مقاله پژوهشی

مجله انفورماتیک سلامت و زیست پزشکی
مرکز تحقیقات انفورماتیک پزشکی
دوره پنجم شماره چهارم. ص ٤٤٦-٤٣٥، ١٣٩٧

استفاده از تکنیک‌های داده‌کاوی جهت تشخیص افتراقی بیماری‌های فقر آهن و بتا–تالاسمی مینور

سمیرا نوفرستی[1]، نرگس شمشادی نژاد[2]*، فاطمه حیدری[3]

• دریافت مقاله: ۹۷/۴/۲۷ • پذیرش مقاله: ۹۷/۸/۲۰

مقدمه: کم‌خونی، فقرآهن یکی از شایع‌ترین انواع کم خونی است که تشخیص افتراقی اصلی آن بتا–تالاسمی مینور می‌باشد. غربالگری سریع و دقیق بتا–تالاسمی مینور جهت مشاوره پزشکی قبل از ازدواج و جلوگیری از تولد نوزادان مبتلا به بتا–تالاسمی ماژور و تمایز آن از فقرآهن برای پیشگیری از تجویز نابه‌جای آهن برای درمان بتا–تالاسمی مینور از اهمیت ویژه‌ای برخوردار است. هدف مطالعه حاضر به‌کارگیری تکنیک‌های داده‌کاوی جهت افتراق فقر آهن از بتا–تالاسمی مینور بر اساس آزمایش‌های CBC به منظور افزایش سرعت تشخیص و کاهش هزینه‌های تشخیصی است.

روش: پژوهش حاضر از نوع گذشته‌نگر و بر روی داده‌های ۱۰۰۰ بیمار در آزمایشگاه دکتر حیدری شهرستان زاهدان انجام گرفت. برای انجام تحقیق از روش استاندارد CRISP-DM و الگوریتم‌های داده‌کاوی ماشین بردار پشتیبان، بیزین ساده، بگینگ، آدابوست و درخت تصمیم استفاده شد. برای تحلیل داده‌ها نرم‌افزار Weka به کار رفت.

نتایج: نتایج ارزیابی‌های انجام گرفته نشان می‌دهد که الگوریتم‌های بگینگ، درخت تصمیم، آدابوست، ماشین بردار پشتیبان و بیزین ساده در افتراق فقر آهن از بتا–تالاسمی مینور به ترتیب به دقت ۹۵/۷۳، ۹۵/۵، ۹۴/۶، ۸۰/۲، ۷۶/۶ درصد دست یافته‌اند.

نتیجه‌گیری: در این تحقیق روشی خودکار مبتنی بر تکنیک‌های داده‌کاوی برای افتراق فقر آهن از بتا–تالاسمی مینور ارائه شد. نتایج ارزیابی‌ها نشان می‌دهد که الگوریتم بگینگ در افتراق فقر آهن از بتا–تالاسمی مینور به دقت بالاتری در مقایسه با سایر الگوریتم‌های داده‌کاوی و شاخص‌های افتراقی دست یافت. همچنین به کمک درخت تصمیم ایجاد شده، قوانینی استخراج شده‌اند که می‌توانند در تشخیص به موقع دو بیماری مذکور توسط پزشک مورد استفاده قرار گیرند.

کلیدواژه‌ها: فقر آهن، بتا–تالاسمی مینور، تشخیص افتراقی، داده‌کاوی، الگوریتم یادگیری جمعی بگینگ

•ارجاع: نوفرستی سمیرا، شمشادی نژاد نرگس، حیدری فاطمه. استفاده از تکنیک‌های داده کاوی جهت تشخیص افتراقی بیماری‌های فقر آهن و بتا–تالاسمی مینور. مجله انفورماتیک سلامت و زیست پزشکی ۱۳۹۷؛ ۵(۴): ۴۳۵-۴۴۶.

۱. دکترای مهندسی کامپیوتر، استادیار گروه مهندسی فناوری اطلاعات، دانشکده مهندسی برق و کامپیوتر، دانشگاه سیستان و بلوچستان، زاهدان، ایران
۲. کارشناسی ارشد مهندسی نرم‌افزار، زاهدان، ایران
۳. متخصص پاتولوژی (آسیب‌شناسی بالینی و تشریحی)، استادیار، دانشکده پزشکی، دانشگاه علوم پزشکی، زاهدان، ایران
*** نویسنده مسئول:** سیستان و بلوچستان، شهر زاهدان، خیابان دانشگاه، دانشگاه ۳۵
• شماره تماس: ۰۹۱۵۵۴۰۶۹۷۵
• **Email:** shemshadi.narges@gmail.com

[Downloaded from jhbmi.com on 2026-11-15]

مقدمه

آنمی میکروسیتیک اختلالی است که در آن اندازه و حجم گلبول‌های قرمز کوچک‌تر از حد طبیعی بوده و این اختلال همراه با کاهش تولید هموگلوبین می‌باشد [۱]. تشخیص بتا-تالاسمی مینور به‌منظور غربالگری قبل از ازدواج، برای جلوگیری از به دنیا آمدن نوزادی با بتا-تالاسمی ماژور استفاده می‌شود [۲]. همچنین ممکن است بتا-تالاسمی مینور با فقر آهن اشتباه گرفته شود و درمان بر اساس مصرف آهن قرار گیرد [۳]، بدین دلیل افتراق فقر آهن از بتا-تالاسمی مینور حائز اهمیت می‌باشد.

روش‌های اصلی غربالگری بتا-تالاسمی مینور عبارت‌اند از: شمارش کامل گلبول‌های قرمز، اندازه‌گیری سطح آهن و فریتین سرم، ظرفیت تام اتصال آهن، بررسی ذخایر آهن مغز استخوان، سطح هموگلوبین A2 و پروتوپورفیرین آزاد اریتروسیتی [۴]. علی‌رغم این که این آزمایش‌ها بسیار مؤثر می‌باشند؛ اما معمولاً وقت‌گیر و پرهزینه بوده و ممکن است در تمام مراکز درمانی قابل اجرا نباشد. درحالی‌که استفاده از تکنیک‌های داده‌کاوی می‌تواند یکی دیگر از روش‌های افتراق فقر آهن از بتا-تالاسمی مینور باشد. تکنیک‌های داده‌کاوی برای تحلیل خودکار داده‌ها، پیش‌بینی و تشخیص سریع و کم هزینه بیماری‌ها، می‌تواند نتایج مناسب‌تری را نسبت به روش‌های سنتی ارائه کند [۶، ۵]. همکاری متخصصان در زمینه کامپیوتر و پزشکی راه‌حل جدیدی در تحلیل داده‌های پزشکی و به دست آوردن الگوهای مفید و کاربردی ارائه می‌دهد که همان داده‌کاوی پزشکی است [۷]. از جمله خدمات داده‌کاوی در پزشکی می‌توان به بررسی میزان تأثیر دارو بر بیماری و اثرات جانبی آن، تعیین نوع درمان، پیش‌بینی انواع بیماری‌ها مانند سرطان، رتبه‌بندی بیمارستان‌ها و کنترل عفونت بیمارستانی اشاره کرد [۹، ۸]. به‌عنوان‌مثال اداره دارو و غذای آمریکا برای کشف دانش درباره عوارض جانبی داروها از الگوریتم داده‌کاوی(Multi-item gamma poisson shrinker) MGPS استفاده کرده است. این روش توانست با موفقیت ۶۷ درصد عوارض جانبی داروها را ۵ سال زودتر از شیوه سنتی شناسایی کند [۱۰].

