



ORIGINAL ARTICLE

Received: 2014/11/22

Accepted: 2015/01/11

Comparing of Decision Tree with Logistic Regression Model in Evaluating Osteoporosis

Mohsen Askarishahi(Ph.D.)¹, Nasime Ghasemi(M.Sc.)², Hossein Flahzadeh(Ph.D.)³, Mohammad Afkhami-Ardekani (Ph.D.)⁴, Arezoo Afkhami- Ardekani(M.Sc.)⁵

1. Assistant Professor, Department of Biostatistics, School of Public Health, Shahid Sadoughi University of Medical Sciences, Yazd, Iran.

2. Corresponding Author: M.Sc. student of Biostatistics, School of Public Health, Shahid Sadoughi University of Medical Sciences, Yazd, Iran. Email: nasimBghasemi@gmail.com Tel: 09135610985

3. Professor, Research center of prevention and Epidemiology of non-communicable disease, School of Public Health, Shahid Sadoughi University of Medical Sciences, Yazd, Iran.

4. Professor, Department of Endocrine Diseases And Metabolism, School of Medical, Shahid Sadoughi University of Medical Sciences, Yazd, Iran.

5. Medical Student, School of Medical, Shahid Sadoughi University of Medical Sciences, Yazd, Iran.

Abstract

Introduction: Early detection of osteoporosis is a key to preventing of it; but recognition, without the use of appropriate diagnostic methods, due to the complexity of risk factors and gradual bone loss process, is problem. The purpose of this study is to develop and efficiency evaluation a predictive model of osteoporosis using decision tree technique as a diagnostic method based on available risk factors; thereby to identify individuals at risk for preventive activities.

Methods: In this study used data from 131 women aged 20 – 40 years. Response variable was amount of BMD (t-score) L1-L4 lumbar region that divided on two group, normal (t-score ≥ -1) and at risk of osteoporosis (t-score < -1). To determine risk factors of osteoporosis used from decision tree model with method of k- fold cross validation $k=4$ and logistic regression. To assess the accuracy prediction of two model, the area under receiver operative characteristic curves (AUROC) was used. Data analysis was performed by R software.

Results: Three variables number of pregnancies, BMI and calcium levels as risk factors for osteoporosis were obtained from the decision tree model and Area under receiver operative characteristic decision tree and logistic regression, respectively 0.665 and 0.686 were obtained.

Conclusion: Area under receiver operative characteristic curve showed advantage superiority of logistic regression that according to advantages of the decision tree applying simultaneously of two models is recommended.

Keywords: Area under receiver operative characteristic, Decision Tree, Osteoporosis, Regression Logistic

Conflict of interest: The authors declared that there is no Conflict interest

**This Paper Should be Cited as:**

Mohsen Askarishahi, Nasime, Ghasemi, Hossein Flahzadeh, Mohammad Afkhami Ardekani, Arezoo Afkhami Ardekani. Comparing of Decision Tree with Logistic Regression Model in Evaluating Osteoporosis..... Toloobehdasht Journal. 2018; 17(1):14-23. [Persian]



مقایسه مدل درخت تصمیم و رگرسیون لوجستیک در ارزیابی پوکی استخوان

نویسندگان: محسن عسکری شاهی^۱، نسیمه قاسمی^۲، حسین فلاح زاده^۳، محمد افخمی اردکانی^۴، آرزو افخمی اردکانی^۵

۱. استادیار آمار زیستی، گروه آمار، دانشکده بهداشت، دانشگاه علوم پزشکی و خدمات بهداشتی درمانی شهید صدوقی یزد، یزد، ایران.

۲. نویسنده مسئول: دانشجوی کارشناسی ارشد آمار زیستی، دانشکده بهداشت، دانشگاه علوم پزشکی و خدمات بهداشتی درمانی شهید صدوقی یزد، یزد، ایران.

تلفن تماس: ۰۹۱۳۵۶۱۰۹۸۵ Email: nasimbghasemi@Gmail.com

۳. استاد مرکز تحقیقات پیشگیری و اپیدمیولوژی بیماری های غیرواگیر، دانشکده بهداشت، دانشگاه علوم پزشکی و خدمات بهداشتی درمانی شهید صدوقی یزد، یزد، ایران.

۴. استادیاریهای غدد داخلی و متابولیسم، دانشکده پزشکی، دانشگاه علوم پزشکی و خدمات بهداشتی درمانی شهید صدوقی یزد، یزد، ایران.

