مجله دانشکده پیراپزشکی دانشگاه علوم پزشکی تهران (پیاورد سلامت)
دوره ۱۱ شماره ۵ آذر و دی ۱۳۹۶، ۵۴۸-۵۴۱

مقاله پژوهشی

مقایسه ی الگوریتم های مختلف طبقه بندی داده ها برای تعیین نوع زردی در نوزادان

دکتر رضا صفدری[۱]، دکتر ملیحه کدیور[۲]، پریناز تبری[۳]، دکتر هالا شاوکی اون[۴]

دریافت مقاله : بهمن ۱۳۹۵
پذیرش مقاله : خرداد ۱۳۹۶

*نویسنده مسئول :
پریناز تبری؛
دانشکده پیراپزشکی دانشگاه علوم پزشکی تهران
Email :
p-tabari@razi.tums.ac.ir

چکیده

زمینه و هدف: زردی در نوزادان مبحثی است که برای متخصصان در سراسر دنیا بسیار مهم تلقی می شود. زیرا این بیماری یکی از عمده ترین وضعیت هایی است که به توجه بالینی نیازمند است. هدف از انجام این پژوهش استفاده از تکنیک های طبقه بندی داده ها برای پیش بینی به موقع نوع زردی نوزادان و در نتیجه پیشگیری از آسیب های جبران ناپذیر به سلامت نوزادان بوده است.

روش بررسی: این پژوهش از نوع توصیفی بوده و با استفاده از مجموعه داده های جمع آوری شده درباره ی زردی نوزادان در شهر قاهره مصر انجام شده است. در این بررسی پس از پیش پردازش داده ها، تکنیک های داده کاوی از قبیل درخت تصمیم، Naïve Bayes و kNN(k نزدیکترین همسایه) در نرم افزار Orange بررسی، مقایسه و تحلیل شده است.

یافته ها: یافته های حاصل از پژوهش نشان داد که الگوریتم درخت تصمیم با دقت ۹۴ درصد، الگوریتم Naïve Bayes با دقت ۹۱ درصد و الگوریتم نزدیک ترین همسایه با دقت ۸۹ درصد نوع زردی در نوزادان را طبقه بندی می کنند. بنابراین بهترین الگوریتم از لحاظ دقت عملکرد در بین روش های طبقه بندی کننده، الگوریتم درخت تصمیم شناخته شد.

نتیجه گیری: استفاده از الگوریتم های طبقه بندی در ساخت سیستم های تصمیم یار می تواند به پزشکان در تصمیم گیری درباره نوع بیماری ها کمک کند و متخصصان می توانند برای رسیدگی به بیماران متناسب با نوع بیماری اقدام کنند که طی آن مخاطرات احتمالی در اثر عدم شناسایی به موقع یا صحیح بیماری کاهش خواهد یافت.

واژهای کلیدی: نوزادان، داده کاوی، طبقه بندی، زردی، هایپربیلی روبینمی

[۱] استاد گروه مدیریت اطلاعات سلامت، دانشکده پیراپزشکی، دانشگاه علوم پزشکی تهران، تهران، ایران
[۲] استاد گروه نوزادان، دانشکده پزشکی، بیمارستان مرکز طبی کودکان، دانشگاه علوم پزشکی تهران، تهران، ایران
[۳] کارشناس ارشد انفورماتیک پزشکی، دانشکده پیراپزشکی، دانشگاه علوم پزشکی تهران، تهران، ایران
[۴] دکترای علم آمار و کامپیوتر، گروه تحقیقات فضایی و خورشیدی، موسسه تحقیقات ملی ژئوفیزیک و ستاره شناسی، هلوان، مصر

مقدمه

زردی(Jaundice) یا هایپر بیلی روبینمی (Hyperbilirubinemia)، وجود سطح بالایی از بیلی روبین در خون است(۲و۱) و معمول ترین ظهور بالینی در نوزادان می باشد. در صورتی که این ضایعه ادامه یابد، دیر درمان شود و یا به موقع تشخیص داده نشود، خطراتی نوزاد را تهدید می کند که ممکن است صدمات جبران ناپذیری به وی وارد کند(۳). اکثر نوزادان(حدود ۶۰ درصد نوزادان ترم و ۸۰ درصد نوزادان پریترم) در هفته ی اول زندگی به زردی مبتلا می شوند(۴). گرچه بیشتر زردی ها در نوزادان بی خطر هستند ولی به علت تاثیر سمی بالقوه ی بیلی روبین، نوزادان باید برای شناسایی موارد هایپربیلی روبینمی شدید و انسفالوپاتی بیلی روبین حاد(Acute bilirubin encephalopathy) یا کرنیکتروس(Kernicterus)، مانیتور شوند(۵).

