‌های ورودی توسط یک تابع به فضای دیگر که غالباً با بعدی بالاتر از بعد ویژگی‌ها است، انتقال می‌یابد و سپس در فضای جدید بهترین صفحه چندبعدی طبقه‌بندی‌کننده ساخته می‌شود، به‌طوری‌که بیشترین فاصله حاشیه‌ای را نسبت به داده‌های دو کلاس داشته باشد. بیشینه بودن حاشیه در صفحه‌های چندبعدی طبقه‌بندی کننده مرزهای تصمیم‌گیری را به‌گونه‌ای تعیین می‌کند که طبقه‌بندی نسبت به داده‌های با مقدارهای غیرقابل پیش‌بینی، به‌مراتب پایدارتر باشد. به‌منظور استفاده از این طبقه‌بندی‌کننده در حالت غیرخطی دو پارامتر ضریب خطا و پارامتر هسته در تابع‌نمایی باید تنظیم شوند (۲۷). روش دیگر طبقه‌بندی کننده نرو- فازی است که در این روش قواعد زبانی با منطق فازی آمیخته می‌شوند. در حقیقت استفاده از قوانین زبانی نسبت تشخیص درست طبقه‌بندی کننده را افزایش می‌دهد (۲۸) و سپس در طراحی شبکه عصبی مورداستفاده قرار می‌گیرند. در این مطالعه از گرادیان مزدوج توزین شده (Scaled Conjugate Gradient) برای بهینه‌سازی پارامترهای غیرخطی طبقه‌بندی‌کننده استفاده شده است. این روش نسبت به روش‌های مشابه دارای سرعت بیشتری است و نیز برای حجم داده‌های وسیع بهتر عمل می‌کند (۲۹). آخرین روش مورد استفاده در این مطالعه شبکه عصبی احتمالاتی است که در آن ویژگی هر داده در ابتدا با گیرنده‌هایی به نام نورون‌های ورودی، برخورد می‌کنند و سپس هر ویژگی وزن خاصی پیدا می‌کند. مقادیر هر ویژگی پس از وزن‌دهی با یکدیگر جمع می‌شود و از یک تابع فعال‌ساز (Activation Function) عبور می‌کند؛ و نتیجه را می‌توان با توجه به فعال شدن و یا نشدن تابع فعال‌ساز مورد بررسی قرار داد. در شبکه عصبی احتمالاتی تابع فعال‌ساز توسط یک تابع نمایی ساخته می‌شود که توانایی طبقه‌بندی داده‌های غیرخطی را دارا است. از مزایای این طبقه‌بندی‌کننده می‌توان به آموزش راحت و بالا بردن پیچیدگی مرز تصمیم‌گیری در صورت لزوم اشاره کرد (۳۰).

نحوه انتخاب ویژگی‌ها

انتخاب ویژگی ازجمله مسایل مهم در مباحث شناسایی الگو است. علت بررسی این موضوع کاهش زمان پردازش، کاهش منابع ورودی، کاهش هزینه و افزایش کارایی سیستم‌های شناسایی است. داده‌های ورودی ممکن است تعداد بسیار زیادی ویژگی به همراه خود داشته باشند و بدون توجه به این‌که کدام ویژگی‌ها بیشتر تأثیر را در صحت طبقه‌بندی‌کننده خواهند گذاشت، طبقه‌بندی کننده طراحی‌شده ناکارامد خواهد بود. حذف ویژگی‌های نامربوط مشکل بعد بالا را از بین برده و عملکرد طبقه‌بندی‌کننده را بهبود می‌دهد (۲۹).

برای انجام فرایند انتخاب ویژگی روش‌های مختلفی وجود دارد که در این مطالعه از روش دنباله‌ای (Sequential) استفاده شده است. روش دنباله‌ای صحت طبقه‌بندی کننده را بر اساس دنباله‌ای از ویژگی‌ها بررسی می‌کند، به‌طوری‌که درنهایت دنباله‌ای که بیشترین کارایی را برای طبقه‌بندی‌کننده ایجاد می‌کند، انتخاب می‌گردد (۳۱).

نحوه ارزیابی طبقه‌بندی‌کننده‌ها

برای ارزیابی نتایج طبقه‌بندی‌کننده از روش ارزیابی متقابل دولایه‌ای (2-Fold Cross-Validation) استفاده کرده‌ایم. بدین‌صورت که برای هر طبقه‌بندی‌کننده داده‌ها به‌طور تصادفی به دو قسمت مساوی با رعایت نسبت به صورت مجموعه داده ۱ و مجموعه داده ۲ تقسیم می‌شوند، از مجموعه داده ۱ برای آموزش طبقه‌بندی کننده و دیگری برای آزمایش آن استفاده می‌شود و این کار به‌صورت بالعکس نیز انجام می‌شود؛ و نتایج کلی بر اساس میانگین نتایج در هر یک از حالات گفته‌شده، گزارش می‌گردد. بر اساس نتایج طبقه‌بندی مقادیر True Positive (تعداد افرادی که به‌درستی غیرعادی تشخیص داده‌شده‌اند)، True

[Downloaded from hsr.mui.ac.ir on 2025-07-09]

Negative (تعداد افرادی که به‌درستی عادی تشخیص داده شده‌اند)، False Positive (تعداد افرادی که به‌اشتباه غیرعادی تشخیص داده شده‌اند) و False Negative (تعداد افرادی که به‌اشتباه عادی تشخیص داده شده‌اند) تعیین و برای ارزیابی عملکرد هر طبقه‌بندی‌کننده از معیارهای حساسیت، ویژگی، صحت و سایر معیارهای مرتبط استفاده خواهد شد که در فرمول‌های (۱-۱) تا (۱-۸) ارایه گردیده است (۳۲).