۵. دانشجوی پزشکی دانشکده پزشکی، دانشگاه علوم پزشکی و خدمات بهداشتی درمانی شهید صدوقی یزد، یزد، ایران.

چکیده

مقدمه: تشخیص زودهنگام استئوپروز کلیدی برای پیشگیری است، لیکن تشخیص، بدون استفاده از روش های تشخیصی مناسب، به علت پیچیدگی عوامل خطر استئوپروز و روند تدریجی از دست رفتن استخوان مشکل است. هدف این مطالعه ارائه و سنجش کارایی یک الگوی پیشگویی استئوپروز با استفاده از تکنیک درخت تصمیم به عنوان یک روش تشخیصی براساس عوامل خطر در دسترس است تا به وسیله آن افراد در معرض خطر برای انجام فعالیت های پیشگیرانه شناسایی شوند.

روش بررسی: برای انجام این مطالعه از داده های ۱۳۱ زن با سن ۲۰ تا ۴۰ سال استفاده شد. متغیر پاسخ مقدار تراکم استخوان (t-score) ناحیه L۴-L۱ کمتری بود که به دو گروه نرمال ($t\text{-score} \geq -1$) و در معرض خطر استئوپروز ($t\text{-score} < -1$) تقسیم شد. به منظور تعیین عوامل خطر استئوپروز از مدل درخت تصمیم به روش ارزیابی مقاطع با $k = 4$ و رگرسیون لوجستیک استفاده شد. برای بررسی دقت پیش بینی دو مدل، سطح زیرمنحنی مشخصه عملکرد (AUROC) به کار گرفته شد. تحلیل داده ها با نرم افزار R انجام شد.

یافته ها: سه متغیر تعداد حاملگی، BMI و میزان کلسیم به عنوان عوامل خطر استئوپروز از مدل درختی به دست آمد و سطح زیر منحنی مشخصه عملکرد درخت تصمیم و رگرسیون لوجستیک به ترتیب برابر ۰/۶۶۵ و ۰/۶۸۶ حاصل شد.

نتیجه گیری: سطح زیر منحنی مشخصه عملکرد برتری نسبی رگرسیون لوجستیک را نشان داد که با توجه به مزایای درخت تصمیم به کار گیری همزمان دو روش پیشنهاد می شود.

واژه های کلیدی: پوکی استخوان، درخت تصمیم، رگرسیون لوجستیک، سطح زیر منحنی مشخصه عملکرد

این مقاله برگرفته از پایان نامه دانشجوی کارشناسی ارشد آمار زیستی، دانشکده بهداشت، دانشگاه علوم پزشکی و خدمات بهداشتی درمانی شهید صدوقی یزد می باشد.

طلوع بهداشت

دو ماهنامه علمی پژوهشی

دانشکده بهداشت یزد

سال هفدهم

شماره: اول

فروردین و اردیبهشت ۱۳۹۷

شماره مسلسل: ۶۷

تاریخ وصول: ۱۳۹۳/۰۹/۰۱

تاریخ پذیرش: ۱۳۹۳/۱۰/۲۱



مقدمه

استئوپروز یا پوکی استخوان یکی از رایج ترین بیماری متابولیک استخوان می باشد که با کاهش توده استخوانی و از بین رفتن ساختار آن مشخص می شود. استئوپروز یک شاخص سلامت اقتصاد در هر کشوری است و تشخیص جمعیت در معرض خطر استئوپروز به عنوان یک مسئله چالش برانگیز باقی مانده است. براساس آخرین گزارش سالیانه سازمان بهداشت جهانی بیش از ۲۰۰ میلیون زن در جهان از استئوپروز رنج می برند. در ایران نیز مطالعات مبتنی بر جمعیت نشان دادند که ۱۰/۶ درصد از جمعیت استئوپروز و ۲۱ درصد از استئوپنی رنج می برند (۱). سازمان سلامت جهانی (WHO) معیاری برای تشخیص استئوپروز براساس اندازه گیری تراکم معدنی استخوان (Bone Mineral Density) به وسیله جذب دوگانه اشعه ایکس (DEXA) تعریف کرده است. که بر این اساس استئوپروز را به صورت کاهش تراکم استخوان $2/5$ انحراف معیار کمتر از متوسط تراکم استخوانی در افراد جوان جامعه و استئوپنی به صورت کاهش تراکم استخوان بین $1 - 2/5$ انحراف معیار کمتر از متوسط تراکم استخوانی در افراد جوان جامعه است (۲،۳). اندازه گیری BMD به وسیله DEXA در بیشتر کشورها دارای محدودیت هایی است چون این روش هزینه بر و تجهیزات آن به طور عموم در دسترس نیست بعلاوه پرتو ایکس در آن هرچند مقدار آن ناچیز است ولی انجام آن برای غربالگری و افراد سالم توصیه نمی شود (۴).