کرنیکتروس یا انسفالوپاتی بیلی روبین، آسیب مغزی غیر قابل بازگشتی است که به وسیله ی جمع شدن بیلی روبین در ناحیه هایی از مغز ایجاد می شود. نوزادان پریترم و بیمار با وضعیت هایپربیلیروبین(بیلیروبین بالا) مستعد قرار گرفتن در این وضعیت هستند(۶).

گرچه زردی نوزادان به وسیله ی نور درمانی، تزریق آلبومین، یا انتقال خون قابل درمان است ولی آسیبی که توسط بیلی روبین به مغز وارد می شود، همیشه قابل جبران نیست و می تواند منجر به فلج مغزی، ناشنوایی یا کاهش شنوایی، خیرگی چشم به سمت بالا و دیسپلازی(Dysplasia) مینای دندان های اصلی شود(۷). همان طور که ذکر شد، سیستم شنوایی به بیلی روبین واکنش نشان می دهد و مقالات متعددی رابطه ی بین هایپر بیلی روبینمی شدید و صدمه به سیستم شنوایی را نشان داده اند و حساسیت سیستم شنوایی به بیلی روبین ثابت شده است. این صدمات و تاثیرات می تواند غیر طبیعی بودن شنوایی و پردازش بیان تا ناشنوایی کامل را شامل شود. به دلیل ریسک شناخته شده ی کاهش شنوایی پس از هایپربیلیروبینمی، این کودکان باید سالها شنوایی سنجی شوند تا تشخیص زود هنگام و مداخلات تهاجمی برای پیشگیری از کاهش شنوایی تسهیل شود(۱). نوزادان مستعد مواجه با این ضایعه در صورتی که به موقع مانیتور و درمان شوند آثار پر خطری را شاهد نخواهیم بود(۸). در این میان استفاده از تکنیک های داده کاوی برای پیش بینی بیماری ها، می تواند نتایج مناسب تری را نسبت به متدهای سنتی ارایه کند(۳). بنابراین می توان با استفاده از این روش ها، زردی نوزادان و نوع آن را نیز به موقع تشخیص داد و از صدمات احتمالی پیشگیری کرد.

روش بررسی

این پژوهش توصیفی با استفاده از مجموعه داده های جمع آوری شده از ۳۲۵ نوزاد دارای زردی و بستری در بخش مراقبت های ویژه نوزادان در بیمارستانی در قاهره مصر در بازه ماه های ژوئن تا دسامبر سال ۲۰۰۷ انجام شده است. این دیتاست مرتبط با زردی نوزادان، مجموعه داده ای جامع همراه با ویژگی های تعیین کننده در زمینه تحقیق، تشخیص و بررسی زردی نوزادان بوده و همواره مورد توجه محققان در این زمینه گردیده است و لازم به ذکر است که در پژوهش های معتبری، داده های جمع آوری شده در این پایگاه داده مورد استفاده قرار گرفته است(۱۱-۹).

در این مجموعه داده، ۱۷ متغیر مربوط به نوزدان جمع آوری شده است که این متغیرها عبارتند از: "جنس، سن(به روز)، سن داخل رحمی، وزن، سن نوزاد در روز بروز زردی، تعداد روزهای بستری، حداکثر میزان بیلی روبین توتال، سن نوزاد در زمان ثبت حداکثر میزان بیلی روبین توتال، میزان بیلی روبین توتال در روز بروز زردی، میزان بیلی روبین دایرکت در روز بروز زردی، میزان بیلی روبین توتال ۲۴ ساعت بعد از بروز اولیه زردی، میزان بیلی روبین دایرکت ۲۴ ساعت بعد از بروز اولیه زردی، میزان بیلی روبین توتال دو روز بعد از بروز اولیه زردی، میزان بیلی روبین دایرکت دو روز بعد از بروز اولیه زردی، میزان بیلی روبین توتال قبل از پذیرش، میزان بیلی روبین دایرکت قبل از پذیرش و الگوی زردی". الگوی زردی به عنوان متغیر خروجی یا مقصد شامل سه کلاس زیر است: زردی غیر مستقیم، زردی غیر مستقیم سپس تبدیل به زردی مستقیم و زردی مستقیم.