$$Sensitivity\,(Se) = \frac{TP}{TP + FN} \quad (۱-۱)$$

$$Specificit\,y\,(Sp) = \frac{TN}{TN + FP} \quad (۱-۲)$$

$$precision\,(\Pr) = \frac{TP}{TP + FP} \quad (۱-۳)$$

$$Type\,I\,Err : FP\,Rate\ (\alpha) = 1 - Sp \quad (۱-۴)$$

$$Type\,II\,Err : FN\,Rate\ (\beta) = 1 - Se \quad (۱-۵)$$

$$Power = 1 - \beta = Se \quad (۱-۶)$$

$$F_1 - Score = 2 \times \frac{precision \times recall}{precision + recall} \quad (۱-۷)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (۱-۸)$$

از آن‌جا که طبقه‌بندی‌کننده هنگام وجود داده‌های ناقص عملکرد مناسبی از خود نشان نمی‌دهند. از این‌رو برای مقایسه‌ای درست میان طبقه‌بندی‌کننده‌ها، ابتدا تمامی داده‌هایی را که حداقل یک ویژگی آن‌ها نقص داشت، از پایگاه اطلاعاتی خارج نموده و تحلیل‌ها با افرادی که دارای اطلاعات کامل در این زمینه هستند انجام پذیرفته است.

یافته‌ها

از مجموع ۱۴۸۸۰ نفر دعوت شده، ۱۳۴۸۶ نفر در مطالعه شرکت کردند. کل یافته‌ها بر اساس اطلاعات ۹۷۹۵ (۴۹/۱۷٪ پسر) از کودکان و نوجوانان ایرانی است که ویژگی‌های آن‌ها به‌صورت کامل ذخیره شده است.

جدول یک توزیع برخی از عوامل تأثیرگذار را در دو گروه افراد نرمال و افراد دارای عارضه چاقی نشان می‌دهد. طبقه‌بندی‌ها با ۲۳ متغیر اولیه از افراد اولیه انجام شده است. ارزیابی عملکرد طبقه‌بندی کننده‌ها برای دو گروه دختر و پسر به‌طور جداگانه بررسی شده است. برای گروه پسر میانگین صحت در ارزیابی متقابل دولایه‌ای برای روش‌های مختلف عبارت هستند از: LDA (۷۸٪)، QDA (۷۴٪)، NB (۷۴٪)، SVM(۸۰٪)، NF(۷۷٪) و PNN(۷۶٪). به همین صورت صحت برای گروه دختر به‌صورت مقابل است: LDA (۷۹٪)، QDA (۷۵٪)، NB(۷۸٪)، SVM (۸۱٪)، NF (۸۱٪) و PNN (۷۵٪). جزییات مربوط به ارزیابی ویژگی‌های مختلف برای دو گروه دختر و پسر به‌صورت جداگانه در جدول ۲ و ۳ ذکر شده است. بر اساس بهترین روش برحسب معیارهای جداول ۲ و ۳ این‌بار با انتخاب ویژگی‌های مناسب بار دیگر کارایی طبقه‌بندی‌کننده برتر در جدول ۴ مورد ارزیابی قرار گرفته است.

ویژگی‌های انتخابی در کل جمعیت مورد مطالعه عبارت هستند از: سن، فعالیت ورزشی، نوع شیرمصرفی در هنگام نوزادی و سابقه دیابت در خانواده. جدول ۵ نتایج طبقه‌بندی‌کننده ماشین بردار پشتیبان را به‌صورت ترکیبی نشان می‌دهد.

[Downloaded from hsr.mui.ac.ir on 2025-07-09]

جدول ۱. اطلاعات آماری متغیرهای موجود در پایگاه داده‌ها

ویژگی		افراد نرمال	افراد چاق
جنسیت (٪)*	پسر	۸۸/۴۷	۲۵/۵۳
سن (پسر و دختر)*	میانگین	۵/۱۴	۵۶/۱۳
	۹۵٪ بازه اطمینان	-۰۳/۱۲	-۱۷/۱۱
		۹۶/۱۶	۹۴/۱۵
وضعیت مصرف صبحانه (٪)*	کامل	۲۰/۶۹	۰۷/۶۳
	گاهی اوقات	۱۲/۱۳	۰۰/۱۴
	اصلاً	۶۸/۱۷	۹۳/۲۲
وضعیت مصرف شیرینی (٪)*	روزانه	۲۰/۳۵	۴۹/۲۹
	هفتگی	۲۶/۴۲	۰۰/۴۲
	به‌ندرت	۵۴/۲۲	۵۱/۲۸
وضعیت مصرف اسنک (٪)*	روزانه	۷۵/۱۲	۱۶/۱۱
	هفتگی	۴۰/۳۶	۴۹/۳۵
	به‌ندرت	۸۵/۵۰	۳۵/۵۳
وضعیت مصرف سبزی (٪)	روزانه	۰۳/۳۶	۹۰/۳۷
	هفتگی	۳۵/۴۶	۸۵/۴۴
	به‌ندرت	۶۲/۱۷	۲۵/۱۷
وضعیت مصرف فست فود (٪)	روزانه	۷۰/۲	۱۷/۳
	هفتگی	۷۲/۲۳	۹۸/۲۳
	به‌ندرت	۵۸/۷۳	۸۵/۷۲
فعالیت ورزشی در طول هفته (٪)	هیچ	۵۰/۸	۹۲/۸
	یک	۶۱/۲۰	۱۳/۲۰
	دو	۳۰/۱۷	۵۵/۱۷
	سه	۲۰/۱۴	۳۰/۱۵
	چهار	۸۰/۸	۹۸/۹
	پنج	۹۲/۴	۸۲/۴
	شش	۵۰/۷	۹۶/۵
	هفت	۱۸/۱۸	۳۴/۱۷
مدت استفاده از تلویزیون و رایانه (٪)*	کمتر از ۴ ساعت	۸۲/۶۴	۷۸/۱۳
	بیشتر از ۴ ساعت	۱۷/۳۶	۲۱/۸۷
دخانیات پذیرا (٪)	بله	۴۵/۴۴	۰۱/۴۵
	خیر	۵۵/۵۵	۹۹/۵۴
استعمال دخانیات (٪)	خیر	۵۵/۹۷	۴۲/۹۷
	سیگار	۲۸/۰	۲۵/۰
	قلیان	۱۲/۲	۳۳/۲
	سایر	۰۵/۰	۰۰/۰
مدت زمان خواب (٪)*	کمتر از ۵ ساعت	۵۷/۰	۹۳/۰
	بین ۵ تا ۸ ساعت	۶۳/۲۲	۷۰/۲۴
	بیشتر از ۸ ساعت	۸۰/۷۶	۳۷/۷۴

