



کاربرد الگوریتم های درخت تصمیم و رگرسیون لجستیک برای پیش بینی ابتلا به لنفادم در بیماران سرطان پستان

ملیحه فاضلی: کارشناسی ارشد مدیریت صنعتی، دانشکده مدیریت، دانشگاه تهران، تهران، ایران

عالیه کاظمی: دانشیار مدیریت صنعتی، دانشکده مدیریت، دانشگاه تهران، تهران، ایران (* نویسنده مسئول) aliyehkazemi@ut.ac.ir

شهرحقیقت: دانشیار پژوهشی، گروه پژوهشی کیفیت زندگی، مرکز تحقیقات سرطان پستان، پژوهشکده معتمد، جهاد دانشگاهی، تهران، ایران

چکیده

کلیدواژه‌ها

داده کاوی،
پیش بینی،
لنفادم،
درخت تصمیم،
رگرسیون لجستیک

زمینه و هدف: لنفادم از عوارض شایع و ناتوان کننده بیماران سرطان پستان است. این مطالعه به پیش بینی و دسته بندی عارضه لنفادم پرداخته است. همچنین شناسایی عوامل مؤثر و کشف الگوهایی برای تشخیص سریع تر و جلوگیری از بروز این عارضه از اهداف دیگر این پژوهش است.

روش کار: اطلاعات حاصل از پرونده های ۱۱۱۳ بیمار مبتلا به سرطان پستان که طی سال های ۱۳۸۸ تا ۱۳۹۶ به کلینیک لنفادم سیدخندان مراجعه کرده بودند، مورد ارزیابی قرار گرفت. تحلیل داده ها با به کارگیری روش شناسایی CRISP-DM و در نرم افزار IBM SPSS Modeler 18 انجام شد و در بخش مدل سازی، الگوریتم های درخت تصمیم و رگرسیون لجستیک استفاده شد.

یافته ها: اطلاعات ۹۳۳ بیمار شامل ۲۵ متغیر پس از پیش پردازش داده ها وارد مدل شدند. احتمال ابتلا به لنفادم برای هر یک از بیماران توسط الگوریتم های درخت تصمیم C5.0، Chaid، C&RT، Quest و الگوریتم رگرسیون لجستیک به ترتیب با حساسیت ۷۹/۳۳٪، ۷۴/۴۱٪، ۷۱/۹۲٪، ۷۲/۶۴٪ و ۷۷/۸۳٪ پیش بینی شد و عوامل خطر لنفادم شناسایی شدند. نسبت تعداد غدد لنفاوی درگیر به تعداد غدد لنفاوی خارج شده، احساس سنگینی، نوع عمل جراحی، مرحله بیماری، سن، شاخص توده بدنی، متاستاز، تعداد دوره های شیمی درمانی، بیماری همراه و تعداد غدد لنفاوی خارج شده به ترتیب مؤثرترین عوامل خطر در پیش بینی ابتلا به لنفادم هستند که توسط الگوریتم درخت تصمیم C5.0 شناسایی شدند.

نتیجه گیری: یافته ها نشان می دهد الگوریتم درخت تصمیم C5.0 با بالاترین حساسیت، مدل برتر برای پیش بینی ابتلا به لنفادم است. با بکارگیری قوانین ایجاد شده برای یک نمونه جدید با ویژگی های مشخص، می توان پیش بینی کرد بیمار احتمالاً مبتلا به لنفادم خواهد شد یا خیر. با در نظر گرفتن شاخص توده بدنی به عنوان یک عامل قابل تغییر، رژیم های کنترل وزن برای این بیماران توصیه می شود. همچنین توجه به احساس سنگینی بیمار در مراحل اولیه ضروری است.

تعارض منافع: گزارش نشده است.

منبع حمایت کننده: حامی مالی نداشته است.

شیوه استناد به این مقاله:

Fazeli M, Kazemi A, Haghighat Sh. Application of decision tree and logistic regression algorithms to predict lymphedema in breast cancer patients. Razi J Med Sci. 2019;25(12):84-95.

*انتشار این مقاله به صورت دسترسی آزاد مطابق با [CC BY-NC-SA 1.0](https://creativecommons.org/licenses/by-nc-sa/4.0/) صورت گرفته است.