از سوی دیگر به منظور رده بندی مشاهدات روش های متفاوتی وجود دارد. رگرسیون، تحلیل ممیزی، مدل های لگ خطی، آنالیز خوشه ای و مدل های لجستیک از جمله روش های رده بندی

کلاسیک است. استفاده از این روش های کلاسیک محدودیتی مانند لزوم در نظر گرفتن یک توزیع احتمال برای هر یک از متغیرهای پاسخ، خطی بودن رابطه ی بین متغیرهای وابسته و متغیرهای مستقل، یکسان بودن واریانس خطاها، عدم وجود هم خطی چندگانه و اثرات متقابل مرتبه بالا می تواند باعث ایجاد خطا در مدل ها، برآوردها و پیش بینی های آن ها گردد. بنابراین نتایج به دست آمده از این روش ها در عمل کاربرد چندانی ندارد (۵،۶).

از این رو جایگزین کردن روش هایی که بتوان به واسطه آن افراد را رده بندی و تعیین کرد کدامیک باید مورد سنجش تراکم استخوانی قرار گیرند مناسب به نظر می رسد. روش های رده بندی اخیر برآنالیزهای چندگانه پارامتری و ناپارامتری متکی است. درخت تصمیم از روش های ساده و قدرتمند آنالیز چند متغیره است که نتایج رده بندی را به وسیله نمودار ساده و قابل درک ارائه می دهد (۷-۹).

ساختار درخت تصمیم مانند یک نوع انتخاب متغیر با هدف تفکیک فضای مشاهدات است. در هر مسئله فرض می شود یک تفکیک از فضای مشاهدات که به درستی همه مشاهدات را طبقه بندی کند وجود دارد و امر مهم پیدا کردن یک درخت برای توصیف آن به طور فشرده است (۱۰).

نقطه شروع درخت رده بندی گره اصلی نامیده می شود که در بالای درخت قرار دارد و کل مجموعه یادگیری را در بر می گیرد (۱۱). داده ها روی متغیری که نتایج، بیشترین اختلاف بین دو گره متوالی را داشته باشد تقسیم می شود. در هر گره داخلی ایجاد شده متغیرها دوباره برای پیدا کردن متغیر پیش بین که نتایج در بهترین تقسیم برای متغیر پاسخ قرار گیرد آزمون



بدنی (BMI)، میزان سرپا ایستادن در روز، سابقه فامیلی، سابقه درد های استخوانی، شغل، منزل و پارامترهای شیمیایی میزان کلسیم، ویتامین D و سطح هورمون پاراتیروئید (PTH). متغیر پاسخ به دو رده سطح تراکم استخوانی بالا ($t\text{-score} \geq -1$) افراد نرمال) با کد ۱ و تراکم استخوانی پایین ($t\text{-score} < -1$) افراد استئوپروز و استئوپنی) با کد ۰ تقسیم شد. تحلیل داده ها به کمک نرم افزار R انجام شد.

ساختن مدل: الگوریتم CART در این مطالعه براساس روش ارائه شده توسط Brieman است. CART اختصار طبقه بندی درختی (Classification Tree) و رگرسیون درختی (Regression Tree) است. وقتی متغیر پاسخ پیوسته باشد رگرسیون درختی و وقتی گسسته باشد درخت طبقه بندی نامیده می شود (۱۳، ۱۴).

برای انتخاب درخت، تابع ناچوری (Impurity Function) با شاخص جینی (Gini Index) با کمترین تغییرات ناچوری در مقدار $0/0001$ انتخاب شد. برای جلوگیری از بیش برآزشی شاخه هایی که در آن ها تعداد مشاهدات کمتر از ۲۰ باشد هرس (Prune) شدند. هرس کردن تکنیکی است که اندازه درخت تصمیم را با حذف بخش هایی از درخت که توان کمی برای مثال مورد طبقه بندی دارد کاهش می دهد. هدف از هرس کردن کاهش پیچیدگی رده بندی نهایی است به طوریکه دقت پیش بینی با کاهش بیش برآزشی و حذف بخشی از رده بندی که ممکن است مبتنی بر اغتشاشات یا داده های نادرست باشد بهبود یابد (۱۳).