زمانی که گلبول های قرمز خون تجزیه می شوند، هموگلوبین آزاد می کنند و ملکول های Heme(از هموگلوبین) تبدیل به بیلی روبین می شود. بیلی روبین(غیر مستقیم یا غیر کونژوگه)، به آلبومین متصل شده و به کبد منتقل می شود و بیلی روبین کونژوگه یا مستقیم در صفرا ریخته می شود. در صورتی که میزان تجمع بیلی روبین غیر کونژوگه(غیر مستقیم) در خون بسیار زیاد باشد ممکن است زردی فیزیولوژیک یا پاتولوژیک در نوزاد رخ دهد(۲). پس تشخیص زردی مستقیم و غیر مستقیم برای پیشگیری از مخاطرات احتمالی در نوزادان بسیار حایز اهمیت است.

در این پژوهش در ابتدا با استفاده از نرم افزار orange بر روی داده ها پیش پردازش انجام شد که طی آن در آغاز، فیلدهایی که دارای مقدار تهی بودند با مقدار میانگین پر شدند. سپس با استفاده از روش انتخاب ویژگی یا feature selection، بر اساس

information gain تعداد ۱۰ عنصر اطلاعاتی مهم و مرتبط از بین عناصر اطلاعاتی پیش بینی کننده انتخاب شد و لازم به ذکر است این عناصر اطلاعاتی از نظر متخصصان داخلی نیز مهم تلقی می شود. نرم افزار با استفاده از این نسبت، اهمیت و میزان مرتبط بودن ویژگی ها با برچسب کلاس(متغیر هدف، که در این مجموعه داده الگوی زردی می باشد) تعیین می کند. به عبارت دیگر با استفاده از این متد، میزان اطلاعات قابل حصول با دانستن مقدار ویژگی(برچسب کلاس) محاسبه شده و در نهایت بهترین عناصر اطلاعاتی تعیین کننده، برای استفاده مورد نظر قرار گرفتند؛ به عبارت دیگر متغیری با اطلاعات بیشتری که برای طبقه بندی مفیدتر است انتخاب می شود.

Information gain یکی از روش های محقق شدن هدف حذف متغیرهای نامرتبط یا مازاد بر احتیاج است و انتروپی(Entropy) به طور عمده در مقیاس تئوری اطلاعات مورد استفاده قرار می گیرد که میزان خالص بودن مجموعه ی دلخواهی از نمونه ها را توصیف می کند. انتروپی، پایه و اساس روش های مختلف درجه بندی و انتخاب خصوصیات در information gain است. این اندازه گیری به عنوان مقیاس غیر قابل پیش گویی سیستم در نظر گرفته می شود. انتروپی Y به این صورت تعریف می شود:

$$H(Y) = -\sum_{y \in Y} p(y) \log_2(p(y))$$

که در اینجا p(y) تابع چگالی احتمال حاشیه ای برای متغیر تصادفی Y است. در صورتی که مقادیر Y در دیتاست آموزش S بر اساس مقادیر متغیر دوم X تقسیم بندی شوند و انتروپی Y با توجه به تقسیمات به وجود آمده توسط X، کمتر از انتروپی Y قبل از تقسیم بندی باشد، بین متغیرهای X و Y وابستگی و ارتباط وجود دارد. بنابراین انتروپی Y بعد از مشاهده ی X به این صورت است:

$$H(Y|X) = -\sum_{x \in X} p(x) \sum_{y \in Y} p(y|x) \log_2(p(y|x))$$

با داشتن انتروپی، مقیاسی تعیین می شود که اطلاعات اضافی درباره ی Y که توسط X به دست آمده است را منعکس می کند که مقداری را نشان می دهد که توسط آن انتروپی Y کاهش می یابد:

$$Information\ Gain = H(Y) - H(Y|X) = H(X) - H(X|Y)$$

این نسبت نشان می دهد که اطلاعات به دست آمده درباره ی Y بعد از مشاهده ی X برابر با اطلاعات به دست آمده درباره ی X بعد از مشاهده ی Y است. نقطه ضعف این تکنیک این است که به نفع متغیرهایی که مقادیر بیشتری دارند سوگیری دارد حتی اگر آنها حاوی اطلاعات مفید کمتری باشند(۱۲).