[Downloaded from hsr.mui.ac.ir on 2025-07-09]

بی‌خوابی (٪)	بله	۵۰/۱۵	۲۳/۱۵
	خیر	۵۰/۸۴	۷۷/۸۴
افسردگی (٪)*	بله	۱۶/۲۰	۲۵/۲۲
	خیر	۸۴/۷۹	۷۵/۷۷
اضطراب (٪)	بله	۶۴/۲۴	۱۵/۲۶
	خیر	۳۶/۷۵	۸۵/۷۳
وضعیت اجتماعی-اقتصادی (٪)*	ضعیف	۱۵/۳۵	۵۸/۲۲
	متوسط	۲۶/۳۳	۶۲/۳۲
	خوب	۵۹/۳۱	۸۰/۴۴
وزن در هنگام تولد (٪)*	کمتر از ۲۵۰۰**	۰۴/۸	۴۷/۶
	بین ۲۵۰۰ تا ۴۰۰۰	۶۵/۷۶	۹۱/۷۷
	بیشتر از ۴۰۰۰	۵۵/۶	۱۷/۹
	وزن نامشخص	۷۶/۸	۴۵/۶
سابقه عارضه چربی در خانواده (٪)*	بله	۴۸/۴۳	۲۵/۵۰
	خیر	۵۲/۵۶	۷۵/۴۹
نوع شیر مصرفی در دوران نوزادی (٪)	شیر مادر	۶/۸۲	۲۰/۸۳
	شیر خشک	۱۵/۴	۴۰/۳
	شیر مادر و خشک	۸۷/۱	۸۲/۱
	شیر گاو	۵۰/۱۰	۱۱/۱۰
	شیر مادر و گاو	۳۷/۰	۳۳/۰
	شیر خشک و گاو	۵۱/۰	۵۴/۰
سابقه چاقی در خانواده (٪)*	بله	۷۶/۴۳	۹۳/۵۶
	خیر	۲۴/۵۶	۰۷/۴۳
سابقه دیابت در خانواده (٪)*	بله	۶۵/۳۵	۸۸/۴۴
	خیر	۳۵/۶۴	۱۲/۵۵
نوع روغن مصرفی (٪)*	جامد	۴۵/۳۸	۷۴/۲۹
	سرخ کردنی	۹۸/۱۸	۶۵/۲۶
	مایع	۸۰/۳۷	۱۵/۳۸
	حیوانی	۱۰/۳	۶۳/۳
	کره	۴۵/۰	۶۵/۰
	پی و دنبه	۲۲/۱	۱۸/۱

تفاوت معنی‌دار ($\alpha=0.05$)

** وزن برحسب گرم می‌باشد.

[Downloaded from hsr.mui.ac.ir on 2025-07-09]

جدول۲. نتایج ارزیابی متقابل دولایه‌ای بر اساس روش‌های مختلف شناسایی در گروه پسران

معیار*	LDA	QDA	NB	SVM	NF	PNN
Se (%)	۷۳/۹	۶۸/۹	۳۷/۶	۷۳/۷	۲۹/۳	۳۶/۲
Sp (%)	۷۸/۸	۷۶/۳	۸۶/۴	۸۱/۹	۹۴/۳	۸۹/۷
Pr (%)	۵۵/۳	۵۰/۸	۴۹/۶	۵۹/۱	۶۴/۲	۵۵/۴
α=Type I err	۰/۲۱	۰/۲۳	۰/۱۳	۰/۱۸	۰/۰۵	۰/۱
Power (%)	۷۳/۹	۶۸/۹	۳۷/۶	۷۳/۷	۲۹/۳	۳۶/۲
Acc (%)	۷۷/۶	۷۴/۴	۷۳/۷	۷۹/۸	۷۷/۳	۷۵/۷
F_1-Score (%)	۶۳/۳	۵۸/۴	۴۲/۸	۶۵/۵	۴۰/۱	۴۳/۸

*تفکیک‌کننده خطی: LDA ، تفکیک‌کننده درجه دو: QDA ، طبقه‌بند بیز: NB، ماشین بردار پشتیبان: SVM ، طبقه‌بند نرو-فازی: NF ، طبقه‌بند شبکه عصبی احتمالاتی:PNN

جدول۳. نتایج ارزیابی متقابل دولایه‌ای بر اساس روش‌های مختلف شناسایی در گروه دختران