برای ارزیابی مدل ایجاد شده به علت کم بودن حجم نمونه از روش اعتبار متقاطع (Cross Validation) $k\text{-fold}$ ، $k = 4$

می شوند. تقسیمات تا وقتی که معیار توقف محقق شود ادامه می یابد. آزمودنی ها در گره پایانی (برگ) برحسب متغیر پاسخ طبقه بندی می شوند (۲، ۱۲).

در این مطالعه قصد ایجاد یک درخت تصمیم براساس متغیرهایی را داریم که اندازه گیری آن ها در هر شرایطی وجود دارد تا بتوان به وسیله آن افراد را براساس این متغیرها به گروه های BMD نرمال و BMD پایین تفکیک کنیم تا افرادی که در معرض خطر استئوپروز هستند با یک غربالگری ساده از عوامل تعیین شده در الگوی درختی شناسایی شوند. همچنین برای بررسی کارایی مدل در پیش بینی، دقت مدل درختی با رگرسیون لوجستیک مورد مقایسه قرار می گیرد.

روش بررسی

داده ها در این مطالعه شامل ۱۳۱ زن ۲۰ تا ۴۰ سال از افراد شرکت کننده در مطالعه طرح سنجش ویتامین D، پارامترهای بیوشیمیایی و سنجش تراکم استخوان در سال ۱۳۸۸ در یزد است. در این مطالعه ۷۰۰ نفر مرد وزن به صورت نمونه گیری خوشه ای انتخاب شدند به این صورت که ۳۵ کد پستی انتخاب شده به صورت تصادفی به عنوان نقطه شروع نمونه گیری در نظر گرفته شد. مشخصات دموگرافیک، آنتروپومتریک و اطلاعات بالینی به دست آمده از پرسشنامه، پارامترهای بیوشیمیایی به دست آمده از آزمایش خون برای کل افراد ثبت شد. نمونه مورد نظر در این بررسی افرادی هستند که علاوه بر اطلاعات بالا، شاخص سنجش استخوان به روش DEXA در ناحیه $L1-L4$ مهره های کمری برای آنها اندازه گیری شده باشد که این تعداد برای زنان ۱۳۱ نفر است. متغیرهای مستقل مورد بررسی سن، سن منارک، مدت شیردهی، تعداد حاملگی، شاخص توده



عملکرد (Area under receiver operative characteristic) برای هر مدل محاسبه شد. اگر این مقدار بین ۰ - ۰/۵ باشد تصادفی بودن کلاس بندی و اگر بین ۰/۵ - ۱ باشد توان بالای مدل بندی را نشان می دهد(۱۵).

استفاده شد. در این روش داده ها به طور تقریبی به ۴ قسمت تقسیم می شوند. ابتدا سه قسمت از داده ها برای ساختن درخت و قسمت چهارم برای آزمون مدل به کار می رود. فرآیند برای ۳ قسمت دیگر تکرار می شود تا وقتی هر قسمت از داده ها در یک زمان به عنوان مجموعه آزمون به کار رود. میانگین ۴ خطای طبقه بندی به عنوان نرخ اعتبار متقاطع گزارش می شود(۱۲). مدل رگرسیون لجستیک با تمام متغیرها برازش داده شد و انتخاب بهترین مدل به روش گام به گام انجام شد. همچنین مقادیر حساسیت، ویژگی، دقت و سطح زیر منحنی مشخصه جدول ۱: شاخص های آمار توصیفی مربوط به متغیرهای کمی و کیفی برای کل افراد، افراد با BMD نرمال و BMD

یافته ها

از ۱۳۱ زن مورد بررسی در این مطالعه ۹۶ (۰/۷۳) نفر نرمال و ۳۲ (۰/۲۴) نفر استئوپنی و ۳ (۰/۰۲) استئوپروز بودند. مشخصات افراد در جدول ۱ خلاصه شده است.