عناصر اطلاعاتی انتخاب شده توسط سیستم شامل سن نوزاد، تعداد روزهای بستری، حداکثر میزان بیلی روبین توتال، بیلی روبین توتال در روز بروز زردی، بیلی روبین دایرکت در روز بروز زردی، بیلی روبین توتال ۲۴ ساعت بعد از بروز اولیه زردی، بیلی روبین دایرکت ۲۴ ساعت بعد از بروز اولیه زردی، بیلی روبین توتال دو روز بعد از بروز اولیه زردی، بیلی روبین دایرکت دو روز بعد از بروز اولیه زردی و بیلی روبین دایرکت قبل از ترخیص می باشد. عنصر اطلاعاتی الگوی زردی نیز به عنوان عنصر هدف و خروجی برای انجام عملیات طبقه بندی مورد استفاده قرار می گیرد.

پس از مرحله ی پیش پردازش، الگوریتم های مختلف طبقه بندی کننده از جمله درخت تصمیم گیری، Naïve Bayes و k نزدیک ترین همسایه بررسی گردید. در این بخش از نمونه گیری تصادفی و با استفاده از تخصیص ۷۰ درصد داده ها برای آموزش و ۳۰ درصد داده ها برای تست سیستم های طبقه بندی استفاده شد و تعداد دفعات تکرار آموزش- تست ۱۰ مرتبه تعیین شد.

یافته ها

یافته های حاصل از پژوهش نشان داد که دقت طبقه بندی با استفاده از درخت تصمیم نسبت به الگوریتم های دیگر بالاتر است. نتایج حاصل از بررسی در جدول ۱ خلاصه شده است:

جدول ۱: تحلیل یافته های حاصل از اعمال الگوریتم های مختلف

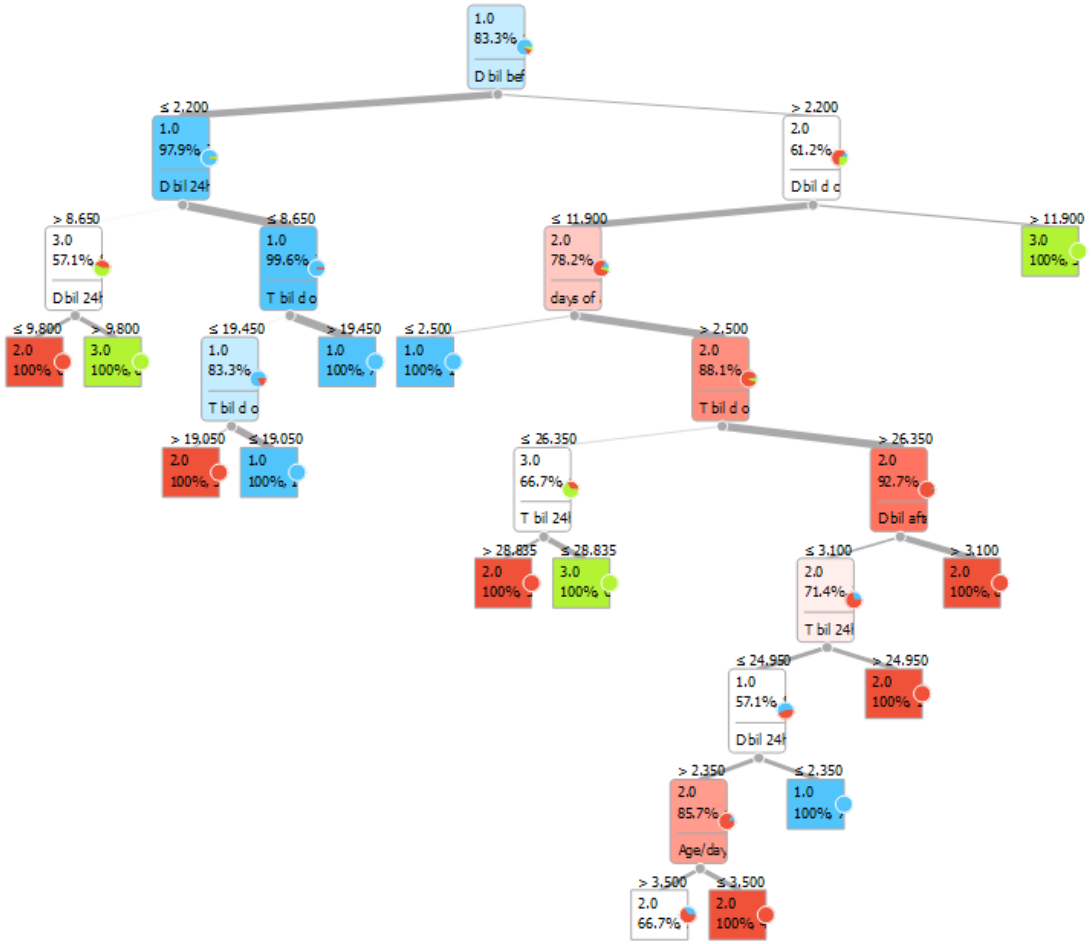
متد طبقه بندی	AUC	CA	F1	Precision	Recall
Decision tree	۰/۸۷۳	۰/۹۴۲	۰/۹۴۰	۰/۹۴۰	۰/۹۴۲
Naïve Bayes	۰/۸۷۹	۰/۸۸۹	۰/۸۹۷	۰/۹۱۴	۰/۸۸۹
kNN	۰/۷۶۷	۰/۹۰۱	۰/۸۹۲	۰/۸۹۶	۰/۹۰۱