معیار*	LDA	QDA	NB	SVM	NF	PNN
Se (%)	۷۵/۲	۶۹/۳	۳۷/۵	۷۶/۶	۳۲/۸	۳۴/۶
Sp (%)	۸۰	۷۶/۷	۸۹/۳	۸۲/۷	۹۴/۸	۸۶/۷
Pr (%)	۵۱/۸	۴۵/۹	۵۰/۲	۵۵/۸	۶۵/۵	۴۲/۶
α=Type I err	۰/۱۹	۰/۲۳	۰/۱	۰/۱۷	۰/۰۵	۰/۱۳
Power (%)	۷۵/۲	۶۹/۳	۳۷/۵	۷۶/۶	۳۲/۸	۳۴/۶
Acc (%)	۷۹	۷۵	۷۷/۸	۸۱/۳	۸۱/۱	۷۵/۱
F_1-Score (%)	۶۱/۴	۵۵/۲	۴۲/۹	۶۴/۶	۴۳/۴	۳۸/۲

*تفکیک‌کننده خطی: LDA ،تفکیک‌کننده درجه دو: QDA ، طبقه‌بند بیز: NB، ماشینبردار پشتیبان: SVM ،طبقه‌بند نرو-فازی:NF ، طبقه‌بند شبکه عصبی احتمالاتی: PNN

[Downloaded from hsr.mui.ac.ir on 2025-07-09]

جدول۴. نتایج ارزیابی متقابل دولایه‌ای برای طبقه‌بندی‌کننده ماشین بردار پشتیبان به‌عنوان بهترین روش بر اساس انتخاب ویژگی (دختر و پسر با انتخاب ویژگی)

گروه	Se (%)	Sp (%)	Pr (%)	A	Acc (%)	F_1-score (%)
پسر	۷۶/۳	۸۵/۹	۶۵/۸	۰/۱۴	۸۳/۴	۷۰/۶
دختر	۷۷/۴	۸۲/۷	۵۶/۱	۰/۱۷	۸۱/۵	۶۵

جدول۵. نتایج ترکیبی کلاسبند ماشین بردار پشتیبان غیرخطی در کل جمعیت مورد مطالعه*

		واقعیت (تاییدشده توسط سازمان جهانی سلامت)	
		بیمار	سالم
پیش‌بینی	بیمار	۹۲۱ (TP)	۵۵۳ (FP)
	سالم	۲۶۰ (FN)	۳۱۶۲ (TN)

* آموزش کلاسبند توسط مجموعه داده ۱ با ۴۸۹۹ مورد و آزمایش توسط مجموعه داده ۲ با ۴۸۹۶ مورد

بحث

در ارزیابی‌های به‌عمل‌آمده ویژگی‌های سن، میزان مصرف شیرینی، میزان فعالیت بدنی، زمان نگاه به نمایشگرهای مختلف، مدت‌زمان خواب، سابقه دیابت در خانواده و نوع شیر مصرفی در دوران نوزادی بیشترین تأثیرگذاری را در ایجاد عارضه چاقی داشته‌اند. مروری بر بررسی‌های پیشین نتایج به‌دست‌آمده را تایید می‌کند. مطالعات پیشین نشان داده‌اند که مصرف شیر مادر می‌تواند از بروز چاقی جلوگیری کند (۳۴-۳۳). برخی مطالعات تأثیرگذاری زمان خواب را تنها محدود به گروه پسر می‌دانند (۳۵) و برخی کاهش زمان خواب را موجب افزایش وزن بدن در کودکان گزارش کرده‌اند (۳۶). مطالعات متعددی افزایش فعالیت ورزشی را عاملی در جهت جلوگیری از عارضه چاقی توصیف کرده‌اند (۳۸-۳۷، ۱۱). افزایش زمان نگاه به نمایشگر و افزایش مصرف شیرینی نیز می‌توانند باعث بروز چاقی شوند (۳۷، ۱۲).

در این مطالعه با استفاده از ۲۳ متغیر اولیه که برای افراد شرکت‌کننده در این مطالعه ثبت شده بود، فرایند طبقه‌بندی باهدف شناسایی ناهنجاری وزنی انجام شده است. بدین منظور ۶ روش طبقه‌بندی شامل LDA، QDA، NB، ماشین بردار پشتیبان، نرو- فازی و PNN به‌عنوان طبقه‌بندی‌کننده‌های اولیه انتخاب و مورد ارزیابی قرارگرفته‌اند. سه طبقه‌بندی کننده LDA، QDA و ماشین بردار پشتیبان در مقایسه با سایرین بهترین نتیجه را در ارزیابی‌ها داشته‌اند. همان‌طور که قبلاً نیز اشاره شد، علم پزشکی و به‌طور کل تصمیم‌گیری‌های حوزه پزشکی همواره با یک عدم قطعیتی همراه بوده است. علت این امر نیز در پیچیدگی فرایندهای بیولوژیکی و ناکافی بودن اطلاعات موجود است. بدین منظور از طبقه‌بندی‌کننده نرو- فازی به‌عنوان طبقه‌بندی‌کننده‌ای که از نظریه سامانه‌های فازی بهره می‌برد، در این مطالعه استفاده شده است. در حقیقت نظریه فازی به‌منظور استفاده در علومی که با ابهام و عدم قطعیت همراه است به وجود آمده است. با استفاده از قوانین مناسب، سیستم فازی یک بستر مناسب برای تجزیه تحلیل اطلاعات بر پایه مدارک، استدلال قیاسی و کلامی به وجود می‌آورد (۴۰-۳۹).