جدول ۱: شاخص های آمار توصیفی مربوط به متغیرهای کمی و کیفی برای کل افراد، افراد با BMD نرمال و BMD

متغیر	کل افراد	افراد با BMD نرمال	افراد با BMD پایین (استئوپنی و استئوپروز)	سطح معنی داری
سن	30.7 ± 3.9	30.1 ± 5.6	32.1 ± 5.8	0.06
شاخص توده بدنی	25 ± 4.5	25.8 ± 4.3	24.8 ± 4.8	0.2
سن منارک	13.68 ± 1.4	12.5 ± 1.2	14.24 ± 1.8	0.03*
تعداد حاملگی	1.8 ± 1.3	1.69 ± 1.2	2.4 ± 1.4	<0.001*
مدت شیردهی (ماه)	30.8 ± 23.8	29.2 ± 23.3	35.3 ± 24.7	0.1
میزان کلسیم	9.5 ± 0.5	9.6 ± 0.5	9.5 ± 0.35	0.9
میزان ویتامین د	27.5 ± 16.5	27.4 ± 16.1	27.6 ± 17.8	0.9
میزان pth	23.2 ± 9.2	22.6 ± 9	25 ± 9.7	0.1
سابقه فامیلی استئوپروز	11 (8%)	8 (7.3%)	3 (2.7%)	1
سابقه دردهای استخوانی	120 (92%)	88 (73%)	32 (27%)	0.2
مدت سر پا ایستادن > 4 ساعت	86 (66%)	61 (71%)	25 (24%)	0.5
در روز < 4 ساعت	45 (34%)	35 (77%)	10 (22%)	
وضعیت شغل	29 (22%)	23 (79%)	6 (21%)	0.5
سایر	102 (78%)	73 (71%)	29 (28%)	

*معنی دار در سطح 5 درصد، † mean ± SD، + فراوانی (درصد)



تعداد حاملگی ۴ و بیشتر در خطر ابتلا به پوکی استخوان قرار دارند با احتمال ۵۳ در صد. و زنانی که کمتر از ۴ تعداد حاملگی دارند و BMI آن ها کمتر از ۱۹ است در معرض خطر پوکی استخوان طبقه بندی شدند با احتمال ۵۷ درصد. متغیرهای تعداد حاملگی، سن منارک، دردهای استخوانی و مدت شیردهی در مدل رگرسیون لوجستیک به عنوان عوامل مهم به دست آمد.

خلاصه ای از ضرایب این مدل در جدول ۲ ارائه شده است. به طور کلی تعداد حاملگی در هر دو مدل درختی و رگرسیون لوجستیک یک عامل خطر مهم شناخته شد.

مقادیر حساسیت، ویژگی و دقت دو مدل در جدول ۳ آورده شده که میزان حساسیت، ویژگی و دقت مدل درختی به ترتیب برابر ۳۷، ۷۴ و ۶۶ درصد به دست آمد و برای مدل رگرسیون لوجستیک به ترتیب ۵۷، ۷۵ و ۷۱ درصد محاسبه شد. برای مقایسه دو مدل، مقدار سطح زیر منحنی مشخصه عملکرد (AUROC) محاسبه شد که برای مدل درختی ۶۶/۵ درصد و برای رگرسیون لوجستیک ۶۸/۶ درصد به دست آمده است.

ارتباط تک تک متغیرها سنجیده شد که نشان می دهد از بین متغیرهای کمی متغیرهای سن منارک و تعداد حاملگی در دو گروه BMD نرمال و BMD پایین از نظر آماری اختلاف معنی داری داشتند به این معنی که میانگین سن منارک و میانگین تعداد حاملگی در افراد در معرض خطر (BMD پایین) بیشتر است. ولی در بقیه متغیرها اختلاف معنی داری دیده نشد. با استفاده آنالیز درخت تصمیم توانسته ایم به درستی جمعیت مورد مطالعه را به زیر گروه هایی با استفاده از متغیرهای پیش بین افراز کنیم. نمودار ۱ نشان می دهد ۳ متغیر تعداد حاملگی، شاخص توده بدنی (BMI) و میزان کلسیم در پیش بینی خطر استئوپروز نقش تعیین کننده دارد و چون تعداد حاملگی در بالای درخت ظاهر شده مهمترین عامل شناخته می شود.

و بعد از آن متغیرهای BMI و میزان کلسیم قرار دارد. نرخ خطای اعتبار متقاطع ۰/۳۳ با انحراف معیار ۰/۱۵. برای این درخت به دست آمد.