همچنین در بیمارستان شهید هاشمی‌نژاد تهران برای تعیین نوع درمان سنگ حالب از راهکار داده‌کاوی استفاده شده است. در این بیمارستان یک الگوریتم درختی با دقت ۷۷ درصد وجود دارد که پزشک بر اساس آن درمانی را که میزان موفقیت بالاتری برای بیمار دارد انتخاب می‌کند [۱۱].

بررسی تحقیقات موجود نشان می‌دهد تاکنون مطالعه‌ای در زمینه استفاده از تکنیک‌های داده‌کاوی برای افتراق فقر آهن از بتا-تالاسمی مینور صورت نگرفته است؛ اما مطالعاتی در زمینه تشخیص بیماری‌های کم‌خونی و به خصوص پیش‌بینی فقر آهن با استفاده از تکنیک‌های داده‌کاوی انجام شده است. صفایی و همکاران [۱۲] برای پیش‌بینی سطح فریتین سرم در جمعیت زنان از درخت تصمیم استفاده کرده‌اند. با توجه به درخت تصمیم به دست آمده، متغیرهای (Mean Corpuscular Hemoglobin)MCH، (Red Blood Cell) RBC، (Mean Corpuscular Hemoglobin Concentration) MCHC، زخم‌های معده-روده و سرطان معده-روده به‌عنوان مهم‌ترین عوامل پیش‌بینی کننده شناخته شده‌اند. قوانین به دست آمده از مدل تصمیم می‌تواند فرآیند تشخیص و درمان بیماران مبتلا به کم‌خونی فقر آهن را بهبود بخشد.

در پژوهش انجام گرفته توسط Abdullah و Al-Asmari به تشخیص نوع کم‌خونی در افراد مبتلا به آنمی با استفاده از تکنیک‌های داده‌کاوی پرداخته شده است. مجموعه داده‌ای مورداستفاده، آزمایش‌های Complete Blood Count یا CBC مربوط به ۴۱ بیمار دارای کم‌خونی است. برای تعیین نوع کم‌خونی از الگوریتم‌های ماشین بردار پشتیبان، شبکه عصبی، بیزین ساده و درخت تصمیم در نرم‌افزار Weka استفاده شده است. نتایج آزمایش‌ها نشان داد که الگوریتم درخت تصمیم J48 به دقت حدود ۹۴ درصد در پیش‌بینی نوع کم‌خونی دست یافته است [۱۳].

هدف تحقیق انجام گرفته توسط Dithy و KrishnaPriya مطالعه شیوع آنمی در زنان باردار و نوجوانان و نیز یافتن رابطه بین فقر آهن و عوامل دموگرافیک است. در این مقاله از الگوریتم‌های داده‌کاوی برای پیش‌بینی کلاس آنمی (خفیف، متوسط، شدید و غیرآنمی) استفاده نموده‌اند [۱۴].

هدف پژوهش حاضر استفاده از الگوریتم‌های داده‌کاوی جهت ساخت مدلی برای افتراق فقر آهن از بتا-تالاسمی مینور است. الگوریتم‌های داده‌کاوی مورداستفاده در این تحقیق عبارت‌اند از: درخت تصمیم، ماشین بردار پشتیبان، بیزین ساده، بگینگ و آدابوست. شرح مفصل این الگوریتم‌ها در کتاب Data mining concepts and techniques آورده شده است [۱۵].

درخت تصمیم C4.5: درخت تصمیم از کاربردی‌ترین و محبوب‌ترین روش‌های طبقه‌بندی محسوب می‌شود. این

[Downloaded from jhbmi.com on 2026-11-15]

درخت یک روش گرافیکی قابل درک است که به کمک مجموعه‌ای از قوانین به پیش‌بینی مقادیر متغیر هدف می‌پردازد. در این مطالعه از الگوریتم درخت تصمیم J48 در نرم‌افزار Weka نسخه 3.9.2 که همان پیاده‌سازی درخت تصمیم معروف به C4.5 است استفاده شد.

ماشین بردار پشتیبان: اساس کاری این الگوریتم دسته‌بندی خطی داده‌ها است که برای داده‌های با ابعاد بالا پیش‌بینی‌های موفقیت‌آمیز دارد و صحت آن نسبت به سایر طبقه‌بندهای شناخته شده بالاتر است.

بیزین ساده: این الگوریتم بر پایه قضیه بیز برای مدل‌سازی پیش‌گویانه ارائه شده است. الگوریتم بیز برای طبقه‌بندی دودویی و چندگانه نتایج دقیقی ارائه می‌دهد.

طبقه‌بند بگینگ: بگینگ یک الگوریتم یادگیری جمعی است که برای بهبود دادن طبقه‌بندی و مدل‌های پس‌رفتی بر حسب پایداری و دقت استفاده می‌شود. این روش همچنین واریانس را کاهش داده و به اجتناب از بیش‌برازش کمک می‌کند.

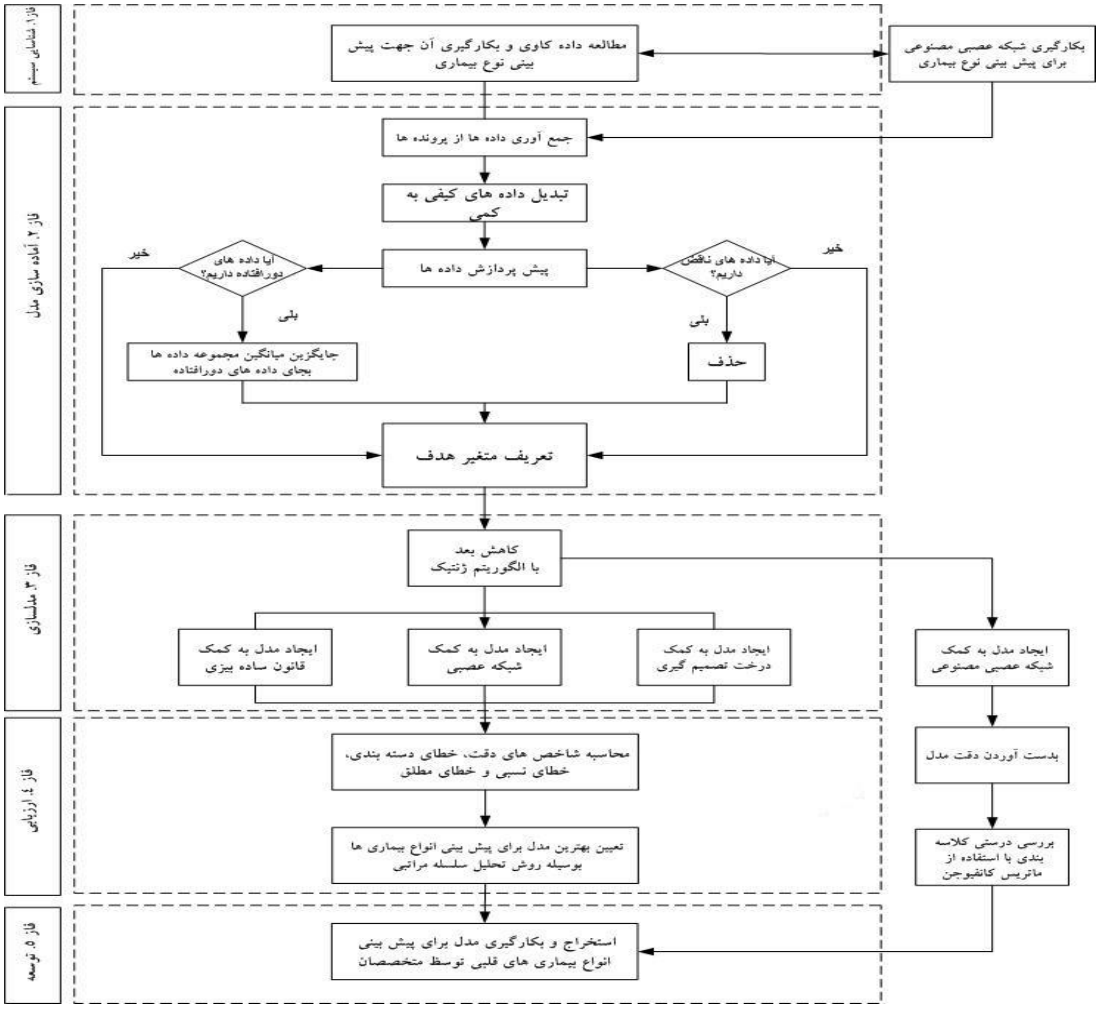
طبقه‌بند آدابوست: آدابوست یک الگوریتم یادگیری جمعی است که مخفف بوستینگ تطبیقی بوده و به‌منظور ارتقاء عملکرد و رفع مشکل دسته‌های نامتوزان، همراه دیگر الگوریتم‌های یادگیری استفاده می‌شود.