[Downloaded from hsr.mui.ac.ir on 2025-07-09]

برای شناسایی عارضه چاقی ابتدا از تمامی متغیرهای ذکر شده در مطالعه استفاده شده است. سپس برای کاهش پارامترهای ورودی و افزایش صحت، برخی از متغیرهای نامربوط توسط روش دنباله‌ای حذف گردید. پس از به‌کارگیری ویژگی‌های اصلاح‌شده، به‌طور میانگین در کلاس‌بندهای LDA، QDA و ماشین بردار پشتیبان شاهد افزایش ۱، ۵ و ۲ درصدی صحت طبقه‌بندی کننده بوده‌ایم. قابل‌ذکر است که درروش دنباله‌ای ارتباط و همبستگی بین ویژگی‌های مختلف در نظر گرفته نمی‌شود (۴۱). استفاده از روش‌های آماری مناسب از قبیل رگرسیون چندگانه منطقی (Multiple Logistic) می‌تواند فرایند انتخاب ویژگی را بهبود دهد. هم‌چنین ویژگی‌های مختلف ممکن است تأثیر مختلفی بر نتیجه داشته باشند و این امر توسط برنامه‌های بهینه‌سازی از قبیل بهینه‌سازی کلونی و با تغییر ضرایب تأثیر ویژگی‌های مختلف محقق می‌شود (۴۲).

برای تعیین بهترین طبقه‌بندی کننده، زمانی که از همه متغیرها استفاده شد، میانگین معیار F-Score برای دو حالت آزمایش طبقه‌بندی‌کننده‌ها در نظر گرفته شد و در این شرایط طبقه‌بندی کننده ماشین بردار پشتیبان به‌عنوان بهترین کلاسبند انتخاب شد، هم چنین پس از انتخاب ویژگی در بین سه طبقه‌بندی‌کننده برتر، دوباره بیشترین کارایی را از خود نشان داده است. با این شرایط طبقه‌بندی‌کننده ماشین بردار پشتیبان غیرخطی به‌عنوان طبقه‌بندی‌کننده کارامد برای تشخیص چاقی معرفی گردید. هرچند طبقه‌بندی‌کننده‌های LDA و QDA نیز عملکرد قابل قبولی داشته‌اند. اگرچه با توجه به مرز تصمیم‌گیری منحنی شکل در QDA عملکرد آن نسبت به LDA بایستی بهتر باشد، اما طبقه‌بندی کننده QDAبا توجه به نتایج آزمایش قدری ضعیف‌تر نسبت به LDA عمل کرده است، علت این امر تغییرات کمتر در مرزهای تصمیم‌گیری LDA است که این کلاسبند را نسبت به تغییرات پایدارتر می‌کند. به بیان دیگر در مرحله آموزش طبقه‌بندی کننده QDA به‌مراتب بهتر از LDA عمل می‌کند، اما در مرحله آزمایش ممکن است برخی از شروط، ازجمله نرمال بودن توزیع داده‌ها در هر کلاس برقرار نباشد و این موضوع عملکرد QDA را مختل می‌کند. در این حالت کلاسبند LDA نسبت به QDA از قدرت پایداری بیشتری برخوردار است. مرز تصمیم‌گیری ساده موجب قدرت پایداری بیشتر در LDA می‌شود و می‌تواند تغییرات بیشتری را نسبت به توزیع داده‌ها در هر کلاس از خود نشان دهد (۴۳)؛ اما QDA به علت مرز تصمیم‌گیری پیچیده‌تر و به عبارتی دقیق‌تر انعطاف کمتری نسبت به تغییرات توزیع داده‌ها در هر کلاس دارد. اگرچه داشتن مرزهای تصمیم‌گیری دقیق و پیچیده این امکان را به QDA می‌دهد تا کلاس‌ها و الگوهای پیچیده‌تری را شناسایی کند (۲۶).

توافق بین ارزیابان استفاده شده بدون انتخاب ویژگی در جداول ۲ و ۳، با استفاده از معیار Cohen's Kappa مورد بررسی قرار گرفت (۴۴). بر این اساس، برای پسران حداکثر ضریب توافق بین LDA و SVM (۰/۸۷) نشان‌دهنده توافق کامل (Perfect Agreement (CC>0.80 & CC<=1.00)) و حداقل آن بین NF و PNN (۰/۱۴) نشان‌دهنده توافق اندک (Slight agreement (CC>0.00 & CC<=0.20)) بوده است. برای دختران، حداکثر ضریب توافق بین LDA و SVM (۰/۸۷) نشان‌دهنده توافق کامل و حداقل آن بین NF و PNN (۰/۲۴) نشان‌دهنده توافق رضایت‌بخش (Fair agreement (CC>=0.21 & CC<=0.40)) می‌باشد. بیشترین توافق بین ارزیابان و استاندارد طلایی، برای پسران مربوط به LDA و SVM و برای دختران، مربوط به LDA، QDA و SVM به‌صورت توافق متوسط (Moderate Agreement (CC>=0.41 & CC<=0.6))می‌باشد. بقیه ارزیابان برای پسران دارای توافق اندک و برای دختران دارای توافق رضایت‌بخش با استاندارد طلایی هستند.