قوانینی را از درخت تصمیم می توان دریافت کرد مثلا زنان با

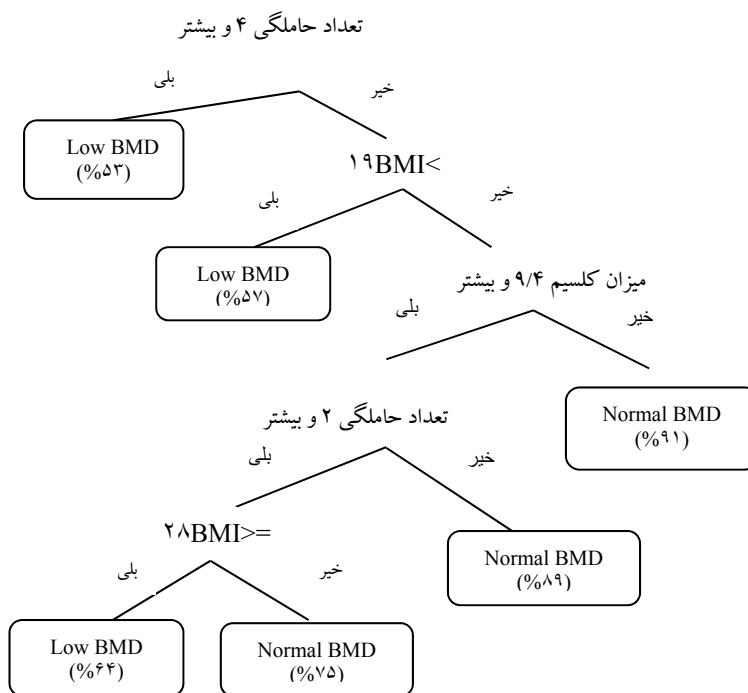
جدول ۲: ضرایب متغیرهای به دست آمده از مدل رگرسیون لوجستیک

ضریب	Exp(ضریب)	سطح معنی داری
-۱/۳۷	۰/۲۵	۰/۰۲*
-۰/۳۴	۰/۷۱	۰/۰۳*
-۰/۸۵	۰/۴۲	۰/۰۹
-۰/۰۰۱	۰/۹۹	۰/۸۶

*معنی دار در سطح ۵ درصد

جدول ۳: مقادیر حساسیت، ویژگی، دقت و سطح زیر منحنی مشخصه عملکرد برای دو مدل

سطح زیر منحنی راک	دقت	ویژگی	حساسیت
۰/۶۶۵	۰/۶۶	۰/۷۴	۰/۳۷
۰/۶۸۶	۰/۷۱	۰/۷۵	۰/۵۷



نمودار ۱: نمودار درخت تصمیم برای طبقه بندی افراد به دو گروه BMD نرمال و BMD پایین

بحث و نتیجه گیری

مطالعات زیادی با روش های مختلف عوامل موثر بر پوکی استخوان را مورد بررسی قرار دادند که متغیرهای مورد استفاده در این مطالعه نیز با این مطالعات منطبق است (۱۷،۱۸). در این مطالعه متغیرهای تعداد حاملگی، شاخص توده بدنی و میزان کلسیم در مدل درختی به عنوان متغیرهای موثر معرفی شدند. متغیر تعداد حاملگی در هر دو مدل به عنوان یک متغیر مهم شناخته شد. در مطالعه باقرلاریجانی و همکاران (۲۰۰۷) که عوامل خطر پوکی استخوان را با رگرسیون لجستیک چندگانه در زنان بررسی کردند عواملی چون BMI و تعداد حاملگی را به عنوان عوامل پوکی استخوان عنوان کردند که با این مطالعه همسو است (۱۹). چپو و همکاران (۲۰۰۶) عوامل خطر پوکی استخوان را با مدل شبکه عصبی به دست آورده اند (۴) و عوامل سن، قد، وزن، BMI، سن منوپوز و مصرف قهوه را از جمله

در این مطالعه به وسیله مدل طبقه بندی درخت تصمیم، مدلی ارائه شد که با استفاده از متغیرهایی که اندازه گیری های آنها در هر شرایطی وجود دارد می توان به پیش بینی خطر پوکی استخوان برای زنان رسید. همچنین مدل رگرسیون لجستیک برای مقایسه با مدل درختی به کار گرفته شد. بررسی فقط بر زنان بین ۲۰ تا ۴۰ سال انجام شده و برخلاف اکثر مطالعات که طبقه بندی افراد را براساس استئوپروز ($t\text{score} \leq -2/5$) در نظر گرفته اند در این مطالعه افراد را براساس استئوپنی به دو گروه با BMD نرمال ($t\text{score} \geq -1$) و افراد با BMD پایین ($t\text{score} < -1$) تقسیم کردیم البته در مطالعه ای که چانگ (۲۰۱۳) و همکاران (۱۶) انجام داده اند از این تقسیم بندی استفاده کرده اند.



