مقاله پژوهشی

تاریخ دریافت: ۸۸/۵/۳
تاریخ پذیرش: ۸۸/۸/۲۸

کاربرد درخت رده بندی برای پیشگویی رتینوپاتی دیابتیک و مقایسه با مدل تابع تشخیص در بیماران دیابتی نوع ۲

دکتر سید محسن حسینی[۱]، دکتر مهدی تذهیبی[۱]، دکتر مسعود امینی[۲]، اصغر زارع[۳]، دکتر حسن جهانی هاشمی[۴]

خلاصه

مقدمه: بیماری دیابت یکی از بیماری‌های شایع در جهان به شمار می‌رود و رتینوپاتی دیابتی به عنوان یک بیماری چشمی، این گروه از بیماران را به شدت درگیر می‌نماید و یکی از عوامل کوری آنان در سنین بالا به شمار می‌رود. با توجه به شیوع بالای دیابت نوع II در جامعه و همچنین خطر بالای رتینوپاتی در این گروه از بیماران، در این مطالعه سعی گردید تا با توجه به مدل رده بندی درختی CART (Classification and Regression Tree) یک الگوی پیش‌گویی و رده‌بندی برای این بیماران معرفی شود.

روش‌ها: این بررسی با توجه به اطلاعات ۳۷۳۴ بیمار دیابتی نوع II انجام شد. بیماران به صورت تصادفی به دو گروه نمونه‌ی آزمون و یادگیری تقسیم شدند. با توجه به نمونه‌ی یادگیری، مدل رده بندی درخت CART معرفی و دقت آن با توجه به نمونه‌ی آزمون مشخص گردید. نتایج به دست آمده از الگوی درختی CART با نتایج به دست آمده از مدل تابع تشخیص برای کل بیماران مقایسه شد.

یافته‌ها: در این مطالعه، مدل رده بندی درختی با دقت ۰/۶۷ با توجه به دوره بیماری، فشار خون سیستولیک، سن، جنسیت، تری‌گلیسیرید، کلسترول و قند خون ناشتا به دست آمد. این الگوی رده بندی از حساسیت ۰/۷۱ و ویژگی ۰/۶۲ برخوردار بو. در این بررسی میزان دقت رده بندی درختی معرفی شده با میزان دقت الگوی رده بندی بر اساس تابع تشخیص تقریبا یکسان بود.

نتیجه‌گیری: در این بررسی مشخص گردید که دوره‌ی بیماری دیابت یکی از مهمترین عوامل تعیین کننده‌ی رتینوپاتی می‌باشد؛ به طوری که خطر رتینوپاتی در بیماران دیابت نوع II که از بیماری دیابت آن‌ها بیشتر از ۷/۵ سال می‌گذرد نسبت به سایر بیماران بیشتر است.

واژگان کلیدی: درخت رده بندی، تحلیل ممیزی (تابع تشخیص)، رتینوپاتی، دیابت نوع II.

[۱] استادیار آمار زیستی، گروه آمار و اپیدمیولوژی، دانشکده‌ی بهداشت، دانشگاه علوم پزشکی اصفهان، اصفهان، ایران.
[۲] استاد، مرکز تحقیقات غدد و متابولیسم، دانشگاه علوم پزشکی اصفهان، اصفهان، ایران.
[۳] کارشناس ارشد،گروه آمار و اپیدمیولوژی، دانشکده‌ی بهداشت، دانشگاه علوم پزشکی اصفهان، اصفهان، ایران.
[۴] دانشیار آمار زیستی، گروه پزشکی اجتماعی، دانشکده‌ی پزشکی، دانشگاه علوم پزشکی قزوین، قزوین، ایران.
نویسنده‌ی مسؤول: دکتر سید محسن حسینی، استادیار آمار زیستی، گروه آمار و اپیدمیولوژی، دانشکده‌ی بهداشت، دانشگاه علوم پزشکی اصفهان، اصفهان، ایران.
E-mail: hosseini@hlth.mui.ac.ir

Using Classification Tree for prediction of Diabetic Retinopathy on Type II Diabetes

Sayed Mohsen Hosseini MSc PhD*, Mehdi Tazhibi MSc PhD*, Massoud Amini MD**, Asghar Zaree MSc***, Hassan Jahani Hashemi MD****

Abstract

Background: Diabetes disease is a common disease in the world and diabetic retinopathy that is an eye disease, involve these patients so hardly that leads patients to blindness in elderly. Because of high incidence of type II diabetes in the society and the danger of retinopathy in this group of patients, in this study we attempted to introduce a classification and predictive model according to Classification and Regression Tree (CART) model for this disease.

Methods: This study was performed according to the information of about 3734 patients with type II diabetes, consulted to Isfahan Metabolic and Endocrine Research Center from 1991 to 2006. According to the CART model, a classification pattern was introduced for predicting of retinopathy in these patients.

Findings: In this study, classification tree model (CART) obtained with accuracy of 67 percents according to duration of disease, age, sex, systolic blood pressure, triglyceride, cholosterol and fasting blood sugar. This classification model had the sensitivity of 71% and specifity of 62%.

Conclusion: By this study founds that the duration of diabetes is one of the most important element of retinopathy, in such a manner that the danger of retinopathy in type II diabetes patients with more than 7.5 years is more than other patients.

Keywords: Diabetic retinopathy, Type II diabetes, Classification tree.