نتایج ارزیابی‌های انجام گرفته نشان می‌دهد الگوریتم‌های داده‌کاوی مذکور به دقت بالایی در افتراق فقر آهن از بتا-تالاسمی مینور دست یافته‌اند. برای نمونه الگوریتم بگینگ به دقت ۹۵/۷۳ درصد رسیده است. همچنین مقایسه الگوریتم بگینگ و شاخص‌های افتراقی بر روی ۵۱۱ بیمار مبتلا به فقر آهن و بتا-تالاسمی مینور نشان دهنده دقت بالاتر الگوریتم بگینگ نسبت به شاخص‌های افتراقی Mentzer، England و Fraser، Srivastava، RDWI (Red Blood Cell Distribution Width Index)، Sirdah و MCI است.

روش

برای انجام این تحقیق از روش (Cross Industry Standard Process for Data Mining) CRISP_DM استفاده شد. چرخه حیات یک پروژه داده‌کاوی در روش CRISP_DM از شش مرحله تشکیل شده است. توالی مراحل مستقیم نیست و حرکت به عقب و جلو بین مراحل مختلف همیشه نیاز است. خروجی هر مرحله مشخص می‌کند که بعد از آن باید چه مرحله‌ای اجرا شود. بردارها وابستگی‌های مهم بین مراحل را مشخص می‌کند. این چرخه در شکل ۱ نمایش داده شد.

[Downloaded from jhbmi.com on 2026-11-15]



شکل ۱: مراحل انجام پژوهش

شناخت مسئله: در این مرحله به شناخت سیستم و بیان اهداف مسئله موردنظر پرداخته شد. تولد نوزادان مبتلا به بتا-تالاسمی ماژور مشکلاتی که این بیماری برای مبتلایان و خانواده‌های آنان به همراه دارد و همچنین موارد مشاهده شده از درمان نادرست بیماری بتا-تالاسمی مینور بر اساس مصرف آهن، مسئله پیش‌بینی و افتراق فقر آهن از بتا-تالاسمی مینور را حائز اهمیت می‌سازد. هدف تحقیق حاضر، ارائه مدل پیش‌بینی کننده‌ای برای افتراق فقر آهن از بتا-تالاسمی مینور به‌منظور افزایش سرعت تشخیص و نیز کاهش هزینه‌های تشخیصی می‌باشد.

جمع‌آوری و درک داده: مطالعه حاضر از نوع گذشته‌نگر بود و داده‌های آن مربوط به پرونده ۱۰۰۰ فرد بیمار و سالم از آزمایشگاه دکتر حیدری شهرستان زاهدان است که در سال ۱۳۹۶ با مراجعه مستقیم پژوهشگر به آزمایشگاه مذکور در قالب فایل اکسل تهیه و موردبررسی قرار گرفت. از تعداد افراد موردبررسی ۴۸۹ فرد سالم، ۳۹۲ فرد مبتلا به فقر آهن و ۱۱۹ فرد مبتلا به بتا-تالاسمی مینور بودند. ۷۰ درصد افراد موردمطالعه را زنان و ۳۰ درصد را مردان تشکیل می‌دهند. همچنین سن بیماران در بازه ۵ ماه تا ۱۰۰ سال قرار دارد.

داده‌های موردبررسی شامل ۱۰ ویژگی درباره بیماران است که با مشورت متخصص مربوطه انتخاب شده و از پرونده بیماران و نتایج آزمایش‌های CBC آن‌ها گردآوری شده‌اند. جدول ۱ مشخصات ویژگی‌های تحقیق، نوع و محدوده مقادیر آن‌ها را نشان می‌دهد. برای پردازش داده‌ها از نرم‌افزار Weka استفاده شد.

[Downloaded from jhbmi.com on 2026-11-15]

جدول ۱: ویژگی‌های موردبررسی و نوع آن‌ها

ردیف	صفت	توضیحات	نوع	محدوده مقادیر
۱	Sex	جنسیت	اسمی	{زن، مرد}
۲	Age	سن	عددی	[۱۰۰...۰]
۳	HB	هموگلوبین	عددی	[۲۳/۹ ...۵/۸]
۴	HCT	هماتوکریت	عددی	[۶۰/۷..۱۸/۴]
۵	RBC	شمارش گلبول قرمز	عددی	[۷/۷۲...۲/۸۱]
۶	MCV	حجم متوسط گلبول قرمز	عددی	[۱۱۹/۳...۵۳]
۷	MCH	هموگلوبین متوسط گلبول‌های قرمز	عددی	[۴۳/۳...۱۳/۷]
۸	MCHC	غلظت متوسط هموگلوبین گلبول‌های	عددی	[۳۸/۴...۲۲/۷]
۹	RDW	وسعت انتشار گلبول قرمز	عددی	[۳۴/۱...۱۱/۲]
۱۰	Ferritin	فریتین	عددی	[۷۱۷/۳...۱/۵۲]

Hematocrit
Red Blood Cell Distribution Width

آماده‌سازی داده: در این مرحله، داده‌ها پاک‌سازی و سپس به شکل مناسب جهت پردازش توسط ماشین در آمدند. در این تحقیق، صحت داده‌ها با بررسی مجاز و منطقی بودن مقادیر آن‌ها توسط متخصص مربوطه کنترل شد. در این راستا مقادیر اشتباه تا حد امکان اصلاح و در صورتی که امکان اصلاح وجود نداشته باشد، حذف شده‌اند. همچنین به دلیل پراکندگی مقادیر صفت سن، عدد سن افراد گرد شده و برای سن افراد زیر یک سال نیز کد صفر در نظر گرفته شد.

مدل‌سازی: در این مرحله، تکنیک‌های مختلف مدل‌سازی انتخاب و به کار گرفته شد. در این تحقیق با استفاده از تکنیک طبقه‌بندی به تولید مدل و الگوی بهینه‌ای برای افتراق فقر آهن از بتا–تالاسمی مینور پرداخته شد. برای ساخت مجموعه آموزش و مجموعه آزمایش از تکنیک رایج ارزیابی متقابل با ۱۰ حلقه (10-Fold Cross Validation) استفاده شد. در این تکنیک مجموعه داده‌ها به‌صورت تصادفی به ۱۰ قسمت مساوی تقسیم شد که در هر تکرار از الگوریتم یک بخش به عنوان آزمایش و ۹ بخش دیگر به عنوان آموزش انتخاب شدند. در واقع در هر اجرا ۹۰ درصد داده‌ها معادل ۹۰۰ رکورد داده‌ای به عنوان داده‌های آموزشی و ۱۰ درصد باقی‌مانده معادل ۱۰۰ رکورد داده‌ای به عنوان آزمایش انتخاب شدند. نهایتاً میانگین ۱۰ بار تکرار الگوریتم به عنوان نتیجه نهایی انتخاب شد.