و رگرسیون لوجستیک ۹۲/۲ درصد نتیجه گرفتند که مدل رگرسیون لوجستیک عملکرد تشخیصی بهتری دارد.

به طور کلی نتایج به دست آمده از مقایسه دو مدل درخت تصمیم و رگرسیون لوجستیک در مطالعات انجام شده یکسان نیست و به طور واقع برتری یک مدل را نشان نمی دهد.

از آن جا که الگوی معرفی شده درخت تصمیم دارای مزایایی چون تفسیر آن برای افراد غیر آماری بسیار ساده است، مدل ها تحت تغییر فضای مشاهدات تغییر ناپذیر هستند، درخت تصمیم به داده های گمشده حساسیت کمتری دارد و نسبت به سایر مدل های طبقه بندی منعطف تر است و به راحتی با یک نگاه به درخت متغیرهای مهم به کار رفته در مدل شناسایی می شود. از سوی دیگر نیازی به بررسی برقراری فرضیات بنیادی نیست بهتر است دو مدل رگرسیون لوجستیک و درخت تصمیم به طور همزمان به کار گرفته شود.

قابل ذکر است که هدف این مطالعه بررسی تک تک متغیرهای مدل لوجستیک نبوده و این مدل صرفاً برای مقایسه با کارایی مدل درختی به کار گرفته شد.

تضاد منافع

نویسندگان این مقاله اعلام می دارند که هیچ گونه تضاد منافی وجود ندارد.

تشکر و قدردانی

نویسندگان از مرکز تحقیقات دیابت یزد به خاطر در اختیار گذاشتن داده ها کمال تشکر و قدردانی را دارند. این مقاله حاصل پایان نامه دانشجویی کارشناسی ارشد دانشگاه علوم پزشکی شهید صدوقی یزد می باشد.

رسیک فاکتورهای پوکی استخوان برشمرند که در این میان BMI با این مطالعه هم خوانی دارد.

مودانی و همکاران (۲۰۱۱) روش جنگل درخت تصمیم را برای شناسایی موارد پوکی استخوان به کار بردند (۱۴) که با دقت ۹۹/۹۲ درصد این مدل دو متغیر BMI و سن را از مهم ترین عوامل مل در پوکی استخوان شناخته است. در مطالعه یوسل و همکاران (۲۰۱۴) که عوامل خطر استئوپروز زنان را با مدل رگرسیون لوجستیک مورد بررسی قرار دادند (۲۰) عواملی چون سن، وزن، میزان کلسیم، ورزش، سن منوپوز و مصرف سیگار را به عنوان عوامل خطر معرفی کرد که از این میان متغیر کلسیم با این مطالعه یکسان است.

در این بررسی نزدیکی مقادیر سطح زیر منحنی راک، ۶۶/۵ درصد برای مدل درختی و ۶۸/۶ درصد برای مدل رگرسیون لوجستیک برتری نسبی مدل رگرسیون لوجستیک را نشان می دهد.

اناجسکو (۲۰۰۷) از چندین روش طبقه بندی از جمله درخت تصمیم و رگرسیون برای ارزیابی پوکی استخوان زنان استفاده کرد (۸)، از مدل درخت تصمیم تعداد حاملگی و BMI به عنوان عوامل خطر پوکی استخوان به دست آمده و بالاتر بودن مقادیر حساسیت و ویژگی مدل درختی در مقایسه بارگرسیون لوجستیک برتری مدل درختی در این مطالعه را نشان داد که این برخلاف این مطالعه است.