* Assistant Professor, Department of Biostatistics and Epidemiology, School of Health, Isfahan University of Medical Sciences, Isfahan, Iran.
** Professor, Endocrine and Metabolism Research Center, Isfahan University of Medical Sciences, Isfahan, Iran.
*** Resident MSc in Biostatistics, Department of Biostatistics and Epidemiology, School of Health, Isfahan University of Medical Sciences, Isfahan, Iran.
**** Associate Professor, Department of Community Medicine, School of Medicine, Isfahan University of Medical Sciences, Isfahan, Iran.
Corresponding Author: Sayed Mohsen Hosseini MSc, PhD, E-mail: hosseini@hlth.mui.ac.ir

مقدمه

دیابت یکی از بیماری‌های شایع و مزمن در جهان به شمار می‌رود (۱) و رتینوپاتی به عنوان یکی از شایع‌ترین عوارض میکرووواسکولر این گروه از بیماران را به شدت گرفتار می‌نماید (۲)؛ به طوری که خطر کوری در دیابتی‌ها ۲۵ برابر نسبت به سایر بیماران می‌باشد (۳). رتینوپاتی دیابتی یکی از مهمترین عواملی است که باعث ایجاد اختلالات بینایی یا کوری در سنین کار و فعالیت در جوامع پیشرفته می‌شود (۵-۴)؛ به طوری که این بیماری یکی از سه عامل مهم کوری در آمریکا به شمار می‌رود و منجر به کوری در سنین ۲۰ تا ۷۴ سالگی می‌گردد (۷-۶). عوامل مختلفی در بروز رتینوپاتی نقش دارند که می‌توان از طول دوره‌ی بیماری دیابت، سن، وجود پروتئین در ادرار، فشار خون، حاملگی، نوع کنترل قند خون و چربی‌های خون نام برد (۸،۹). با توجه به شیوع بالای دیابت نوع II در جامعه و همچنین خطر بالای رتینوپاتی دیابتی در بیماران مبتلا، در این بررسی سعی گردید تا با توجه به عوامل مؤثر در رتینوپاتی یک الگوی رده بندی درختی برای این بیماران معرفی گردد.

درخت رده بندی CART (Classification and Regression Tree) در زمینه‌ی تشخیص و پیش‌گویی پزشکی به طور گسترده‌تری در مقایسه با انواع دیگر درخت رده بندی مورد استفاده قرار می‌گیرد (۱۲-۱۰). روش‌های معمول آماری به سه دلیل عمده نمی‌توانند در این زمینه مورد استفاده قرار گیرند؛ اول این که به طور معمول، در اطلاعات پزشکی با مقادیر بسیاری از متغیرهای کمکی سر و کار داریم که این حجم زیاد متغیرها استفاده از روش‌های معمول را با مشکل و پیچیدگی‌هایی روبه‌رو می‌سازد؛ به عنوان نمونه اثرات متقابل در این داده‌ها به قدری پیچیده است که نمی‌توان با روش‌های معمول آماری آن‌ها را بررسی کرد. دوم این که روش‌های معمول آماری نیازمند وجود بعضی از فرضیات نظیر نرمال بودن توزیع و واریانس‌های برابر می‌باشند که در داده‌های پزشکی اغلب چنین فرضیاتی برقرار نخواهد بود. درخت تصمیم با توجه به استفاده از شیوه‌های ناپارامتری نیازی به داشتن توزیع خاصی برای مشاهدات ندارد (۱۵-۱۳، ۱۰). سوم این که نتایج به دست آمده از روش‌های معمول آماری به سادگی قابل استفاده نیست. درخت تصمیم یکی از شیوه‌های مناسب رده‌بندی می‌باشد که با وجود مشکلات ذکر شده می‌تواند به خوبی عمل نماید (۱۶). به علاوه درخت تصمیم نسبت به داده‌های پرت حساس نمی‌باشد و برای اطلاعات با حجم زیاد بسیار مناسب است (۱۸-۱۷، ۱۵-۱۳).

روش‌ها

تحقیقات دیابت اصفهان مراجعه کرده بودند، انجام گرفت. از بین این بیماران، کسانی که وضعیت رتینوپاتی مشخص شده داشتند، برای مطالعه انتخاب شدند؛ تعداد این بیماران ۳۷۳۴ نفر بود. برای این بیماران اطلاعاتی نظیر سن، جنسیت، مدت بیماری، شاخص توده‌ی بدنی (BMI)، وضعیت سیگاری بودن، میزان قند خون ناشتا، سطح کلسترول، تری‌گلیسیرید خون، فشار خون سیستولیک و دیاستولیک، وجود پروتئین در ادرار و هموگلوبین A_1C در نظر گرفته شد.

در این مطالعه برای معرفی یک الگوی رده بندی از بررسی آماری CART به کمک نرم‌افزار $STATISTICA_6$ استفاده گردید. CART یکی از انواع درخت رده بندی می‌باشد که توسط Breiman و همکاران در سال ۱۹۸۴ میلادی معرفی گردید. این مدل

یک گراف غیر چرخشی شبیه درخت با تقسیمات دوتایی بر اساس متغیرهای کمکی را برای معرفی یک الگوی رده بندی و تشخیصی معرفی می‌نماید (۲۲-۱۹، ۱۳، ۱۱).

یک درخت تصمیم از سه جز اصلی شامل ریشه، گره‌ی داخلی و برگ تشکیل شده و روند بدین گونه است که ابتدا یک متغیر کمکی به عنوان ریشه انتخاب و با توجه به اهداف مطالعه به چندین گره‌ی داخلی تقسیم می‌شود. هر گره‌ی داخلی نیز مانند ریشه به گره‌های دیگری تقسیم می‌شود تا در نهایت به هر گره یک رده از متغیر پاسخ منتسب گردد؛ این گره‌ها برگ نامیده می‌شود (۲۴-۲۳، ۱۷، ۱۰). نمودار ۱ نمونه‌ای از درخت رده‌بندی CART می‌باشد.