در تفکیک‌کننده خطی استاندارد، ویژگی‌های ورودی باید در هر گروه دارای توزیع نرمال بوده و ماتریس کوواریانس یکسان داشته باشند (۲۶). این شرایط به‌هیچ‌وجه برای داده‌های مورد بررسی، صدق نمی‌کند (آزمون Kolmogorov-Smirnov

[Downloaded from hsr.mui.ac.ir on 2025-07-09]

و آنالیز واریانس Levene). بااین‌حال به دلیل استفاده از تابع متمایزکننده خطی Fisher در روش LDA، بررسی دو شرط بالا موردنیاز نمی‌باشد. در روش QDA فرض نرمال بودن داده‌ها مورد نیاز است و این می‌تواند عملکرد ضعیف‌تر QDA نسبت به LDA را توجیه کند. اگرچه به دلیل ایجاد نواحی تصمیم‌گیری بیضوی، هیپربولیک، و پارابولیک توسط QDA در شرایط غیر نرمال نیز، کارایی نسبتاً مناسبی از خود نشان می‌دهد (۴۵). لازم به ذکر است که فرض نرمال بودن از نسخه بیزین QDA حذف می‌گردد و این روش جایگزین مناسبی برای QDA متداول محسوب می‌شود (۴۶). روش‌های NB، NF و PNN زمانی بالاترین کارایی خود را نشان می‌دهند که ویژگی‌های هر گروه از همدیگر مستقل باشند (۳۰، ۲۶). شرط دیگر برای NF، کم بودن تعداد متغیرهای ورودی است (۴۷). اگر چه روش‌های مزبور در داده‌های همبسته و نویزی رفتار نسبتاً مقاومی از خود نشان می‌دهند (۴۸). در گروه نرمال، هیچ‌گونه همبستگی قابل ملاحظه‌ای بین ویژگی‌های ورودی مشاهده نشد (قدر مطلق ضریب همبستگی از ۰/۳ کمتر بود). در افراد گروه چاق، تنها در یک مورد و آن‌هم بین مصرف اسنک و فست فود، همبستگی معنی‌دار ۰/۳۱ دیده شد. در روش SVM که نسخه پیشرفته‌تری از LDA است، قابل تمایز بودن کلاس‌ها مستلزم برقراری شروط Karush-Kuhn-Tucker می‌باشد که به دلیل تمایز خطی مناسب در روش LDA، این شروط برقرار است (۴۹). عملکرد مناسب SVM نیز به‌صورت غیرمستقیم، نشان‌دهنده برقراری شروط مزبور می‌باشد.

طبقه‌بندی داده‌های ترتیبی در مقایسه با داده‌های بازه‌ای و کیفی از سختی بالاتری برخوردار است. در این مطالعه داده‌ها به‌صورت ترکیبی از داده‌های ترتیبی، بازه‌ای و کیفی وجود دارند و این موضوع فرایند طبقه‌بندی و تخمین فاصله را سخت می‌کند. به‌عنوان مثال سن یک متغیر بازه‌ای است، درحالی‌که فعالیت ورزشی یک متغیر ترتیبی است؛ و از طرفی دیگر سابقه دیابت در خانواده یک متغیر کیفی است؛ بنابراین تنظیم و تخمین فاصله بین متغیرها در این مطالعه کار آسانی نبوده است (۳۲).

نتیجه‌گیری

در این مطالعه عارضه چاقی به‌طور قابل قبولی بر اساس پارامترهای ورودی شناسایی شد و سامانه‌ای برای تشخیص بر اساس پارامتر ورودی طراحی گردید. برخی از ویژگی‌های انتخابی شامل سن، میزان مصرف شیرینی، میزان فعالیت بدنی، زمان نگاه به نمایشگرهای مختلف، مدت زمان خواب، سابقه دیابت در خانواده و نوع شیر مصرفی در دوران نوزادی بیشترین تأثیر را در ایجاد عارضه چاقی داشته‌اند. این سیستم این امکان را می‌دهد تا در حوزه سلامت بررسی‌های لازم به عمل آید و با تجزیه و تحلیل پارامترهای تأثیرگذار بر چاقی برنامه مناسبی برای پیشگیری از بروز چاقی در میان کودکان و نوجوانان پی‌ریزی شود.

تشکر و قدردانی

بدین‌وسیله از گروه بزرگ مطالعات کاسپین که از دانشگاه‌های مختلف در سراسر ایران مسؤولیت جمع‌آوری داده را بر عهده داشته‌اند کمال قدردانی و تشکر را به عمل می‌آوریم. هم‌چنین از تمامی سازمان‌ها، افراد و خانواده‌های گرامی که در انجام این مطالعه شرکت داشته‌اند قدردانی می‌کنیم.

References

1. Must A, Strauss RS. Risks and consequences of childhood and adolescent obesity. International journal of obesity and related metabolic disorders . Journal of the International Association for the Study of Obesity 1999;23(Suppl 2):S2-11.
2. Ul-Haq Z, Mackay DF, Fenwick E, Pell JP. Meta-analysis of the association between body mass index and health-related quality of life among children and adolescents, assessed using the pediatric quality of life inventory index. The Journal of pediatrics 2013;162(2):280-6.

[Downloaded from hsr.mui.ac.ir on 2025-07-09]