Original Article

Application of decision tree and logistic regression algorithms to predict lymphedema in breast cancer patients

Malihe Fazeli, MSc of Industrial Management, Faculty of Management, University of Tehran, Tehran, Iran
***Aliyeh Kazemi**, Associate Professor of Industrial Management, Faculty of Management, University of Tehran, Tehran, Iran (* Corresponding author) aliyehkazemi@ut.ac.ir
Shahpar Haghghat, Research Associate Professor, Quality of Life Department, Breast Cancer Research Center, Motamed Cancer Institute, ACECR, Tehran, Iran

Abstract

Background: Lymphadenitis is a common and debilitating complication of breast cancer patients. This study has predicted and classified lymphedema complications. Moreover, identifying effective factors and discovering patterns for faster diagnosis and prevention of this complication is another goal of this study.

Methods: Data from 1113 patients with breast cancer who were referred to Seyed Khandan Lymphedema clinic during 2009 to 2017 were evaluated. Data analysis was performed by using IBM SPSS Modeler software version 18 and CRISP-DM methodology and in the modeling section, logistic regression and decision tree algorithms were used.

Results: Data from 933 patients including 25 variables were entered into the model after pre-processing. Probability of catching of Lymphedema for each patient predicted by logistic regression algorithm and different decision tree algorithms consist of C5.0, Chaid, C&RT, and Quest with the sensitivity of 79.33%, 74.41%, 71.92%, 72.64% and 77.83%, respectively and finally the factors related to Lymphedema were identified. Ratio of the involved lymph node numbers to the removed lymph node numbers, heaviness, type of surgery, stage of disease, age, body mass index, metastasis, number of chemotherapy courses, comorbidity, number of removed lymph nodes respectively.

Conclusion: The results show that C5.0 decision tree algorithm with the highest sensitivity is the best model for predicting Lymphedema. By applying the rules created for a new sample with specific characteristics, it can be predicted that the patient will probably suffer from Lymphedema or not. Considering that BMI is a changeable factor, weight control regimens are recommended for these patients. In addition, it is necessary to pay attention to the patient's heavy feeling in the early stages.

Conflicts of interest: None

Funding: None

Keywords

Data mining,
prediction,
Lymphedema,
Decision tree,
Logistic regression

Received: 21/10/2018

Accepted: 13/01/2019

Cite this article as:

Fazeli M, Kazemi A, Haghghat Sh. Application of decision tree and logistic regression algorithms to predict lymphedema in breast cancer patients. Razi J Med Sci. 2019;25(12):84-95.

This work is published under [CC BY-NC-SA 1.0 licence](https://creativecommons.org/licenses/by-nc-sa/4.0/).

در مطالعه کن و همکاران سن، سطح تحصیلات پایین، چاقی، اندازه تومور (Tumor size)، تعداد غدد لنفاوی درگیر و پرتودرمانی با تجزیه و تحلیل رگرسیون لجستیک به عنوان مهم‌ترین عوامل خطر لنفادم شناسایی شدند (۱۹). ماستکتومی (Mastectomy)، میزان نمونه‌برداری از زیر بغل، پرتودرمانی و حضور گره‌های لنفاوی مثبت، در یک مطالعه فراتحلیل (Study of meta-analysis) به عنوان عوامل خطر ابتلا به لنفادم بازو پس از درمان سرطان پستان شناخته شد (۲۰). نتایج مدل رگرسیون لجستیک چندمتغیره برای شناسایی عوامل خطر (Risk factors) نشان داد سن در زمان تشخیص، سابقه فشارخون، شاخص توده بدنی بیش‌تر از ۳۰ و عمل جراحی که در آن ۱۰ یا تعداد بیش‌تری گره لنفاوی خارج شده باشد، با بروز لنفادم در ارتباط است (۲۱).

در مطالعه صورت گرفته توسط لی و همکاران، شاخص توده بدنی بالا، شیمی‌درمانی، نوع جراحی زیربغل و پرتودرمانی ارتباط مستقیمی با بروز لنفادم دارند (۲۲).

هیچ یک از درمان‌ها و روش‌های دارویی و غیر دارویی موجود موجب درمان قطعی لنفادم نمی‌گردند و فقط شدت آن را کاهش می‌دهند (۱۷، ۲۰، ۲۳). در تمامی مطالعات مورد بررسی، از تحلیل‌های آماری برای شناخت مهم‌ترین عوامل استفاده شده است و مطالعه ای که از روش‌های داده‌کاوی (Data mining) برای پیش‌بینی لنفادم و شناسایی عوامل خطر استفاده کند انجام نشده است؛ لذا این پژوهش با به کارگیری روش‌های داده‌کاوی، احتمال ابتلا به لنفادم را برای هر بیمار تحت درمان سرطان پستان پیش‌بینی و همچنین مهم‌ترین عوامل خطر ابتلا به لنفادم را شناسایی می‌کند تا از این طریق بتوان اقدامات پیش‌گیرانه را برای بیماران انجام داد.