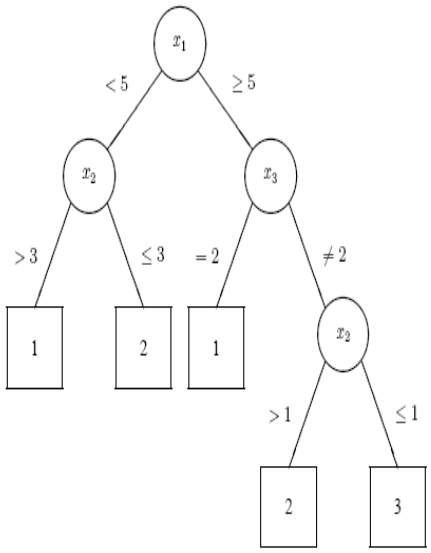
در این بررسی، مشاهدات به صورت تصادفی به دو گروه نمونه‌ی آزمون و یادگیری تقسیم شدند. نمونه‌ی یادگیری (فراگیر) زیرمجموعه‌ای از مشاهدات می‌باشد که درخت رده بندی بر اساس آن طراحی می‌گردد؛ نمونه‌ی آزمون نیز زیرمجموعه‌ای از داده‌هاست که به منظور تعیین دقت درخت طراحی شده با نمونه‌ی یادگیری، مورد استفاده قرار می‌گیرد (۲۴، ۱۷-۱۶).

به منظور انتخاب متغیرهای مهم در الگوی رده بندی درختی، در این بررسی از تابعی تحت عنوان تابع ناجوری (Impurity Function) و شاخصی به نام جینی (Gini) استفاده گردید. تابع ناجوری برای گره‌ای مانند t و متغیر وابسته با k رده ($c_1,\ldots,c_k$) به صورت زیر تعریف می‌شود:

$$i(t) = \Phi[p(c = c_1 \mid t),\ldots, p(c = c_k \mid t)]$$

شاخص جینی اغلب در مدل‌های درختی با تقسیمات دوتایی در هر گره مورد استفاده قرار می‌گیرد و به صورت زیر تعریف می‌شود:

$$i(t) = gini(t) = 1 - \sum_{j=1}^{k} p^2[c = c_j \mid T]$$
$$= \sum_{k \neq l} p(c = c_k \mid T).p(c = c_l \mid T)$$



نمودار ۱. نمونه‌ای از درخت رده بندی CART

این رابطه هنگامی که مشاهدات فقط متعلق به یک رده باشند، برابر صفر است و زمانی که احتمال هر رده برابر باشد، بیشترین مقدار ممکن را اختیار می‌نماید. با در نظر گرفتن متغیر کمکی x که بر اساس آن گره‌ی t به n زیر شاخه تقسیم می‌گردد (هر زیر شاخه با T_j برای j از ۱ تا n نشان داده می‌شود)، یک کاهش در تابع ناجوری خواهیم داشت که بر اساس شاخص جینی به صورت زیر تعریف می‌شود:

$$gini \quad gain = GG(T, X) = gini(t) - \sum_{j=1}^{n} p(T_j \mid T).gini(T_j)$$

از بین چندین متغیر، متغیری مناسب است که مقدار بیشتری را برای $GG(T, X)$ اختیار نماید؛ این ملاکی برای انتخاب بهترین متغیر از بین چندین متغیر می‌باشد (۲۵، ۱۳، ۱۱-۱۰). از این رو با توجه به تابع ناجوری و شاخص جینی ابتدا مقدار تابع ناجوری در حالت کلی برای متغیر پاسخ محاسبه می‌گردد. در مرحله‌ی بعد برای

تمام متغیرهای کمکی، با توجه به بهترین تقسیمات دوتایی برای متغیر پاسخ، مقدار تابع ناجوری در هر یک از دو زیر مجموعه‌ی ایجاد شده محاسبه و میانگین وزنی آن‌ها از مقدار تابع ناجوری کل کم می‌گردد.

از بین متغیرهای کمکی، متغیری که دارای بیشترین مقدار برای این رابطه باشد، در گام اول برای رده بندی درختی انتخاب می‌شود. در برخورد با متغیرهای کمی، از تقسیمات دوتایی استفاده نموده، نقطه‌ای مانند a (نقطه‌ی برش) را تعیین می‌کنیم. لازم به ذکر است که نقطه‌ی برش در بسیاری از الگوهای رده بندی درختی توسط خود شاخص به کار برده شده (در این جا شاخص جینی)، مشخص می‌شود. در برخورد با متغیر کیفی، هر سطح متغیر به عنوان یک زیر شاخه‌ی درخت رده بندی در نظر گرفته می‌شود (۲۹-۲۶، ۱۲).

در مدل CART برای انتخاب اندازه‌ی مناسب از درخت رده بندی، از روشی تحت عنوان ارزش پیچیدگی (Cost complexity) استفاده می‌شود. یک الگوی درختی زمانی مناسب است که علاوه بر این که برای مشاهدات موجود (نمونه‌ی فراگیری) خوب عمل می‌نماید، برای مشاهدات جدید (نمونه‌ی آزمون) نیز مناسب باشد. با وجود این که با بزرگ‌تر شدن اندازه‌ی درخت رده بندی، دقت رده بندی در مورد نمونه‌ی یادگیری افزایش می‌یابد، این دقت برای نمونه‌ی آزمون از اندازه‌ای به بعد کاهش می‌یابد. روش ذکر شده در واقع تعادلی بین دقت درخت رده بندی و اندازه‌ی آن برقرار می‌نماید و با توجه به اندازه‌ی خطا و تعداد گره‌های درخت، تصمیم گیری می‌نماید که کدام گره از درخت را باید حذف نمود (۲۲-۱۹، ۱۳، ۱۱).