AUC (Area Under the Curve) = سطح زیر منحنی راک/ CA(Classification Accuracy)= صحت طبقه بندی/F1 = میانگین هارمونیک وزن دار دقت و recall

Precision = نسبت مثبت صحیح(True Positive) در بین نمونه های طبقه بندی شده به عنوان مثبت.

Recall = نسبت مثبت صحیح در بین تمام نمونه های مثبت در داده ها.

در این قسمت درخت حاصل از الگوریتم درخت تصمیم نشان داده شده است:



شکل ۱: درخت تصمیم

در این بخش ماتریس ابهام مربوط به هریک از الگوریتم های طبقه بندی نشان داده شده است.

جدول ۲: ماتریس ابهام الگوریتم درخت تصمیم

		پیش بینی شده		
		کلاس ۱	کلاس ۲	کلاس ۳
واقعی	کلاس ۱	٪۹۶/۳	٪۹/۳	٪۲/۷
	کلاس ۲	٪۳	٪۷۹/۴	٪۸/۱
	کلاس ۳	٪۰/۷	٪۱۱/۲	٪۸۹/۲

جدول ۳: ماتریس ابهام الگوریتم Naïve Bayes

		پیش بینی شده		
		کلاس ۱	کلاس ۲	کلاس ۳
واقعی	کلاس ۱	٪۹۸/۶	٪۳۳/۹	٪۱۹/۲
	کلاس ۲	٪۱/۴	٪۵۴/۲	٪۲۱/۲
	کلاس ۳	٪۰	٪۱۱/۹	٪۵۶/۶

جدول ۴: ماتریس ابهام الگوریتم k نزدیکترین همسایه

		پیش بینی شده		
		کلاس ۱	کلاس ۲	کلاس ۳
واقعی	کلاس ۱	٪۹۲/۹	٪۱۲/۸	٪۴/۵
	کلاس ۲	٪۶/۲	٪۶۲/۸	٪۰
	کلاس ۳	٪۰/۸	٪۲۴/۵	٪۹۵/۵

در جداول ۲، ۳ و ۴ مقادیر متغیر هدف یا برچسب که در اینجا متغیر الگوی زردی است، با عناوین کلاس ۱(زردی غیر مستقیم)، کلاس ۲(زردی غیر مستقیم سپس تبدیل به زردی مستقیم) و کلاس ۳(زردی مستقیم) معرفی شده است. همان طور که در جدول ۲ مشاهده می شود، ماتریس ابهام الگوریتم درخت تصمیم به این صورت تعریف می شود که حدود ۹۶ درصد داده هایی که در کلاس ۱ پیش بینی شده اند، واقعاً متعلق به کلاس ۱ بودند(میزان مثبت صحیح یا True positive) و این نسبت برای کلاس ۲، حدود ۷۹ درصد و برای کلاس ۳ حدود ۸۹ درصد می باشد؛ که نتیجه می گیریم الگوریتم درخت تصمیم در تشخیص کلاس ۱ بهتر عمل کرده و بر خلاف آن برای تشخیص داده های کلاس ۲ ضعیف عمل می کند.

این رویه تشخیص برای الگوریتم Naive Bayes نیز صادق است. در الگوریتم k نزدیکترین همسایه، میزان پیش بینی صحیح کلاس ۳ بیشتر از دو کلاس دیگر است و الگوریتم در تشخیص صحیح داده های کلاس ۲ ضعیف عمل می کند(حدود ۶۲ درصد).