روش کار

در این پژوهش از روش‌شناسی Cross (CRISP-DM) که Industry Standard Process for Data Mining

لنفادم یک عارضه شایع و مزمن درمان سرطان پستان می‌باشد که سبب بروز اختلالات مختلف و کاهش کیفیت زندگی بیمار می‌شود (۳-۱). لنفادم به دنبال سرطان پستان ناشی از انباشت غیر طبیعی مایع غنی از پروتئین در فضای بینابینی به علت تخلیه لنفاوی نامناسب است (۴ و ۵) که نرخ بروز آن در مطالعات مختلف حدود ۳ تا ۸۷ درصد است (۹-۶).

مطالعات نشان می‌دهد، درمان‌های متفاوتی نظیر عمل جراحی (Surgery) (۱۰ و ۱۱)، پرتودرمانی (Radiotherapy)، شیمی‌درمانی (Chemotherapy) (۱۲ و ۱۳) و هورمون درمانی (Hormonal therapy) (۱۴ و ۱۵) و عوامل خطر دیگر نظیر چاقی (Obesity)، سن در زمان تشخیص (۱۲ و ۱۳)، وجود بیماری همراه و کاهش فعالیت (۱۰)، سبب اشکال در آناتومی (Anatomy) و در نتیجه عملکرد سیستم لنفاوی (Lymphatic system) شده و فرد را در معرض خطر ابتلا به لنفادم (Lymphedema) قرار می‌دهد (۱۴ و ۱۶).

آنالیز تک‌متغیری و چندمتغیری با به کارگیری مدل رگرسیون لجستیک (Logistic Regression) بر روی ۱۷۰ بیمار نشان داد چاقی بر اساس شاخص توده بدنی (Body Mass Index) به صورت معناداری بر ایجاد لنفادم مؤثر بود و میانگین غدد لنفاوی درگیر در بیماران واجد و فاقد لنفادم تفاوت معنی‌داری داشت (۱۷).

حجم توده بدنی بالا، تعداد غدد لنفاوی درگیر و مدت زمانی که از جراحی می‌گذشت در آنالیز لجستیک چند متغیره ۴۱۰ بیمار سرطان سینه از سه مرکز در تهران و مشهد که در ۱۲۳ مورد آن‌ها لنفادم توسعه یافته است، رابطه معنی‌داری با بروز لنفادم نشان دادند (۱۸). حداد و همکاران نیز با به کارگیری آزمون‌های آماری (آزمون t-test) و رگرسیون لجستیک فاصله زمانی بعد از عمل، پرتودرمانی و تکنیک مورد استفاده برای درمان، از جمله عوامل مؤثر بروز لنفادم شناخته شد (۱).

با توجه به این که وجود داده‌های مفقوده از قدرت آماری آزمون‌های آماری می‌کاهد و خطای اندازه‌گیری را افزایش می‌دهد باید این داده‌ها شناسایی و مدیریت شوند. همچنین متغیرهایی که دارای مقادیر مفقوده هستند، شناسایی شدند و با به کارگیری الگوریتم Classification and Regression Tree (C&RT)، ابتدا داده‌های مفقوده برای متغیرهای کم اهمیت و پس از آن برای متغیرهای با اهمیت، جایگزین شدند.

متغیر سن در زمان تشخیص از اختلاف بین دو تاریخ تشخیص بیماری و تاریخ تولد به دست آمد. با توجه به توزیع متغیر سن در مجموعه داده‌ها، این متغیر نیز به بازه‌هایی در کلاس‌های کم‌تر از ۴۰ سال، ۴۰ تا ۵۰ سال، ۵۰ تا ۶۰ سال و بالای ۶۰ سال تفکیک شد.

همچنین متغیرهای کد، نام خانوادگی، تاریخ تولد، تاریخ شروع و پایان شیمی‌درمانی، تاریخ پایان پرتودرمانی، تاریخ جراحی، تاریخ مراجعه و تاریخ تشخیص بیماری به علت بی‌تاثیربودن در خروجی از مجموعه ورودی‌ها حذف شدند.