در این مطالعه بعد از تعیین الگوی درختی بر اساس مدل CART، برای تعیین مناسب‌ترین مدل، میزان گروه بندی اشتباه، حساسیت و ویژگی الگوی معرفی بر مبنای تشخیص بالینی برای دو نمونه‌ی آزمون و یادگیری محاسبه و با استفاده از آزمون χ^2 مقایسه گردید. به علاوه با روشی مشابه، نتایج حاصل از الگوی درختی معرفی شده با نتایج به دست آمده از روش معمول آماری تابع تشخیص نیز مقایسه گردید.

یافته‌ها

در بین ۳۷۳۴ بیمار تحت بررسی، ۲۰۰۱ بیمار (۵۳/۶ درصد) دارای رتینوپاتی و ۱۷۳۳ بیمار دیگر (۴۶/۴ درصد) بدون رتینوپاتی بودند. میانگین سنی این بیماران برابر ۱۰/۴۴ ± ۵۲/۱۵ سال و میانگین دوره‌ی بیماری دیابت ۶/۰۱ ± ۷/۰۴ سال بود. ۶۴/۳ درصد بیماران را زنان تشکیل می‌دادند و از بین این بیماران تنها ۱۲/۳ درصد سیگاری بودند. جدول ۱ وضعیت بیماران را برای سن، دوره‌ی بیماری، شاخص توده‌ی بدنی، میزان قند خون، کلسترول و تری‌گلیسرید خون، فشار خون سیستولی و دیاستولی و هموگلوبین A_1C را در کل بیماران و در دو گروه دارای رتینوپاتی و بدون رتینوپاتی نشان می‌دهد. با توجه به جدول ۱، به جز تری‌گلیسرید و کلسترول خون، از نظر آماری بین میانگین خصوصیات کمی بررسی شده در دو گروه دارای رتینوپاتی و بدون رتینوپاتی تفاوت آماری معنی‌دار وجود داشت. جدول ۲ فراوانی و درصد فراوانی بیماران را در دو گروه دارای رتینوپاتی و بدون رتینوپاتی با توجه به جنسیت، سیگاری بودن و وجود پروتئین در ادرار نشان می‌دهد. با توجه به جدول ۲، دو گروه دارای رتینوپاتی و بدون رتینوپاتی از نظر درصد فراوانی جنسیت، سیگاری بودن و وجود پروتئین در ادرار تفاوت معنی‌دار داشتند.

با توجه به الگوی رده بندی درختی CART، چهار الگوی درختی با معیارهای مختلف برای ساخت درخت رده بندی معرفی گردید. به کمک میزان گروه‌بندی اشتباه، حساسیت و ویژگی مدل‌های معرفی شده، مناسب‌ترین مدل با توجه به آزمون χ^2 انتخاب گردید. مدل رده بندی درختی به دست آمده مطابق نمودار ۲ بود.

جدول ۱. مقایسه‌ی مقادیر کمی خصوصیات بررسی شده در کل بیماران و دو گروه دارای رتینوپاتی و بدون رتینوپاتی

متغیر	کل بیماران		دارای رتینوپاتی		بدون رتینوپاتی		P-value
	میانگین	SD	میانگین	SD	میانگین	SD	
سن	۵۲/۱۵	۱۰/۴۴	۵۳/۹۱	۱۰/۰۵	۵۰/۱۱	۱۰/۵۱	< ۰/۰۰۱
دوره‌ی بیماری	۷/۰۴	۶/۰۱	۸/۸۳	۶/۴۷	۴/۹۷	۴/۷	< ۰/۰۰۱
شاخص توده‌ی بدنی	۲۷/۲۷	۴/۴۸	۲۶/۶۷	۴/۲۳	۲۷/۹۶	۴/۶۶	< ۰/۰۰۱
قند خون ناشتا	۱۷۵/۵۹	۴۵/۱۳	۱۸۱/۱۴	۴۶/۵۳	۱۶۹/۳۲	۴۲/۶۵	< ۰/۰۰۱
کلسترول	۲۱۸/۵۷	۳۸/۴۸	۲۱۸/۵۲	۳۸/۷۷	۲۱۸/۶۲	۳۸/۱۷	۰/۹۴۲
تری‌گلیسیرید	۲۱۲/۹۳	۱۰۸/۲۸	۲۱۲/۳۵	۱۰۴/۸۴	۲۱۳/۵۹	۱۱۲/۰۵	۰/۷۳۳
فشار خون سیستولی	۱۲۹/۴۵	۱۶	۱۳۱/۸۶	۱۶/۰۷	۱۲۶/۶۳	۱۵/۴۶	< ۰/۰۰۱
فشار خون دیاستولی	۸۰/۵۱	۷/۵۶	۸۰/۷۴	۷/۵۴	۸۰/۲۴	۷/۵۸	۰/۰۵۱
هموگلوبین A_1C	۸/۸۹	۱/۹۴	۹/۰۷	۱/۹۸	۸/۶۹	۱/۸۸	< ۰/۰۰۱

جدول ۲. فراوانی و درصد فراوانی در دو گروه رتینوپاتی و بدون رتینوپاتی برای جنسیت، سیگاری بودن و وجود پروتئین در ادرار