بحث

در مطالعه ای که در سال ۹۳ توسط باقری و همکاران انجام شد، با استفاده از الگوریتم های داده کاوی عوامل موثر بر پیش بینی وضعیت بدو تولد نوزادان بررسی گردید. در میان الگوریتم های مورد استفاده ی CART، QUEST، CHAID و C5.0، الگوریتم C5.0 با نرخ تشخیص ٪۹۴/۴۴ به عنوان بهترین الگوریتم شناخته شد(۱۳). در مطالعه ی دیگری که توسط فیروزی جهانتیغ و همکاران در سال ۹۴ انجام شد، با استفاده از تکنیک های داده کاوی به شناسایی، بررسی و رتبه بندی عوامل خطر وزن کم نوزادان در زمان تولد پرداخته شده است. درخت تصمیم به خوبی وزن کم نوزادان در زمان تولد را مشخص می کند و تکنیک جنگل تصادفی در تشخیص بیماری نقش مهمی را ایفا می کند(۱۴). در مطالعه ای که توسط Seidman و همکاران در سال ۱۹۹۹ انجام شده است، هدف پژوهش تعیین توانایی پیش بینی آینده نگر هایپربیلیروبینمی شدید در نوزادان ترم سالم بوده است. در این مطالعه، ۱۱۱۷ نوزاد ترم سالم مورد بررسی قرار گرفتند. با استفاده از آنالیز رگرسیون لوجستیک چند گانه، زردی نوزادان به خوبی پیش بینی شد($P<۰/۰۰۰۱$) که این کار به وسیله ی اندازه گیری بیلی روبین سرم روز اول و تغییر در سرم بیلی روبین بین روز اول و روز دوم انجام شده است. مدل پیش بینی زردی نوزادان بر اساس فاکتورهایی مثل گروه خونی، سابقه زردی در هم نیاها، جنس و وزن نوزاد، دیابت ملیتوس، فشار خون، مصرف سیگار و ... در مادر، دارای حساسیت ٪۸۱/۸، ویژگی ٪۸۲/۸، مثبت کاذب ٪۸۰/۲ و منفی کاذب ٪۱/۱ بوده است(۱۵). در سال ۲۰۰۶، Srinivasan و همکاران، سیستم تصمیم یار بالینی برای تشخیص زردی نوزادان ارایه دادند. که هدف اصلی این سیستم توسعه یک ماژول سیستمی است که از شروط تصمیمات بالینی استفاده می کند و برای ارایه مشورت

به پرسنل نیمه حرفه ای طراحی شده که احتیاج به راهنمایی برای تصمیم گیری دارند. در این مطالعه از یک سیستم مبتنی بر قاعده برای ایجاد درخت تصمیم استفاده شده است. در طراحی این سیستم از Microsoft .net framework استفاده شده که از ۳۹ مورد بیمار در ۴ ماه ۳۳ مورد به درستی تشخیص داده شدند و به آنها برنامه پیگیری ارایه شد. ایراد اصلی سیستم این است که قابلیت یادگیری به وسیله ی خود را ندارد(۱۶). مطالعه ی Own و Abraham در سال ۲۰۱۲ با استفاده از دیتاست زردی نوزادان که در مصر جمع آوری شده، انجام شده است. در این پژوهش متد جدیدی از طبقه بندی وزن دار برای کلاسه بندی زردی نوزادان با استفاده از ۱۶ متغیر مورد استفاده قرار گرفته است که در نهایت با استفاده از الگوریتم های کاهش ویژگی، با پنج متغیر قواعد سیستم تولید شده است. روش طبقه بندی مورد استفاده در این پژوهش نسبت به الگوریتم های SVM وزن دار و درخت تصمیم مناسب تر عمل کرده است(۱۰). در مطالعه دیگری که در سال ۲۰۱۲ توسط Ferreira و همکاران انجام شد، از تکنیک های داده کاوی برای بهبود تشخیص در زردی نوزادان استفاده شده است. طبق نظر نویسندگان، در فیلدهای مختلف پزشکی، داده کاوی برای بهبود نتایج به دست آمده از متدولوژی های دیگر مورد استفاده قرار می گیرد. متدولوژی این مطالعه استفاده از فاز های مختلف Cross Industry Standard Process برای مدل داده کاوی می باشد. این تحقیق از ماه فوریه تا مارس ۲۰۱۱ بر روی ۲۲۷ نوزاد سالم با سن بارداری ۳۵ هفته یا بیشتر انجام شده است. در این مطالعه بیش از ۷۰ متغیر جمع آوری و آنالیز شده است. زیر مجموعه های خصوصیتی مختلفی برای مدل های تقسیم بندی آموزش و آزمایش با استفاده از الگوریتم هایی که در نرم افزار داده کاوی Weka قرار دارد، مورد استفاده قرار گرفته است، مثل درخت های تصمیم(J۴۸) و شبکه های عصبی(پرسپترون چند لایه). در ۲۴ ساعت بعد از تولد، دقت پیش بینی هایپربیلی روبینمی ۸۹٪ می باشد. بهترین نتایج به وسیله ی الگوریتم هایی مثل بیزین، پرسپترون چند لایه و لوجستیک ساده به دست آمده است. نتایج این مطالعه نشان داده است که رویکردهای جدیدی مثل داده کاوی ممکن است از تصمیمات بالینی برای بهبود تشخیص زردی نوزادان پشتیبانی کند(۳). بسیاری از متغیرهای مهم در مطالعات ذکر شده مرتبط با زردی، با عناصر اطلاعاتی مطالعه ی حاضر همخوانی دارند که این مساله اهمیت این متغیرها در زمینه تشخیص زردی در نوزادان را نشان می دهد. علاوه بر این استفاده از تکنیک های مناسب طبقه بندی داده ها و انواع مختلف روش های داده کاوی در مطالعات پیشین، نتایج مناسبی را برای طبقه بندی نوع زردی در نوزادان و همچنین پیش بینی به موقع دیگر وضعیت ها به همراه داشته است که پژوهش حاضر نیز با استفاده از تکنیک های مناسبی به بررسی این موضوع پرداخته است.