در مطالعه حاضر سه برچسب برای متغیر هدف در نظر گرفته شد که عبارت‌اند از: افراد سالم، افراد مبتلا به فقر آهن، افراد مبتلا به بتا–تالاسمی مینور. برچسب داده‌های موردبررسی توسط متخصص مربوطه مشخص و مورد تأیید قرار گرفت. با توجه به این‌که متغیر هدف دارای سه برچسب است، نیاز به الگوریتم طبقه‌بندی است که بتواند با برچسب‌های غیر دودویی کار کند. بدین دلیل برای مدل‌سازی از الگوریتم‌های بیزین ساده، ماشین بردار پشتیبان، درخت تصمیم، بگینگ و آدابوست استفاده شد.

ارزیابی مدل: به منظور مقایسه مدل‌های ساخته شده از معیار دقت استفاده شده است که به صورت فرمول ۱ تعریف می‌شود:

$$\text{دقت} = \frac{TP+TN}{TP+FP+TN+FN} \quad (۱)$$

[Downloaded from jhbmi.com on 2026-11-15]

متغیر TP یا «مثبت صحیح» به تعداد رکوردهای مثبت از مجموعه آزمایش اشاره دارد که توسط الگوریتم داده‌کاوی به‌درستی دسته‌بندی شده‌اند یعنی برچسب مثبت گرفته‌اند. به‌طور مشابه متغیر TN یا «منفی صحیح» به تعداد رکوردهای منفی از مجموعه آزمایش اشاره دارد که توسط الگوریتم داده‌کاوی به درستی دسته‌بندی شده‌اند، یعنی برچسب منفی گرفته‌اند. متغیر FP یا «مثبت کاذب» نیز تعداد رکوردهای منفی از مجموعه آزمایش که اشتباهاً به عنوان مثبت دسته‌بندی شده‌اند را نشان می‌دهد. FN یا «منفی کاذب» تعداد رکوردهای مثبت مجموعه آزمایش که به اشتباه برچسب منفی خورده‌اند را نشان می‌دهد. علاوه بر دقت، شاخص‌های حساسیت، ویژگی، ارزش اخباری مثبت و ارزش اخباری منفی نیز برای مقایسه مدل‌ها مورداستفاده قرار گرفته‌اند که در ادامه فرمول محاسبه هرکدام ارائه شده است.

$$\text{حساسیت} = \frac{TP}{TP+FN} \quad (۲)$$

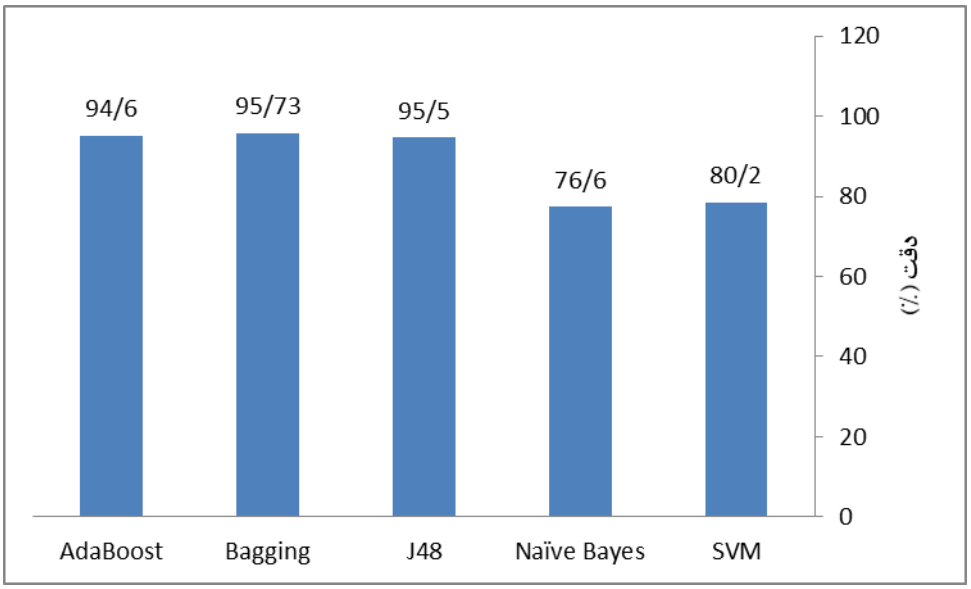
$$\text{ویژگی} = \frac{TN}{TN+FP} \quad (۳)$$

$$\text{ارزش اخباری مثبت} = \frac{TP}{TP+FP} \quad (۴)$$

$$\text{ارزش اخباری منفی} = \frac{TN}{TN+FN} \quad (۵)$$

نتایج

دقت طبقه‌بندهای درخت تصمیم، ماشین بردار پشتیبان، بیزین ساده و الگوریتم‌های یادگیری جمعی بگینگ و آدابوست بر روی مجموعه داده‌ای توصیف شده در این مطالعه مورد آزمایش قرار گرفت. شکل ۲ نشان می‌دهد که به ترتیب الگوریتم‌های یادگیری جمعی بگینگ، درخت تصمیم و آدابوست به بالاترین دقت در افتراق فقر آهن از بتا–تالاسمی مینور دست یافته‌اند.



شکل ۲: مقایسه دقت الگوریتم‌های طبقه‌بندی

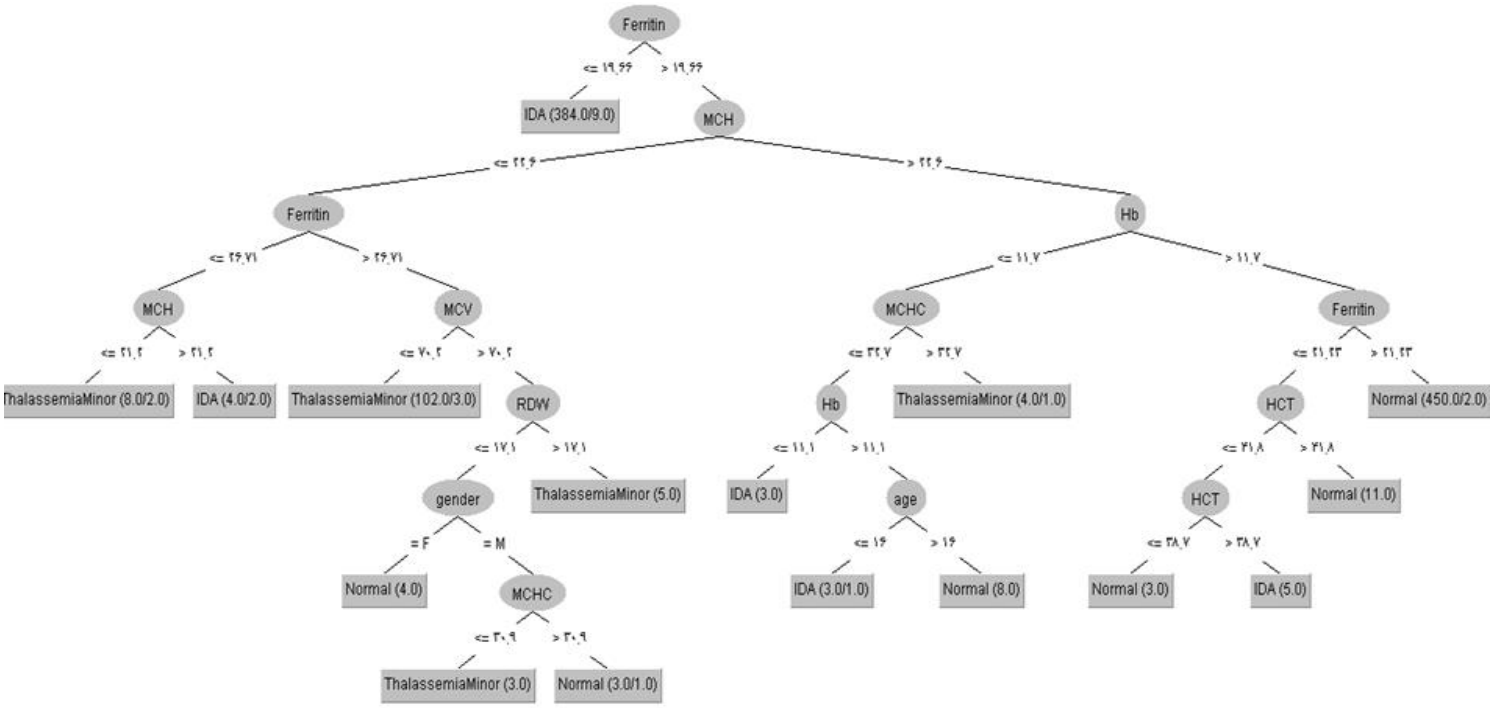
در جدول ۲، الگوریتم‌های طبقه‌بندی بر اساس معیارهای دقت، حساسیت، ویژگی، ارزش اخباری مثبت و ارزش اخباری منفی مقایسه شده‌اند. الگوریتم بگینگ بر اساس همه معیارها (به جزء ارزش اخباری منفی در مقایسه با درخت تصمیم که نتایج بسیار نزدیک است) بهتر از سایر الگوریتم‌ها عمل کرده است؛ بنابراین الگوریتم بگینگ مدلی مناسب برای پیش‌بینی و افتراق فقر آهن از بتا–تالاسمی مینور است.