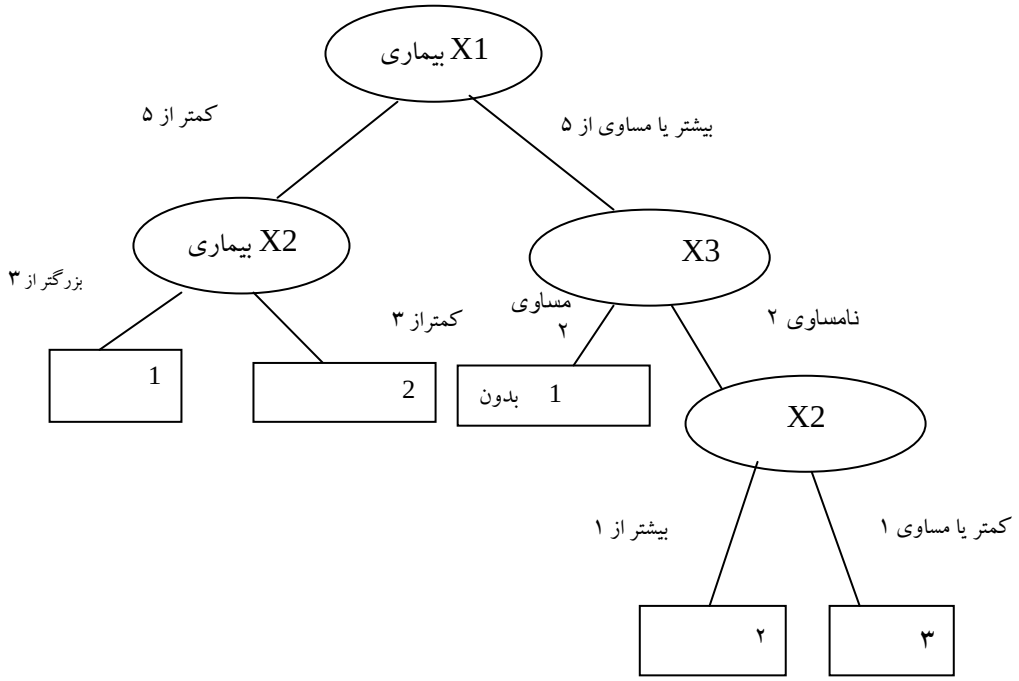
		کل بیماران		دارای رتینوپاتی		بدون رتینوپاتی		P-value
		فراوانی	درصد	فراوانی	درصد	فراوانی	درصد	
جنسیت	مرد	۱۳۳۰	۳۵/۷	۷۷۹	۵۸/۶	۵۵۱	۴۱/۴	< ۰/۰۰۱
	زن	۲۳۹۹	۶۴/۳	۱۲۱۹	۵۰/۸	۱۱۸۰	۴۹/۲	
وضعیت سیگاری	سیگاری	۴۱۳	۱۲/۲	۲۴۵	۵۹/۳	۱۶۸	۴۰/۷	۰/۰۲۳
	غیر سیگاری	۲۹۶۷	۸۷/۸	۱۵۸۵	۵۳/۴	۱۳۸۳	۴۶/۶	
وجود پروتئین در ادرار	دارد	۸۹۲	۳۰/۹	۶۲۹	۷۰/۵	۲۶۳	۲۹/۵	< ۰/۰۰۱
	ندارد	۱۹۹۳	۶۹/۱	۹۸۱	۴۹/۲	۱۰۱۲	۵۰/۸	

جدول ۳. مقایسه‌ی گروه‌بندی اشتباه، حساسیت و ویژگی در نمونه‌ی آزمون و یادگیری

	نمونه‌ی آزمون	نمونه‌ی فراگیر	P-value
گروه‌بندی اشتباه	۰/۳۳	۰/۳۱	۰/۳۴۲
حساسیت	۰/۷۱	۰/۷۲	۰/۷۲۹
ویژگی	۰/۶۲	۰/۶۵	۰/۲۷۶

بر اساس الگوی رده‌بندی درختی معرفی شده، بیماران به نه رده‌ی متمایز دسته بندی می‌شوند. با توجه به این مدل درختی، دوره‌ی بیماری دیابت مهمترین نقش را برای تعیین رتینوپاتی دیابتی ایفا می‌کند.

جدول ۳ معیارهای صحت الگوی رده بندی درختی را برای نمونه‌ی فراگیر و آزمون نشان می‌دهد. این جدول نشان می‌دهد که بین مقادیر گروه‌بندی اشتباه، حساسیت و ویژگی در دو نمونه‌ی آزمون و یادگیری تفاوت معنی‌داری از نظر آماری وجود نداشت. از این رو می‌توان نتیجه گرفت که الگوی درختی معرفی شده برای مشاهدات جدید می‌تواند با دقتی در حدود ۶۷ درصد تصمیم‌گیری نماید که این تصمیم‌گیری از حساسیت ۷۱ و ویژگی ۶۲ درصد برخوردار می‌باشد.