شاخص BMI ابزاری مناسب است تا سلامت وزن فرد با توجه به قدش تخمین زده شود. بنابراین با استفاده از دو متغیر قد و وزن بیماران، این متغیر جدید به مجموعه داده‌ها اضافه شد.

برای متغیر شاخص توده بدنی می‌توان داده‌ها را جهت تفسیرپذیری بهتر به طبقه‌های کمبود وزن و عادی ($BMI \leq 25$)، اضافه وزن ($25 < BMI < 30$) و چاق ($BMI > 30$) تبدیل کرد.

استفاده از شاخص‌های نسبی به عنوان ورودی مدل می‌تواند کیفیت مدل را بالاتر ببرد؛ بنابراین از نسبت متغیر تعداد گره‌های لنفاوی درگیر به تعداد گره‌های لنفاوی خارج‌شده متغیر جدیدی وارد مجموعه داده‌ها شد.

با بررسی مطالعات انجام شده و با تایید پزشک متخصص کلینیک، بیمارانی که حجم تورم آن‌ها بیش‌تر از ۲۰۰ سی‌سی است، دچار عارضه لنفادم شدند که کلاس یک نشان‌دهنده بیمارانی است که حجم تورم آن‌ها بالای ۲۰۰ سی‌سی است و کلاس صفر نشان‌دهنده بیمارانی است که حجم تورم آن‌ها پایین‌تر از ۲۰۰ سی‌سی است (۲۵).

یک روش استاندارد داده‌کاوی است، برای انجام فرآیند تحقیق استفاده شده است (۲۴).

مجموعه داده‌ها از اطلاعات بیماران مراجعه‌کننده به کلینیک لنفادم سیدخندان طی سال‌های ۱۳۸۸ تا ۱۳۹۶ جمع‌آوری شده است که این مجموعه داده‌ها شامل ۳۶ متغیر برای ۱۱۱۳ بیمار است. اطلاعات بیماران شامل کد، نام خانوادگی، تاریخ تولد، تاریخ شروع و پایان شیمی‌درمانی، تاریخ پایان پرتودرمانی، تاریخ جراحی، تاریخ مراجعه و تاریخ تشخیص بیماری، قد، وزن، شغل، تحصیلات، وضعیت تاهل، دست غالب و درگیر، مرحله بیماری، محل تومور، اندازه تومور، وجود یا عدم وجود متاستاز و بیماری همراه، تعداد غدد لنفاوی درگیر و برداشته شده، انجام یا عدم انجام شیمی‌درمانی، پرتودرمانی و هورمون‌درمانی، تعداد دوره‌های شیمی‌درمانی، نوع رژیم شیمی‌درمانی، دامنه حرکت اندام درگیر، میزان فعالیت فیزیکی، درد، احساس سنگینی و احساس سوزن سوزن شدن، نوع عمل جراحی و نمونه‌برداری است که به عنوان ورودی مدل مورد استفاده قرار گرفته است.

با بررسی تاریخ مراجعه بیماران به کلینیک، تاریخ عمل جراحی و تاریخ دوره‌های شیمی‌درمانی و پرتودرمانی، مقطع پیش‌بینی برای بیماران، پایان دوره درمان اولیه در نظر گرفته شد. با توجه به اینکه طول دوره درمان اولیه بیماران بعد از تاریخ جراحی حدود ۷ ماه در نظر گرفته شده است، بنابراین بیمارانی که فاصله تاریخ تشخیص تا تاریخ مراجعه آن‌ها کم‌تر از ۷ ماه است، به غیر از بیمارانی که شرایط زیر را دارند، از مطالعه حذف گردیدند.

• بیمارانی که هیچ یک از دوره‌های درمانی برای آن‌ها تجویز نشده باشد.

• بیمارانی که دوره‌های درمانی را پیش از تاریخ تشخیص گذرانده باشند.

• بیمارانی که از تاریخ اتمام شیمی‌درمانی آن‌ها بیش از یک ماه گذشته باشد؛ تا بتوان برای داده‌هایی که مقدار پرتودرمانی آن‌ها مفقود است، دوره پرتودرمانی را در صورت لزوم در نظر گرفت.

همچنین با بررسی مجموعه داده‌ها تعداد ۷ داده تکراری شناسایی شد. در مجموع ۱۴۳ داده از مجموعه داده‌ها حذف گردیدند.