[Downloaded from jhbmi.com on 2026-11-15]

جدول ۲: مقایسه طبقه‌بندهای مورداستفاده بر اساس شاخص‌های مختلف

الگوریتم	دقت	ویژگی	حساسیت	ارزش اخباری مثبت	ارزش اخباری منفی
ماشین بردار پشتیبان	۸۰/۲	۸۶/۵۶	۸۰/۲	۸۰/۹۹	۸۵/۰۶
بیزین ساده	۷۶/۶	۸۴/۱۹	۷۶/۶	۷۶/۷۴	۸۲/۷۳
درخت تصمیم	۹۵/۵	۹۷/۲	۹۵/۵	۹۵/۵۱	۹۷/۰۹
بگینگ	۹۵/۷۳	۹۷/۳۴	۹۵/۷۳	۹۵/۷۳	۹۶/۹
آدابوست	۹۴/۶	۹۶/۱۲	۹۴/۶	۹۴/۵۹	۹۶/۶

نتایج درخت تصمیم برای پیش‌بینی افتراق فقر آهن از بتا–تالاسمی مینور در شکل ۳ نشان داده شد. همچنین قوانین مستخرج از درخت تصمیم که از پیمایش تک‌تک شاخه‌های درخت به دست می‌آید در شکل ۴ آورده شد. با توجه به تعداد زیاد قوانین تولید شده، می‌توان تنها تعدادی از آن‌ها را انتخاب و در تصمیم‌گیری‌های پزشکی مورداستفاده قرار داد. برای انتخاب قوانین دو تکنیک رایج عبارت‌اند از: (۱) انتخاب قوانین دارای خطای کمتر و (۲) انتخاب قوانین مورد تأیید پزشک متخصص مربوطه. در درخت تصمیم ترسیم شده در شکل ۳ در مستطیل‌های انتهای هر شاخه علاوه بر برچسب آن شاخه (سالم، فقر آهن و بتا–تالاسمی مینور) درون پرانتز دو عدد آورده شد که اولین عدد تعداد نمونه‌های پوشش داده شده توسط قانون متناظر با آن شاخه و عدد دوم تعداد نمونه‌های خطا یعنی نمونه‌هایی که قانون به آن‌ها برچسب اشتباه داده است را نشان می‌دهد. با تقسیم تعداد نمونه‌های خطا بر تعداد نمونه‌های پوشش داده شده می‌توان خطای قانون را محاسبه کرد و تنها قوانین دارای خطای کمتر را برگزید. با این وجود انتخاب و ساده‌سازی قوانین توسط متخصص رایج‌تر است.



شکل ۳: درخت تصمیم برای افتراق کم‌خونی فقر آهن از بتا–تالاسمی مینور

[Downloaded from jhbmi.com on 2026-11-15]

```
Ferritin <= ۱۹/۶۶: IDA (384.0/9.0)
Ferritin > ۱۹/۶۶
|MCH <= ۲۲/۶
| |Ferritin <= ۲۶/۷۱
| | |MCH <= ۲۱/۲: ThalassemiaMinor (8.0/2.0)
| | |MCH > ۲۱/۲: IDA (4.0/2.0)
| |Ferritin > ۲۶/۷۱
| | |MCV <= ۷۰/۲: ThalassemiaMinor (102.0/3.0)
| | |MCV > ۷۰/۲
| | | |RDW <= ۱۷/۱
| | | | |gender = F: Normal (4.0)
| | | | |gender = M
| | | | | |MCHC <= ۳۰/۹: ThalassemiaMinor (3.0)
| | | | | |MCHC > ۳۰/۹: Normal (3.0/1.0)
| | | |RDW > ۱۷/۱: ThalassemiaMinor (5.0)
|MCH > ۲۲/۶
| |Hb <= ۱۱/۷
| | |MCHC <= ۳۲/۷
| | | |Hb <= ۱۱/۱: IDA (3.0)
| | | |Hb > ۱۱/۱
| | | | |age <= ۱۶: IDA (3.0/1.0)
| | | | |age > ۱۶: Normal (8.0)
| | |MCHC > ۳۲/۷: ThalassemiaMinor (4.0/1.0)
| |Hb > ۱۱/۷
| | |Ferritin <= ۲۱/۲۳
| | | |HCT <= ۴۱/۸
| | | | |HCT <= ۳۸/۷: Normal (3.0)
| | | | |HCT > ۳۸/۷: IDA (5.0)
| | | |HCT > ۴۱/۸: Normal (11.0)
| | |Ferritin > ۲۱/۲۳: Normal (450.0/2.0)
```

شکل ۴: قوانین استخراج شده از درخت تصمیم

در جدول ۳ تعدادی از قوانین تأیید شده توسط پزشک متخصص نمایش داده شد. احتمالاتی که در شرح هر قانون ذکر شد، در واقع دقت آن قانون را نشان می‌دهد. برای محاسبه دقت یک قانون مطابق روش توصیف شده در بخش مدل‌سازی عمل شده است. به این صورت که درخت تصمیم بر روی مجموعه آموزش ساخته شد. با کمک درخت تصمیم حاصل مجموعه قوانین استخراج شده‌اند و دقت هر قانون به صورت مجزا بر روی داده‌های مجموعه آزمایش محاسبه شد. دقت یک قانون عبارت است از درصد داده‌ای مجموعه آزمایش که برچسب سالم، فقر آهن یا بتا-تالاسمی مینور آن به درستی توسط آن قانون پیش‌بینی شده باشد.