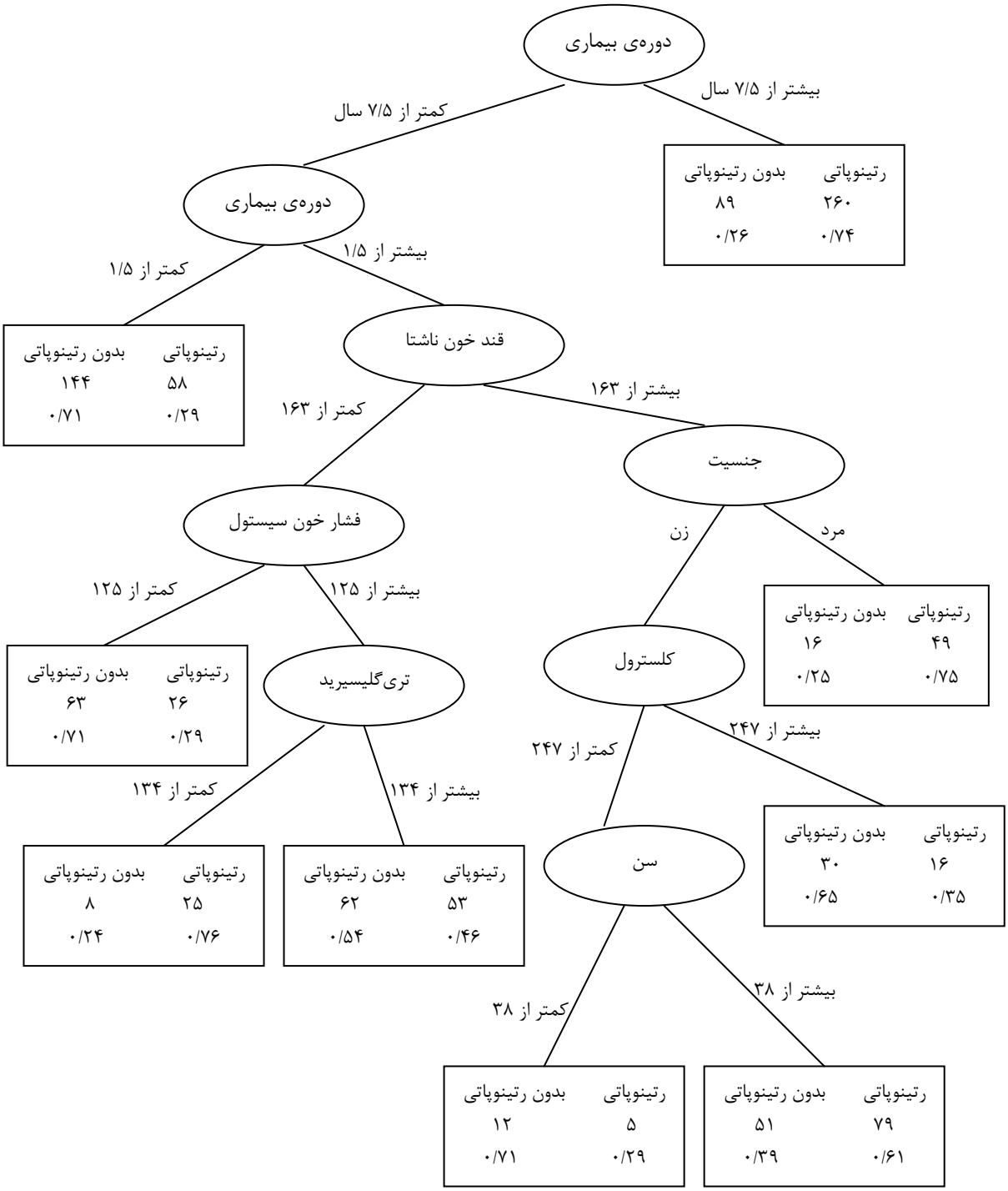


نمودار ۱: نمونه ای از درخت رده‌بندی CART

جدول ۴. ضرایب استاندارد شده‌ی متغیرهای مهم در تابع تشخیص

متغیر	ضریب استاندارد شده
جنسیت	-۰/۱۴۴
دوره‌ی بیماری	۰/۶۰۸
شاخص توده‌ی بدنی	-۰/۲۲۲
قند خون ناشتا	۰/۲۴۷
فشار خون سیستولی	۰/۶۱۵
فشار خون دیاستولی	-۰/۳۸۴
وجود پروتئین در ادرار	۰/۳۱۹

در این بررسی برای اطمینان از کارآیی مدل درختی، الگوی رده بندی بر اساس روش کلاسیک آماری تابع تشخیص نیز به دست آمد. تابع تشخیص با توجه به متغیرهای جنسیت، دوره‌ی بیماری، شاخص توده‌ی بدنی، میزان قند خون ناشتا، فشار خون سیستولی و دیاستولی و وجود پروتئین در ادرار، یک الگوی رده‌بندی خطی با ضرایب استاندارد شده مطابق جدول ۴ را معرفی نمود.



نمودار ۲. درخت رده‌بندی با اندازه‌ی نه بر اساس نمونه‌ی فراگیر و آزمون

بحث

با توجه به یکسان بودن میزان گروه‌بندی اشتباه در نمونه‌ی آزمون و یادگیری می‌توان نتیجه گرفت که مدل درختی معرفی شده می‌تواند برای مشاهدات جدید با دقتی معادل ۶۷ درصد تصمیم‌گیری نماید. بر اساس مدل معرفی شده، دوره‌ی بیماری دیابت به عنوان مهم‌ترین عامل برای رده بندی شناخته شد؛ به طوری که بیماران دیابتی نوع II که از سابقه‌ی بیماری آن‌ها بیشتر از ۷/۵ سال می‌گذرد، نسبت به گروه‌های دیگر، در معرض خطر بیشتری برای رتینوپاتی هستند. بسیاری از مطالعات دیگر نیز رابطه‌ی مستقیمی بین دوره‌ی بیماری دیابت و شیوه‌ی رتینوپاتی نشان داده و دوره‌ی بیماری به عنوان یکی از مهم‌ترین عوامل تأثیرگذار در رتینوپاتی شناخته شده است (۳۳-۳۰، ۶). با این حال، الگوی معرفی شده توسط درخت تصمیم، یک نمودار قابل درک و ساده برای رده بندی معرفی می‌نماید (۳۵-۳۴). به علاوه در الگوهای درخت تصمیم، نیازی به در نظر گرفتن اثرات متقابل نیست. در این الگوها اگر اثر متقابل وجود داشته باشد، در درخت تصمیم بدون در نظر گرفتن ظاهر می‌گردد. همچنین شدت عوامل مؤثر بر ابتلا به بیماری را تعیین می‌کند (۱۶، ۱۰).