نتیجه گیری

زردی در نوزادان همواره مساله ی مهمی در حوزه ی بالین محسوب می شود چرا که اکثر نوزادان به این ضایعه مبتلا می شوند. در صورتی که میزان بیلی روبین به حدی زیاد شود، می تواند منجر به آنسفالوپاتی بیلی روبین یا کرنیکتروس شود و به مغز صدمات جبران ناپذیری وارد کند. بنابراین شناسایی روشی برای پیش بینی به موقع زردی نوزادان و نوع آن می تواند از این مورد پیشگیری کند. روش های طبقه بندی مختلفی می توانند نوع زردی در نوزادان را پیش بینی کنند که در این پژوهش و با استفاده های داده های موجود در مصر، درخت تصمیم توانست بهترین عملکرد را در طبقه بندی ارایه دهد. همچنین عناصر اطلاعاتی مثل میزان بیلی روبین در روز بروز و تا دو روز بعد از بروز، متغیرهای مهمی محسوب می شوند. پیشنهاد می شود سیستم تصمیم یار بالینی از طریق نتایج این بررسی طراحی شده تا پزشکان بتوانند با استفاده از محیطی مناسب به بررسی و پیش بینی نوع زردی در نوزادان بپردازند و در نتیجه از بروز صدمات جبران ناپذیر در نوزادان پیشگیری نمایند.

تشکر و قدردانی

این مقاله حاصل بخشی از پایان نامه ی کارشناسی ارشد انفورماتیک پزشکی با عنوان "طراحی و ایجاد سیستم تصمیم یار بالینی پیش بینی احتمال بروز زردی پاتولوژیک در نوزادان" و شماره ی ۳۴/الف/۲۸۰/۳ در دانشکده پیراپزشکی دانشگاه علوم پزشکی تهران می باشد، لذا از تمام استادان گروه مدیریت اطلاعات سلامت این دانشکده که ما را یاری کردند صمیمانه تشکر می شود.

منابع

1. Mesic I, Milas V, Medimurec M & Rimar Z. Unconjugated pathological jaundice in newborns. Collegium Antropologicum 2014; 38(1): 173-8.

2. Mohammed AE, Behiry EG, El-Sadek AE, Abdulghany WE, Mahmoud DM & Elkholy AA. Case-controlled study on indirect hyperbilirubinemia in exclusively breast fed neonates and mutations of the bilirubin uridine diphosphate-glucuronyl transferase gene 1a1. Annals of Medicine and Surgery 2017; 13(1): 6-12.

3. Ferreira D, Oliveira A & Freitas A. Applying data mining techniques to improve diagnosis in neonatal jaundice. BMC Medical Informatics and Decision Making 2012; 12(1): 143.

4. National Collaborating Centre for Women's and Children's Health (UK). Neonatal jaundice. Available at: https://www.ncbi.nlm.nih.gov/pubmed/22132434. 2010.

5. American Academy of Pediatrics Clinical Practice Guideline Subcommittee on Hyperbilirubinemia. Management of hyperbilirubinemia in the newborn infant 35 or more weeks of gestation. Pediatrics 2004; 114(1): 297-316.