جدول ۳: قوانین انتخاب شده توسط متخصص

ردیف	قانون
۱	اگر فریتین بالاتر از ۱۹، MCH بالاتر از ۲۲، HB کمتر از ۱۱ و MCHC کمتر از ۳۲ باشد فرد با احتمال ۱۰۰٪ مبتلا به فقر آهن است.
۲	اگر فریتین بالاتر از ۲۶، MCH کمتر از ۲۲، MCV بالاتر از ۷۰، RDW کمتر از ۱۷، جنسیت مرد و MCHC کمتر از ۳۰ باشد فرد با احتمال ۱۰۰٪ مبتلا به بتا-تالاسمی مینور است.
۳	اگر فریتین مابین ۱۹-۲۶، MCH کمتر از ۲۲ و MCV کمتر از ۷۰ باشد فرد با احتمال ۹۷٪ مبتلا به بتا-تالاسمی مینور است.

در آزمایش پایانی دقت الگوریتم بگینگ با تعدادی از شاخص‌های رایج افتراق فقر آهن از بتا-تالاسمی مینور که در جدول ۴ معرفی شده‌اند، مقایسه شد. برای انجام این آزمایش در ابتدا افراد سالم از مجموعه داده‌ای حذف شده‌اند و بر روی ۵۱۱ نفر باقی‌مانده (۳۹۲ نفر مبتلا به فقر آهن و ۱۱۹ نفر مبتلا به بتا-تالاسمی مینور) ابتدا شاخص‌های افتراقی اعمال شده و تعداد تشخیص‌های درست هر شاخص ثبت شده است. سپس الگوریتم بگینگ به افتراق فقر آهن و بتا-تالاسمی مینور بر

[Downloaded from jhbmi.com on 2026-11-15]

روی این مجموعه داده‌ای پرداخته و نتایج حاصل ثبت شد. جدول ۵ نشان می‌دهد که الگوریتم بگینگ قادر به افتراق صحیح ۴۹۱ مورد فقر آهن و بتا-تالاسمی مینور است درحالی‌که از میان شاخص‌های افتراقی بهترین نتیجه حاصل متعلق به شاخص پیشنهادی مطالعه Ehsani و همکاران [۱۹] است که توانسته است ۴۵۰ مورد از دو بیماری مذکور را به درستی تشخیص دهد.

جدول ۴: شاخص‌های افتراقی فقر آهن و بتا-تالاسمی مینور و فرمول محاسبه آن‌ها

ردیف	نام شاخص	فرمول	به نفع فقر آهن	به نفع تالاسمی مینور
۱	Mentzer [۱۶]	MCV/RBC	>۱۳	<۱۳
۲	England و Fraser [۱۷]	MCV-RBC-5Hb-3.4	>۰	<۰
۳	Srivastava [۱۸]	MCH/RBC	>۳,۸	<۳,۸
۴	Ehsani و همکاران [۱۹]	MCV-(10*RBC)	>۱۵	<۱۵
۵	RDWI [۲۰]	MCV*RDW/RBC	>۲۲۰	<۲۲۰
۶	Sirdah [۲۱]	MCV−RBC− (3*Hb)	>۲۷	<۲۷
۷	MCI [۲۲]	(1.91*RBC)+(0.44*MCHC)	<۲۳/۸۵	>۲۳/۸۵

جدول ۵: تشخیص صحیح موارد فقر آهن و بتا-تالاسمی مینور

شاخص افتراقی	تعداد بیماران درست تشخیص داده شده			دقت
	فقر آهن (تعداد کل بیماران مبتلا به فقر آهن =۳۹۲)	بتا-تالاسمی مینور (تعداد کل بیماران مبتلا به بتا-تالاسمی مینور =۱۱۹)	مجموع (تعداد کل بیماران=۵۱۱)	
Mentzer [۱۶]	۳۵۰	۹۸	۴۴۸	۸۷/۶۷
England و Fraser [۱۷]	۳۳۱	۶۳	۳۹۴	۷۷/۱
Srivastava [۱۸]	۳۵۷	۸۹	۴۴۶	۸۷/۲۸
Ehsani و همکاران [۱۹]	۳۵۸	۹۲	۴۵۰	۸۸/۰۶
RDWI [۲۰]	۲۷۶	۹۸	۳۷۴	۷۳/۱۹
Sirdah [۲۱]	۳۶۳	۶۹	۴۳۲	۸۴/۵۴
MCI [۲۲]	۲۳۶	۸۲	۳۱۸	۶۲/۲۳
الگوریتم بگینگ (روش پیشنهادی)	۳۷۹	۱۱۲	۴۹۱	۹۸/۲۳

بحث و نتیجه‌گیری

در این پژوهش به افتراق فقر آهن از تالاسمی مینور با استفاده از تکنیک‌های داده‌کاوی پرداخته شد. تکنیک‌های استفاده شده در این پژوهش، الگوریتم‌های یادگیری جمعی (بگینگ و آدابوست) و طبقه‌بندهای تکی (درخت تصمیم، ماشین بردار پشتیبان و بیزین ساده) است. با توجه به نتایج ارزیابی، دقت الگوریتم‌های بگینگ، درخت تصمیم و آدابوست به‌ترتیب عبارت‌اند از: ۹۵/۷۳، ۹۵/۵ و ۹۴/۶ درصد که در مقایسه با الگوریتم‌های ماشین بردار پشتیبان و بیزین ساده به دقت بیشتری رسیده‌اند.

برای هرکدام از طبقه‌بندهای مذکور علاوه بر دقت، معیارهای حساسیت، ویژگی، ارزش اخباری مثبت و ارزش اخباری منفی محاسبه گردید. نتایج ارزیابی نشان می‌دهد الگوریتم بگینگ بر اساس اغلب معیارها بهتر از سایر الگوریتم‌ها عمل کرده است و به ویژگی ۹۷/۳۴ درصد، حساسیت ۹۵/۷۳ درصد، ارزش اخباری مثبت ۹۵/۷۳ درصد و ارزش اخباری منفی ۹۶/۹ درصد دست یافته است که برای فقر آهن و بتا-تالاسمی مینور قابل‌قبول به نظر می‌رسد.

همچنین به کمک درخت تصمیم ایجاد شده، قوانینی استخراج شد که در تشخیص به موقع دو بیماری فقر آهن و بتا-تالاسمی مینور توسط پزشک قابل استفاده می‌باشند. همان‌طور که در جدول ۳ مشاهده شد قوانینی که به تأیید متخصص رسیده‌اند قادرند با دقت بیشتر از ۹۷ درصد به افتراق این دو بیماری بپردازند. به‌علاوه بالا بودن حجم نمونه آماری به نتایج مطالعه حاضر اعتبار می‌بخشد. در مطالعه حاضر مدل‌سازی بر روی مجموعه داده‌ای با ۱۰۰۰ رکورد انجام گرفت، درحالی‌که در اغلب مطالعات پیشین تعداد نمونه‌های موردبررسی کمتر از ۵۰۰ و گاهی زیر ۲۰۰ هستند.

با توجه به اهمیت افتراق فقر آهن از بتا-تالاسمی مینور، مطالعاتی توسط پژوهشگران در این حوزه انجام گرفته است. تمرکز مطالعات پیشین بر روی افتراق سریع این دو بیماری با استفاده از شاخص‌های گلبول قرمز بوده است. برای نمونه غفوری و همکاران [۲۳] به بررسی مقایسه‌ای شاخص‌های

[Downloaded from jhbmi.com on 2026-11-15]

CBC بتا-تالاسمی مینور و فقر آهن پرداخته‌اند و حساسیت و ویژگی را در ۴ شاخص Shine و Lal، Mentzer، England و Fraser و تراکم متوسط هموگلوبین در لیتر خون مورد بررسی قرار داده‌اند. نتایج نشان داد شاخص Mentzer با حساسیت ۹۰/۹ درصد و ویژگی ۸۰/۳ بهترین شاخص در شناسایی بیماران تالاسمی محسوب می‌شود. همچنین مشخص شد میزان حساسیت شاخص‌های تراکم متوسط هموگلوبین در لیتر خون و شاین برای تشخیص بتا-تالاسمی مینور بسیار کمتر از آن است که بتوانند نقش یک آزمایش غربالگر را ایفا کنند.