با توجه به مطالعه‌ی انجام شده در آمریکا، با گروه‌بندی بر روی دوره‌ی بیماری تفاوتی بین شیوع رتینوپاتی در دیابت نوع II و نوع I دیده نشد (۳۶).

الگوی درختی معرفی شده در این بررسی، فشار خون سیستولیک را نیز به عنوان یک عامل برای رده بندی معرفی می‌نماید که تأثیرات فشار خون، به خصوص فشار خون سیستول، بر روی رتینوپاتی در چندین مطالعه قید شده است و فشار خون به عنوان یک ریسک فاکتور رتینوپاتی شناخته می‌شود (۳۹-۳۷، ۹، ۶). بر اساس نتایج به دست آمده از مدل درختی، بیماران با سطح پایین تری‌گلیسیرید و کلسترول نسبت به گروه دیگر بیماران در خطر بیشتری قرار دارند؛ این نکته در بعضی از مطالعات ذکر شده ولی اختلاف معنی‌داری بین دو گروه دیده نشده است. با توجه به ارتباط مستقیم شاخص توده‌ی بدنی با این دو عامل، نتایج به دست آمده در این مطالعه با نتایج دیگر مطالعات (۳۸-۳۷، ۹، ۵) همخوانی دارد.

مدل درختی معرفی شده در این بررسی، حاکی از اختلاف شیوع رتینوپاتی در مدل تابع تشخیص معرفی شده است. این مدل به منظور رده بندی رتینوپاتی از میزان دقت ۰/۶۷ با حساسیت ۰/۷۱ و ویژگی ۰/۶۴ برخوردار می‌باشد.

باید توجه داشت که در مقایسه‌ی مدل درخت رده بندی و تابع تشخیص، هدف مقایسه‌ی تک تک متغیرها نمی‌باشد؛ بلکه همان طور که اشاره شد، برای اطمینان از کارآیی مدل درختی، میزان دقت درخت رده‌بندی و تابع تشخیص مقایسه گردید که از نظر آماری اختلاف معنی‌داری بین آن‌ها دیده نشد.

مدل درختی معرفی شده در این بررسی، حاکی از اختلاف شیوع رتینوپاتی بین مردان و زنان بود. با این که در مطالعات مختلف شیوع بالاتر رتینوپاتی در مردان نسبت به زنان عنوان شده ولی در اکثر مطالعات جنسیت به عنوان یک عامل تأثیرگذار شناخته نشده است (۳۷-۳۶، ۹، ۶، ۳).

نتیجه‌گیری

در این بررسی مشخص گردید که دوره‌ی بیماری دیابت یکی از مهم‌ترین عوامل تعیین کننده‌ی رتینوپاتی می‌باشد؛ به طوری که خطر رتینوپاتی در بیماران دیابت

نوع II، که از بیماری آن‌ها بیشتر از ۷/۵ سال می‌گذرد، نسبت به سایر بیماران بیشتر می‌باشد. همچنین بر اساس مدل CART می‌توان به خوبی مدل تابع تشخیص، در مورد بیماری رتینوپاتی در بیماران دیابت نوع II تصمیم‌گیری نمود.

هر چند در این مطالعه دوره‌ی بیماری به عنوان مهم‌ترین عامل رتینوپاتی شناخته شد اما بایستی توجه داشت که تعریف دوره‌ی بیماری، مدت زمان تشخیص دیابت بیمار تا شروع مطالعه بوده و در آن، دوره‌ی کمون لحاظ نشده است. همچنین یکی دیگر از محدودیت‌های مطالعه‌ی حاضر آن بود که درخت‌های رده‌بندی اغلب در هر مرحله از تقسیم، فقط از یک متغیر بهره می‌برند (۱۴-۱۳).