یاداوی (۲۰۱۳) و همکاران درخت تصمیم و رگرسیون لوجستیک را برای پیش بینی بیماری قلبی به کار بردند (۲۱) با استفاده سطح زیر منحنی راک برای درخت تصمیم ۸۶/۶ درصد



References

- 1-Goodarzi H,Larijani B,Keshtkar A ,Khashayar P. Prevalance and Assosiate Factors of T-Score Discordance Between Lumbar Spine and Femoral Neck in Postmenopausal Women: A 11-Year Irostepos Study. Iranian journal of Diabetes and Metabolism.2014;13(2):234-44.[Persian]
- 2-White SC, Taguchi A, Kao D,Wu S,Yoon D, Swei Y. Clinical and panoramic predictors of femur bone mineral density. Osteoporosis international.2005;16(3):339-46.
- 3-Lee J,Hwang Y,Park S,Kim S.Diagnosis of Osteoporosis by Quantification of Trabecular Microarchitectures from Hip Radiographs Using Artificial Neural Networks.Bio-Inspired Computing-Theories and Applications: Springer.2014: 247-50.
- 4-Chiu J,Li Y,Yu F,Wang Y.Applying an artificial neural network to predict osteoporosis in the elderly. Studies in health technology and informatics.2006;124:609.
- 5-Johnson R,Wichern D,Education P.Applied multivariate statistical analysis: Prentice hall Englewood Cliffs, NJ.1992:644-46
- 6- Sedehi M, Kazemnezhad A, Majd V. Design of artifical neural networks to model the response of two mixed-variable and its application to medical data Journal of Epidemiology IRAN. 2011;6(6):28-39.[Persian]
- 7-Lewis R. An introduction to classification and regression tree (CART) analysis. Annual Meeting of the Society for Academic Emergency Medicine in San Francisco, California.2000.
- 8-Enachescu C.Evaluation of some statistical methods for referring women for bone densitometry. Biocybernetics and Biomedical Engineering.2007;27(3):43.
- 9-Reby D,Lek S,Dimopoulos I,Joachim J,Lauga J,Aulagnier S.Artificial neural networks as a classification method in the behavioural sciences. Behavioural processes.1997;40(1):35-43.
- 10-Kemp F. Modern Applied Statistics with S. Journal of the Royal Statistical Society: Series D The Statistician.2003;52(4):704-5.
- 11-Siotani T,Fujikoshi Y,Hayakawa T.Modern multivariate statistical analysis, a graduate course and handbook: American Sciences Press.1985;281-90.
- 12-Allali F Rostom S,Bennani L,Abouqal R,Hajjaj-Hassouni N.Educational level and osteoporosis risk in postmenopausal Moroccan women: a classification tree analysis. Clinical rheumatology. 2010;29(11):1269-75



- 13- Patil N, Lathi R, Chitre V. Comparison of C5.0 & CART classification algorithms using pruning technique. *International Journal of Engineering Research and Technology*. 2012;1(4):1-5.
- 14- Moudani W, Shahin A, Chakik F, Rajab D. Intelligent Predictive Osteoporosis System. *International Journal of Computer Applications*. 2011;32(5):28-37.
- 15- Ferri C, Flach P, Hernández-Orallo J. Learning decision trees using the area under the ROC curve. *nICML*. 2002:139-146.
- 16- Chang H, Chiu Y, Kao H, Yang C, Ho W. Comparison of classification algorithms with wrapper-based feature selection for predicting osteoporosis outcome based on genetic factors in a taiwanese women population. *International journal of endocrinology*. 2013;2013:8.
- 17- Napoli N, Schwartz AV, Palermo L, Jin JJ, Wustrack R, Cauley JA, et al. Risk factors for subtrochanteric and diaphyseal fractures: the study of osteoporotic fractures. *The Journal of Clinical Endocrinology & Metabolism*. 2013;98(2):659-67.
- 18- Suzuki T. Risk factors for osteoporosis in Asia. *Journal of bone and mineral metabolism*. 2001;19(3):133-41.
- 19- Keramat A, Larijani B, Adibi H, Hosseinijad A, Chopra A, Patwardhan B. Risk factors for osteoporosis in postmenopausal women in urban. *Journal of Science and Health*. 2007;2(3):36-41. [Persian]
- 20- Yücel AS, Korkmaz M. Studying the Factors Affecting Osteoporosis in Women with the Logistic Regression Analysis. *Middle-East Journal of Scientific Research*. 2014;20(8):893-9.
- 21- Yadav P, Jaiswal K, Patel S, Shukla D. Intelligent Heart Disease Prediction Model Using Classification Algorithms. *IJCSMC*. 2013;3(08):102-7