در پژوهش دیگری که کیخایی و همکاران [۲۴] با هدف مقایسه شاخص‌های مختلف جهت کمک به تشخیص افتراقی فقر آهن از بتا-تالاسمی مینور انجام دادند، شاخص‌های Mentzer، England و Fraser [۱۷]، Srivastava [۱۸]، Green و King [۲۵]، Shin و Lal [۲۶]، RBC (Red Blood Cell)، RDWI (Red Blood Cell Distribution Width Index) [۲۰]، MCHD و MDHL [۲۷]، محاسبه شده است. در نهایت مشخص شد که هیچ‌کدام از آزمون‌ها، صحت و دقت ۱۰۰ درصد نداشته است؛ ولی در سن زیر ۱۰ سال، آزمون Shine و Lal و شمارش گلبول قرمز و در سن بالای ۱۰ سال، شاخص وسعت انتشار گلبول قرمز RDWI و شمارش گلبول قرمز از صحت و اعتبار بیشتری برخوردار بوده است.

مطالعه‌ای که Ehsani و همکاران [۱۹] بر روی ۲۸۴ بیمار انجام دادند، نتایج نشان داد که دقت شاخص‌های Mentzer، England و Fraser [۱۷]، Srivastava [۱۸] و Klee [۲۸]، در افتراق فقر آهن و بتا-تالاسمی مینور به ترتیب ۸۳/۰۹، ۹۴/۷۱، ۸۶/۹۷ و ۹۲/۶۱ است. در این تحقیق شاخص جدیدی بر مبنای شاخص Mentzer معرفی شده است که در تشخیص صحیح موارد فقر آهن و بتا-تالاسمی مینور به دقت ۹۲/۹۶ درصد رسیده است.

Arora و همکاران [۱] در ارزیابی‌هایی که بر روی شاخص‌های Mentzer، RWDI و MCI انجام دادند به این نتیجه رسیدند که شاخص Mentzer با حساسیت ۹۷/۶۲ درصد و ویژگی ۶۶/۶۷ درصد بیشترین کارایی را در افتراق فقر آهن و بتا-تالاسمی مینور دارد.

مقایسه دقت الگوریتم بگینگ با بهترین دقت حاصل شده از روش‌های مبتنی بر شاخص‌های گلبول قرمز (شاخص‌های افتراقی) کارایی روش پیشنهادی را می‌رساند. در مجموع با به‌کارگیری روش پیشنهادی جهت تشخیص و پیشگیری به‌موقع فقر آهن و بتا-تالاسمی مینور در مشاوره پزشکی قبل از ازدواج می‌تواند از تولد فرزندان تالاسمی ماژور و درمان بتا-تالاسمی مینور بر اساس مصرف آهن جلوگیری شود و بدین‌وسیله مسیری برای انجام بررسی‌های گسترده‌تر در این زمینه فراهم گردد. به‌طور خلاصه نقاط قوت روش پیشنهادی به شرح زیر است:

۱-دقت روش پیشنهادی در مقایسه با شاخص‌های افتراقی معرفی شده در مطالعات پیشین بالاتر است. بهترین دقت حاصل شده توسط شاخص‌های افتراقی معرفی شده در جدول ۵ برابر ۸۸/۰۶ درصد بوده است، در حالی که در روش پیشنهادی الگوریتم بگینگ به دقت ۹۸/۲۳ درصد دست یافته است.

۲- در الگوریتم‌های داده‌کاوی پیشنهادی اثر همه متغیرهای پیشگویی کننده (شاخص‌های CBC) به‌صورت هم‌زمان بررسی می‌شود و همه متغیرها در افتراق بیماران حائز اهمیت هستند. درصورتی‌که مطالعات پیشین تنها بر روی شاخص‌های افتراقی تمرکز داشتند که هر شاخص تنها متکی بر یک یا چند متغیر پیشگویی کننده است. به‌بیان‌دیگر این شاخص‌ها اثر هم‌زمان همه متغیرها را در نظر نمی‌گیرند.

۳- قوانین مستخرج از درخت تصمیم رسم شده در پژوهش به دلیل قابلیت تفسیر ساده می‌تواند به‌آسانی توسط پزشکان مورداستفاده قرار گیرد. در واقع با استفاده از الگوریتم‌های پیش‌بینی می‌توان یک سامانه کمک دستیار پزشک جهت افتراق فقر آهن از بتا-تالاسمی مینور طراحی کرد.

۴- روش پیشنهادی در مقایسه با روش‌های سنتی سرعت تشخیص بالاتر و هزینه‌های تشخیصی کمتری دارد.

۵- روش پیشنهادی کاملاً خودکار است و نیاز به دخالت انسان ندارد.

از جمله محدودیت‌هایی که طی انجام تحقیق حاضر وجود داشت، مشکلات فراوان در مسیر جمع‌آوری داده بود. ثبت سیستمی اطلاعات بیماران به جای استفاده از پرونده می‌تواند تا حد زیادی از مشکلات جمع‌آوری و آماده‌سازی داده و نیز خطای انسانی ناشی از ثبت دستی اطلاعات بکاهد و باعث صرفه‌جویی در زمان شود.

در تحقیقات آینده می‌توان از سایر مشخصه‌های آزمایشگاهی و بالینی نیز برای افتراق فقر آهن از بتا-تالاسمی مینور استفاده کرد. همچنین می‌توان دیگر الگوریتم‌های داده‌کاوی را نیز برای پیش‌بینی دو بیماری مذکور مورداستفاده قرار دارد. به‌علاوه با

[Downloaded from jhbmi.com on 2026-11-15]

به‌کارگیری تکنیک‌های هرس درخت تصمیم می‌توان درخت حاصل را هرس کرد تا به قوانین ساده‌تری دست یافت.

تشکر و قدردانی

بدین‌وسیله از زحمات خانم معصومه دهمرده که با راهنمایی‌های بی‌دریغشان بنده را یاری نمودند و همچنین از پرسنل محترم آزمایشگاه دکتر حیدری زاهدان کمال تشکر و قدردانی را دارم. پژوهش حاضر حاصل یک طرح تحقیقاتی است و بدون حمایت مالی می‌باشد.

تعارض منافع

در پژوهش حاضر هیچ‌گونه تعارض منافعی وجود ندارد.

References

1. Arora S, Rana D, Kolte S, Dawson L, Dhawan I. Validation of new indices for differentiation between iron deficiency anemia and beta thalessemia trait, a study in pregnant females. Int J Sci Rep 2018; 4(2): 26-30.

2. Yasumasa A, Fumio K, Sunil KB, Niriksha BM, Prakash RP, Vijay P, et al. A Study of B Thalassemia Screening Using an Automated Hematology Analyzer. Sysmex Journal Interational 1998;8(2).

3. Morshedi M, Khorshidi S, Johari H, Raafat E. Effect of iron deficiency anemia on amount of HbA2 and comparison to minor thalassemia. Zahedan J Res Med Sci 2012; 13(suppl 1): 60.

4. Madan N, Sikka M, Sharma S, Rusia U, Kela K. Red cell indices and discriminant functions in the detection of beta-thalassaemia trait in a population with high prevalence of iron deficiency anaemia. Indian J Pathol Microbiol 1999;42(1):55-61.

5. Ferreira D, Oliveira A, Freitas A. Applying data mining techniques to improve diagnosis in neonatal jaundice. BMC Med Inform Decis Mak 2012;12:143.