6. Brito MA, Pereira P, Barroso C, Aronica E & Brites D. New autopsy findings in different brain regions of a preterm neonate with kernicterus: Neurovascular alterations and up-regulation of efflux transporters. Pediatric Neurology 2013; 49(6): 431-8.

7. Wright MO & Robicsek A. Clinical decision support systems and infection prevention: To know is not enough. American Journal of Infection Control 2015; 43(6): 554-8.

8. Bhutani VK & Johnson L. Kernicterus in the 21st century: Frequently asked questions. Journal of Perinatology 2009; 29(1): 20-4.

9. Banu PKN, Inbarani HH, Azar AT, Own HS & Hassanien AE. Rough set based feature selection for egyptian neonatal jaundice. Advanced machine learning technologies and applications, communications in computer and information science, Egypt: Second International Conference, AMLTA, 2014.

10. Own HS & Abraham A. A new weighted rough set framework based classification for Egyptian neonatal jaundice. Applied Soft Computing 2012; 12(3): 999-1005.

11. Azar AT, Inbarani HH, Kumar SU & Own HS. Hybrid system based on bijective soft and neural network for Egyptian neonatal jaundice diagnosis. Int J Intell Eng Inform 2016; 4(1): 71-90.

12. Novakovic J. Using information gain attribute evaluation to classify sonar targets. Available at: http://2009.telfor.rs/files/radovi/10_60.pdf. 2009.

13. Bagheri F, Alizadeh Majd H, Mehrbakhsh Z & Ziaratban M. Use of data mining algorithms in assessing the affecting factors on predicting the health status of newborns. Hakim Jorjani Journal 2014; 2(2): 59-68[Article in Persian].

14. Firouzi Jahantigh F, Nazarnejad R & Firouzijahantigh M. Investigating the risk factors for low birth weight using data mining: A case study of Imam Ali hospital, Zahedan, Iran. Journal of Mazandaran University of Medical Sciences 2016; 133(1): 171-88[Article in Persian].

15. Seidman DS, Ergaz Z, Paz I, Laor A, Revel-Vilk S, Stevenson DK, et al. Predicting the risk of jaundice in full-term healthy newborns: A prospective population-based study. Journal of Perinatology 1999; 19(1): 564-7.

16. Srinivasan S, Mital DP & Haque S. A point of care clinical decision support system for the diagnosis of neonatal jaundice by medical field personnel. Journal of Applied Sciences 2006; 6(5): 1003-8.

Comparison of Data Classification Algorithms to Determine the Type of Neonatal Jaundice

Safdari Reza[1] (Ph.D.) - Kadivar Maliheh[2] (M.D.) - Tabari Parinaz[3] (M.S.) - Shawky Own Hala[4] (Ph.D.)

1 Professor, Health Information Management Department, School of Allied Medical Sciences, Tehran University of Medical Sciences, Tehran, Iran

2 Professor, Neonatology Department, School of Medicine, Children's Medical Center Hospital, Tehran University of Medical Sciences, Tehran, Iran

3 Master of Science in Medical Informatics, School of Allied Medical Sciences, Tehran University of Medical Sciences, Tehran, Iran

4 Ph.D. in Statistics and Computer Science, Solar and Space Research Department, National Research Institute of Astronomy and Geophysics, Helwan, Egypt

Abstract

Received: Jan 2017

Accepted: May 2017

Background and Aim: Neonatal jaundice is a matter that is very important for clinicians all over the world because this disease is one of the most common cases that requires clinical care. The aim of this study is to use data classification algorithms to predict the type of jaundice in neonates, and therefore, to prevent irreparable damages in future.

Materials and Methods: This is a descriptive study and is done with the use of neonatal jaundice dataset that has been collected in Cairo, Egypt. In this study, after preprocessing the data, classification algorithms such as decision tree, Naïve Bayes, and kNN (k-Nearest Neighbors) were used, compared and analyzed in Orange application.

Results: Based on the findings, decision tree with precision of 94%, Naïve Bayes with precision of 91%, and kNN with precision of 89% can classify the types of neonatal jaundice. So, among these types, the most precise classification algorithm is decision tree.

Conclusion: Classification algorithms can be used in clinical decision support systems to help physicians make decisions about the types of special diseases; therefore, physicians can look after patients appropriately. So the probable risks for patients can be decreased.

Keywords: Neonatal, Data Mining, Classification, Jaundice, Hyperbilirubinemia

*** Corresponding Author:**
Tabari P;
Email:
p-tabari@razi.tums.ac.ir