3. De Onis M, Organization WH. WHO child growth standards: Length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age: Methods and development. *Geneva:* World Health Organization, 2006.
4. Kelishadi R, Marashinia F, Heshmat R, Motlagh ME, Qorbani M, Taslimi M, et al. First report on body image and weight control in a nationally representative sample of a pediatric population in the Middle East and North Africa: the CASPIAN-III study. Archives of medical science : AMS. 2013 Apr 20;9(2):210-7.
5. Popkin BM, Gordon-Larsen P. The nutrition transition: worldwide obesity dynamics and their determinants. International journal of obesity 2004;28:S2-S9.
6. DeLaune S, Ladner P. Fundamentals of nursing. 4th ed. Clifton Park, NY: Cengage Learning; 2010. p 1456.
7. Tol A, Tavassoli E, Shariferad GR, Shojaeezadeh D. Health-promoting lifestyle and quality of life among undergraduate students at school of health, Isfahan university of medical sciences. Journal of education and health promotion 2013;2:11.
8. Lee HA, Lee WK, Kong KA, Chang N, Ha EH, Hong YS, et al. The effect of eating behavior on being overweight or obese during preadolescence. Journal of preventive medicine and public health = Yebang Uihakhoe chi 2011;44(5):226-33.
9. Steinbeck KS. The importance of physical activity in the prevention of overweight and obesity in childhood: A review and an opinion. Obesity Reviews. 2001;2(2):117-30.
10. Kelishadi R, Ardalan G, Gheiratmand R, Gouya MM, Razaghi EM, Delavari A, et al. Association of physical activity and dietary behaviours in relation to the body mass index in a national sample of Iranian children and adolescents: CASPIAN Study. Bulletin of the World Health Organization 2007;85(1):19-26.
11. Lahti-Koski M, Pietinen P, Heliövaara M, Vartiainen E. Associations of body mass index and obesity with physical activity, food choices, alcohol intake, and smoking in the 1982–1997 FINRISK Studies. The American journal of clinical nutrition 2002;75(5):809-17.
12. Malik VS, Schulze MB, Hu FB. Intake of sugar-sweetened beverages and weight gain: a systematic review. The American journal of clinical nutrition 2006;84(2):274-88.
13. McLaren L. Socioeconomic status and obesity. Epidemiologic reviews 2007;29(1):29-48.
14. Sobal J, Stunkard A J. Socioeconomic status and obesity: A review of the literature. Psychological bulletin 1989;105(2):260-75.
15. Castellino RA. Computer aided detection (CAD): An overview. Cancer Imaging 2005;5(1):9-17.
16. Rangayyan RM. Biomedical signal analysis: IEEE Standards Office. New York, NY :Wiley-*IEEE* Press; 2001.
17. Rogers W, Ryack B, Moeller G. Computer-aided medical diagnosis: Literature review. International journal of bio-medical computing 1979;10(4):267-89.
18. Jari M, Qorbani M, Motlagh ME, Heshmat R, Ardalan G, Kelishadi R. A Nationwide Survey on the Daily Screen Time of Iranian Children and Adolescents: The CASPIAN - IV Study. International journal of preventive medicine 2014;5(2):224-9.
19. Kelishadi R, Ardalan G, Qorbani M, Ataie-Jafari A, Bahreynian M, Taslimi M, et al. Methodology and Early Findings of the Fourth Survey of Childhood and Adolescence Surveillance and Prevention of Adult Non-Communicable Disease in Iran: The CASPIAN-IV Study. International journal of preventive medicine. 2013;4(12):1451-60.
20. Kelishadi R, Majdzadeh R, Motlagh M-E, Heshmat R, Aminaee T, Ardalan G, et al. Development and evaluation of a questionnaire for assessment of determinants of weight disorders among children and adolescents: The Caspian-IV study. International journal of preventive medicine 2012;3(10):699.
21. Kohavi R. Scaling Up the Accuracy of Naive-Bayes Classifiers: A Decision-Tree Hybrid. KDD; 1996.

[Downloaded from hsr.mui.ac.ir on 2025-07-09]

22. Maglogiannis I, Zafiropoulos E, Anagnostopoulos I. An intelligent system for automated breast cancer diagnosis and prognosis using SVM based classifiers. Applied intelligence 2009;30(1):24-36.
23. Polat K, Güneş S. An expert system approach based on principal component analysis and adaptive neuro-fuzzy inference system to diagnosis of diabetes disease. Digital Signal Processing;17(4):702-10.
24. Sboner A, Eccher C, Blanzieri E, Bauer P, Cristofolini M, Zumiani G, et al. A multiple classifier system for early melanoma diagnosis. Artificial Intelligence in Medicine 2003;27(1):29-44.
25. Mansourian M, Marateb HR, Kelishadi R, Motlagh ME, Aminaee T, Taslimi M, et al. First growth curves based on the World Health Organization reference in a Nationally-Representative Sample of Pediatric Population in the Middle East and North Africa (MENA): The CASPIAN-III study. BMC Pediatr 2012;12:149.
26. Webb AR. Statistical pattern recognition. New York: John Wiley & Sons; 2003.
27. Hsu C-W, Chang C-C, Lin C-J. A practical guide to support vector classification [Technical Report]. Taipai: National Taiwan University; 2003.
28. Cetisli B. Development of an adaptive neuro-fuzzy classifier using linguistic hedges: Part 1. Expert systems with Applications. 2010;37(8):6093-101.
29. Cetişli B, Barkana A. Speeding up the scaled conjugate gradient algorithm and its application in neuro-fuzzy classifier training. Soft computing 2010;14(4):365-78.
30. Specht DF. Probabilistic neural networks. Neural networks 1990;3(1):109-18.
31. Aha DW, Bankert RL. A comparative evaluation of sequential feature selection algorithms. Learning from Data. New York; NY: Springer: 1996. pp. 199-206.
32. Marateb HR, Mansourian M, Adibi P, Farina D. Manipulating measurement scales in medical statistical analysis and data mining: A review of methodologies. Journal of research in medical sciences: the official journal of Isfahan University of Medical Sciences. 2014;19(1): *47*-56.
33. Arenz S, Rückerl R, Koletzko B, von Kries R. Breast-feeding and childhood obesity—a systematic review. International journal of obesity 2004;28(10):1247-56.
34. Owen CG, Martin RM, Whincup PH, Davey-Smith G, Gillman MW, Cook DG. The effect of breastfeeding on mean body mass index throughout life: A quantitative review of published and unpublished observational evidence. The American journal of clinical nutrition 2005;82(6):1298-307.
35. Storfer-Isser A, Patel SR, Babineau DC, Redline S. Relation between sleep duration and BMI varies by age and sex in youth age 8-19. Pediatric obesity 2012;7(1):53-64.
36. Cappuccio FP, Taggart FM, Kandala NB, Currie A, Peile E, Stranges S, et al. Meta-analysis of short sleep duration and obesity in children and adults. Sleep 2008 May;31(5):619-26.
37. Andersen RE, Crespo CJ, Bartlett SJ, Cheskin LJ, Pratt M. Relationship of physical activity and television watching with body weight and level of fatness among children: Results from the third national health and nutrition examination survey. JAMA 1998;279(12):938-42.
38. Fogelholm M, Kukkonen-Harjula K. Does physical activity prevent weight gain – a systematic review. Obesity Reviews 2000;1(2):95-111.
39. Steimann F, Adlassnig KP. Fuzzy medical diagnosis. Bristol Philadelphia: Handbook of Fuzzy Computation; 1998:G13.
40. Teodorescu H-NL, Kandel A, Jain LC. Fuzzy and neuro-fuzzy systems in medicine . Florida, US: CRC Press; 1998.
41. Liu H, Motoda H. Feature selection for knowledge discovery and data mining: New York: Springer; 1998.
42. Sousa T, Silva A, Neves A. Particle swarm based data mining algorithms for classification tasks. Parallel Computing 2004;30(5):767-83.
43. Marateb HR, Mansourian M, Faghihimani E, Amini M, Farina D. A hybrid intelligent system for diagnosing microalbuminuria in type 2 diabetes patients without having to measure urinary albumin. Computers in biology and medicine 2014;45:34-42.