References

1. Vrijhoef HJ, Diederiks JP, Spreeuwenberg C. Effects on quality of care for patients with NIDDM or COPD when the specialised nurse has a central role: a literature review. Patient Educ Couns 2000; 41(3): 243-50.
2. Rodriguez J, Sanchez R, Munoz B, West SK, Broman A, Snyder RW, et al. Causes of blindness and visual impairment in a population-based sample of U.S. Hispanics. Ophthalmology 2002; 109(4): 737-43.
3. Jamal-u-Din, Qureshi MB, Khan AJ, Khan MD, Ahmad K. Prevalence of diabetic retinopathy among individuals screened positive for diabetes in five community-based eye camps in northern Karachi, Pakistan. J Ayub Med Coll Abbottabad 2006; 18(3): 40-3.
4. Gupta S, Ambade A. Prevalence of diabetic retinopathy and influencing factors amongst type 2 diabetics from central India. Int J Diab Dev Ctries 2004; 24:75-8.
5. Rema M, Deepa R, Mohan V. Prevalence of retinopathy at diagnosis among type 2 diabetic patients attending a diabetic centre in South India. Br J Ophthalmol 2000; 84(9): 1058-60.
6. Brown JB, Pedula KL, Summers KH. Diabetic retinopathy: contemporary prevalence in a well-controlled population. Diabetes Care 2003; 26(9): 2637-42.
7. Fonseca V, Munshi M, Merin LM, Bradford JD. Diabetic retinopathy: a review for the primary care physician. South Med J 1996; 89(9): 839-50.
8. Marshall G, Garg SK, Jackson WE, Holmes DL, Chase HP. Factors influencing the onset and progression of diabetic retinopathy in subjects with insulin-dependent diabetes mellitus. Ophthalmology 1993; 100(8): 1133-9.
9. Harney F. Diabetic retinopathy. Medicine-Abingdon 2006; 34(3): 95-8.
10. Podgorelec V, Kokol P, Stiglic B, Rozman I. Decision trees: an overview and their use in medicine. J Med Syst 2002; 26(5): 445-63.
11. Breiman L, Friedman J, Stone CJ, Olshen RA. Classification and regression trees. 1st ed. London: Chapman and Hall/CRC; 1984.
12. Marshall RJ. The use of classification and regression trees in clinical epidemiology. J Clin Epidemiol 2001; 54(6): 603-9.
13. Timofeev R. Classification and patients attending a diabetic center Regression Trees (CART) Theory and application. [Thesis]. Berlin, Humboldt Uni; 2004
14. Tan PN, Steinbach M, Kumar V. Introduction to data mining. Canada: Pearson Addison Wesley; 2006. p. 145-206.
15. Yohannes Y, Webb P. Classification and regression trees, CART: a user manual for identifying indicators of vulnerability to famine and chronic food insecurity. Washington, DC: International Food Policy Research Institute; 1999.
16. Lewis RJ. An Introduction to Classification and Regression Tree (CART) Analysis," presented at Annual Meeting of the Society for Academic Emergency Medicine. Annual Meeting of the Society of Academic Emergency Medicine. 2000. Available from: URL: [http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.95.4103]
17. Rema M, Premkumar S, Anitha B, Deepa R, Pradeepa R, Mohan V. Prevalence of diabetic retinopathy in urban India: the Chennai Urban Rural Epidemiology Study (CURES) eye study, I. Invest Ophthalmol Vis Sci 2005; 46(7): 2328-33.
18. Safavian SR, Landgrebe D. A survey of decision tree classifier methodology. IEEE Trans Systems, Man, & Cybernetics 1991; 21(3): 660-7.
19. Mitra S, Acharya T. Data Mining: multimedia, soft computing, and bioinformatics. London: John Wiley and Sons; 2003.
20. Wilkinson L. Tree structured data analysis: AID, CHAID, and CART. Proceedings of Sawtooth

Software Conference; 1992; Sun Valley.
21. Kim H, Loh WY. Classification trees with unbiased multiway split. J Amer Statist Assoc 2001; 98: 598-604.
22. Scoh C, Willett R, Nowak R. CORT: Classification and regression Trees. Proceedings of the 6th International Conference of IEEE; April 2003.
23. Utgoff PE. Decision trees. In: Wilson RA, Keil FC, Editors. The MIT encyclopedia of the cognitive sciences. Bradford: MIT Press; 1998.
24. Quinlan JR. C4.5: programs for machine learning. San Francisco: Morgan Kaufmann Publishers Inc; 1994. p. 235-40.
25. Buntine W, Niblett T. A further comparison of splitting rules for decision-tree induction. Machine Learning 1992; 8(1): 75-85.
26. Quinlan JR. Induction of decision trees. Machine Learning 1986; 1(1): 81-106.
27. Shih YS. Selecting the best splits for classification trees with categorical variables. Statistics & Probability Letters 2001; 54(4): 341-5.
28. Loh WY, Shih YS. Split selection methods for classication trees. Statistica Sinica 1997; 7: 815-40.
29. Hoare R. Using CHAID for classification problems. Proceedings of New Zealand Statistical Association Conference on machine learning; 1997. New Zealand.
30. Dobra AV. Scalable classification and regression tree construction. New York: Cornell University; 2003.
31. Kull CE, Abrahasson M. Prevalence of retinopathy difference in a population with type 1 diabetes. Diabet Med 2002; 19(11): 923-4.
32. Porta M, Bandello F. Diabetic retinopathyA clinical update. Diabetologia 2002; 45(12): 1617-34.
33. Dandona L, Dandona R, Naduvilath TJ, McCarty CA, Rao GN. Population based assessment of diabetic retinopathy in an urban population in southern India. Br J Ophthalmol 1999; 83(8): 937-40.
34. Frank E, Wang Y, Inglis S, Holmes G, Witten IH. Using model trees for classification. Machine Learning 1998; 32(1): 63-76.
35. Frank E. Pruning decision trees and lists. Newzealand: Waikato Univ; 2000.
36. Varma R, Torres M, Pena F, Klein R, Azen SP. Prevalence of diabetic retinopathy in adult Latinos: the Los Angeles Latino eye study. Ophthalmology 2004; 111(7): 1298-306.
37. Janghorbani M, Amini M, Ghanbari H, Safaiee H. Incidence of and risk factors for diabetic retinopathy in Isfahan, Iran. Ophthalmic Epidemiol 2003; 10(2): 81-95.
38. Abdollahi A, Malekmadani MH, Mansoori MR, Bostak A, Abbaszadeh MR, Mirshahi A. Inflammatory markers and diabetic retinopathy in type 1 diabetes. Acta Medica Iranica, 2006; 44(6): 415-9.
39. Izuora KE, Chase HP, Jackson WE, Coll JR, Osberg IM, Gottlieb PA, et al. Inflammatory markers and diabetic retinopathy in type 1 diabetes. Diabetes Care 2005; 28(3): 714-5.