6. Baron-Epel O, Heymann AD, Friedman N, Kaplan G. Development of an unsupportive social interaction scale for patients with diabetes. Patient Preference and Adherence 2015; 9: 1033.

7. Jurian N, Ashoori M. Predicting the effectiveness of preeclampsia medications based on dose and method of drug consumption using data mining. The Iranian Journal of Obstetrics, Gynecology and Infertility 2014;17(123):13-22. Persian

8. Ameri H, Alizadeh S, Barzegari A. Identification of influencing factors for heart attack in diabetic patients using C & R algorithm. Daneshvar Medicine 2014;21(112):71-82. Persian

9. Gholamhosseini L, Damroodi M. Evaluation of data mining applications in the health system. Paramedical Sciences and Military Health 2015;10(1):39-48. Persian

10. Canlas RD. Data mining in healthcare: Current applications and issues [dissertation]. Australia: Carnegie Mellon University; 2009.

11. Sepehri MM, Rahnama P, Shadpour P, Teimourpour B. A data mining based model for selecting type of treatment for kidney stone patient. Tehran Univ Med J 2009; 67(6): 421-7. Persian

12. Safaee P, Noorossana R, Heidari K, Soleimani P. Using decision tree to predict serum ferritin level in women with anemia. Tehran Univ Med J 2016;74(1):50-7. Persian

13. Abdullah M, Al-Asmari S. Anemia Types Prediction Based on Data Mining Classification Algorithms. 1th ed. London CRC Press; 2016.

14. Dithy MD, KrishnaPriya V. Predicting anemia in pregnant women by using gausnominal classification algorithm. International Journal of Pure and Applied Mathematics 2018; 118(20): 3343-9.

15. Gnanapriya S, Suganya R, Devi GS, Kumar MS. Data mining concepts and techniques. Data Mining and Knowledge Engineering 2010; 2(9): 256-63.

16. Mentzer WC. Differentiation of iron deficiency from thalassaemia trait. Lancet 1973;1(7808):882.

17. England JM, Fraser PM. Differentiation of iron deficiency from thalassaemia trait by routine blood-count. Lancet 1973;1(7801):449-52.

18. Srivastava PC. Differentiation of thalassaemia minor from iron deficiency. Lancet 1973;2(7821):154-5.

19. Ehsani MA, Shahgholi E, Rahiminejad MS, Seighali F, Rashidi A. A new index for discrimination between iron deficiency anemia and beta-thalassemia minor: results in 284 patients. Pak J Biol Sci 2009;12(5):473-5.

20. Jayabose S, Giavanelli J, Levendoglu-Tugal O, Sandoval C, Özkaynak F, Visintainer P. Differentiating iron deficiency anemia from Thalassemia minor by using an RDW-based index. J Pediatr Hematol 1999;21:314.

21. Sirdah M, Tarazi I, Al Najjar E, Al Haddad R. Evaluation of the diagnostic reliability of different RBC indices and formulas in the differentiation of the beta-thalassaemia minor from iron deficiency in Palestinian population. Int J Lab Hematol 2008;30(4):324-330.

22. Matos JF, Dusse LM, Borges KB, de Castro RL, Coura-Vital W, Carvalho M. A new index to discriminate between iron deficiency anemia and thalassemia trait. Rev Bras Hematol Hemoter 2016;38(3):214-9.

23. Ghafouri M, Mostaan Sefat L, Sharifi S, Hosseini Gohari L, Attarchi Z. Comparison of cell counter

[Downloaded from jhbmi.com on 2026-11-15]

indices in differentiation of beta thalassemia minor from iron deficiency anemia. Sci J Iran Blood Transfus Organ 2006;2(7): 385-9. Persian.
24. Keikhaei B, Rahim F, Zandian K, Pedram M. Comparison of different indices for better differential diagnosis of iron deficiency anemia from beta thalassemia trait. Sci J Iran Blood Transfus Organ 2007;4(2): 95-104. Persian
25. Green R, King R. A New Red Cell Discriminant Incorporating Volume Dispersion for Differentiating Iron Deficiency Anemia from Thalassemia Minor. Blood Cells 1989;15(3):481-95.
26. Shine I, Lal S. A strategy to detect beta-thalassemia minor. Lancet 1977; 1: 692−4.
27. Telmissanil OA, Khalil S, George TR. Mean density of hemoglobin per liter of blood: a new hematologic parameter with an inherent discriminant function. Lab Haematol 1999;149-52.
28. Klee GG, Fairbanks VF, Pierre RV, O'Sullivan MB. Routine erythrocyte measurements in diagnosis of irondeficiency anemia and thalassemia minor. Am J Clin Pathol 1976;66(5): 870-7.

[Downloaded from jhbmi.com on 2026-11-15]

Original Article

Journal of Health and Biomedical Informatics
Medical Informatics Research Center
2019; 5(4): 435-446

Using Data Mining Models for Differential Diagnosis of Iron Deficiency Anemia and β-thalassemia Minor

Noferesti Samira[1], Shemshadi Nejad Narges*[2], Heydari Fatemeh[3]

• **Received:** 18 Jul, 2018 • **Accepted:** 11 Nov, 2018

Introduction: One of the most common types of anemia is Iron deficiency anemia that its main differential diagnosis is β-thalassemia minor. The rapid and accurate screening of β-thalassemia minor has particular importance for pre-marriage medical counseling and the prevention of the birth of neonates with β-thalassemia major and differentiating it from iron deficiency anemia to avoid unnecessary prescription of iron. The aim of this study was to apply data mining techniques to differentiate iron deficiency anemia from β-thalassemia minor based on CBC test in order to increase the diagnostic speed and to reduce diagnostic costs.

Method: The present study was a retrospective study and was performed on 1000 patients referred to Dr. Heidari laboratory of Zahedan city. To conduct research, CRISP-DM standard methodology and support vector machine data mining algorithms, naive *Bayes*, Bagging, Adaboots and decision tree have been used. WEKA software was used to analyze the data.

Results: The results of the evaluations show that Bagging, Decision tree, Adaboots, support vector machine, and naive *Bayes* algorithms had respectively 95.73%, 95.5%, 94.6%, 80.2% and 76.6% accuracy in differentiating iron deficiency anemia from β-thalassemia minor.

Conclusion: In this study, an automatic method based on data mining techniques for differentiation of iron deficiency anemia from β-thalassemia minor is presented. The results of the evaluations show that Bagging algorithm has higher accuracy compared to other data mining algorithms and differential indices. Also, with the help of the decision tree, rules have been extracted that can be used by the physician in timely diagnosis of the two diseases.

Keywords: Iron deficiency anemia, β-thalassemia minor, Differential diagnosis, Data mining, Bagging ensemble learning algorithm

• **Citation:** Noferesti S, Shemshadi Nejad N, Heydari F. Using Data Mining Models for Differential Diagnosis of Iron Deficiency Anemia and β-thalassemia Minor. Journal of Health and Biomedical Informatics 2019; 5(4): 435-446.

1. PhD in Computer Engineering, Assistant Professor, Information Technology Dept., Faculty of Electrical and Computer Engineering, University of Sistan and Baluchestan, Zahedan, Iran
2. MSc in Software Engineering, Zahedan, Iran
3. Pathologist (Clinical and Anatomical Disease), Assistant Professor, Faculty of Medicine, Zahedan University of Medical Sciences, Iran

***Correspondence:** Sistan and Baloochestan Province, Zahedan City, University Street, University 35
• **Tel:** 09155406975 • **Email:** shemshadi.narges@gmail.com

[Downloaded from jhbmi.com on 2026-11-15]