[Downloaded from hsr.mui.ac.ir on 2025-07-09]

44. Cyr L, Francis K. Measures of clinical greement for nominal and categorical data: the kappa coefficient. Computers in biology and medicine 1992;22(4):239-46.
45. R.O. Duda, P.E. Hart, D.G. Stork. Pattern Classification. 2nd ed. NewYork: Wiley; 2000.
46. Srivastava S, Gupta MR, Frigyik A. Bayesian Quadratic Discriminant Analysis. J Mach Learn Res. 2007;8:1277-305.
47. Hosseini MS, Zekri M. Review of Medical Image Classification using the Adaptive Neuro-Fuzzy Inference System. Journal of medical signals and sensors 2012;2(1):49-60.
48. S.B. Kotsiantis, Supervised Machine Learning: A Review of Classification Techniques. Netherlands: IOS Press; 2007. pp. 3–24.
49. V.N. Vapnik. Statistical Learning Theory. New York: Wiley; 1998

[Downloaded from hsr.mui.ac.ir on 2025-07-09]

A computer-aided Diagnosis System for the Prediction of Overweighs Using Life Style Factors, Socio-economic Status and Family History of Disorders in Children

Morteza Farahi [1], Hamid Reza Marateb [2],
Roya Kelishadi [3], Mohammad Esmaeil Motlagh [4]

Original Article

Abstract

Background: Obesity represents one of the most important nutritional problems worldwide.Obesity in childhood can cause variety of health issues such as orthopedic, neurological, pulmonary and gastroenterological disorders in the future, although no side effects were reported from malignant obesity during childhood. In this paper, we presented a computer-aided diagnosis system to predict the obesity based on input features obtained from the life style and other factors of the subjects.

Methods: The total number of 9795 subjects (49.17% boy) aged 6 to 18 years taken from the CASPIAN IV study participated in this study. The input parameters of the proposed system were taken from the dietary habit, physical activity, family history, social economic status, and other features. Then, the obesity was predicted using the data mining and artificial intelligence techniques. Feature Selection (FS) methods were also used to improve the performance of the proposed system. The performance of the diagnosis system was assessed based on the hold-out validation framework.

Findings: The performance of the classifications method has been validated by hold-out cross validation. Among the different classification techniques tested, SVM with FS showed the best performance. The accuracy and precision of this method were 63.3% and 83.7%, respectively. Some features such as age, physical activities, birth feeding and family history of diabetes mellitus detected as the most effectivefactors with obesity in both gender.

Conclusion: Designing of an intelligentdiagnosis system with the input parameter such as life-style, socioeconomic status and genetic information can help predict obesity in children to modify their life-style to improve theirqualityof life in the future. A web-based version of this intelligent system can easily provide the obesity prediction facilities for the families at home.

Key Words: Classification; Obesity; Data Mining; Medical Diagnosis System; Artificial Intelligence

Citation: Farahi M, Marateb HR, Kelishadi R, Motlagh M E. **A computer-aided Diagnosis System for the Prediction of Overweighs Using Life Style Factors, Socio-economic Status and Family History of Disorders in Children.** J Health Syst Res 2014; 10(3):571-586

Received date: 14.09.2014 **Accept date: 11.10.2014**

1. Student, Biomedical Engineering Department, Faculty of Engineering, University of Isfahan, Iran
2. Assistant Professor, Biomedical Engineering Department, Faculty of Engineering, University of Isfahan, Iran
3. Professor of Pediatrics, Pediatrics Department, Faculty of Medicine and Child Growth and Development Research Center, Isfahan University of Medical Sciences, Isfahan, Iran (Corresponding Author) Email: roya_kelishadi@gmail.com
4. Department of School Health, Bureau of Population, Family and School Health, Ministry of Health and Medical Education, Tehran, Iran

[Downloaded from hsr.mui.ac.ir on 2025-07-09]