که کلاس یک نشان دهنده بیماری است که دست غالب و درگیر آن‌ها هم‌جهت می‌باشد و کلاس دو نشان دهنده بیماری است که دست غالب و درگیر آن‌ها در خلاف جهت هم باشند.

در برخی از متغیرها تعداد داده‌ها برای برخی از گروه‌ها به نسبت کم می‌باشد که این موضوع سبب پراکندگی و عدم تعمیم‌پذیری قوانین به دست آمده از مدل‌ها خواهد شد. بنابراین گروهی از داده‌های یک متغیر که امکان ترکیب با گروه دیگر را داشتند، با یکدیگر ادغام شدند.

به منظور تشخیص نقاط پرت با به کارگیری مدل شناسایی نقاط پرت (Anomaly Detection) با تنظیمات پیش فرض در نرم افزار IBM SPSS Modeler 18.0 هر رکورد با رکوردهای دیگر مقایسه شد. بر اساس این مدل مواردی با مقدار شاخص بزرگ‌تر از دو می‌تواند کاندیدهای خوبی برای مغایرت باشند. در این مطالعه نیز متغیرهای کمی به عنوان ورودی به این مدل وارد و پس از اجرای مدل داده‌های پرت شناسایی شدند. شاخص Anomaly برای ۳۷ رکورد بزرگ‌تر از ۲ بود. بنابراین این رکوردها از مجموعه داده‌ها حذف شدند (شکل ۱).

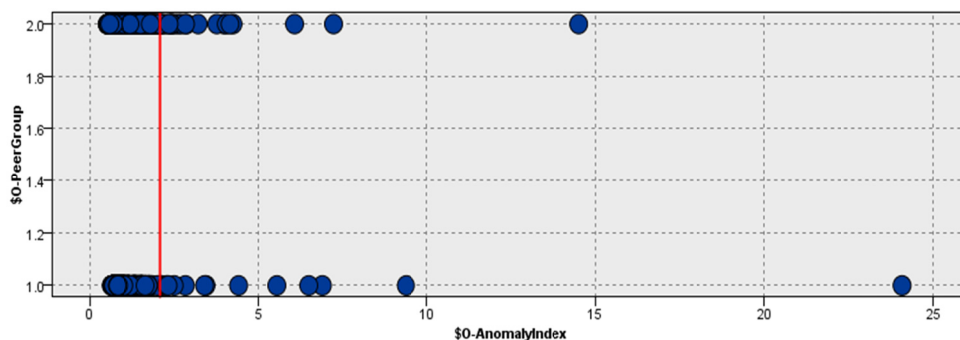
برای متغیر وضعیت تاهل، تعداد کمی از داده‌ها در کلاس بیوه و مطلقه قرار دارند، بنابراین به جهت از بین بردن پراکندگی و نویز، این دو کلاس با کلاس افراد مجرد ادغام شد و با عنوان کلاس مجرد وارد مدل شدند.

در متغیر نمونه‌برداری تعداد مشاهدات در برخی از روش‌ها بسیار کم است. به همین علت روش‌های مذکور در دو دسته Invasive ductal Ca (IDC) و others خلاصه شد.

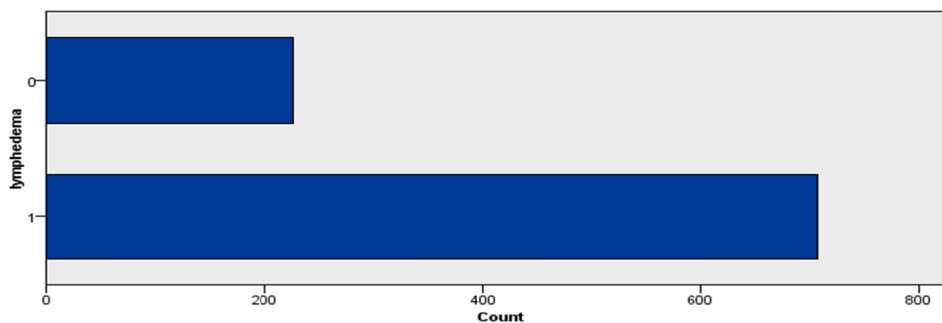
افرادی که دارای تحصیلات پزشکی و غیرپزشکی بودند، در یک گروه و با عنوان تحصیلات دانشگاهی ادغام شدند.

در مورد متغیر نوع رژیم شیمی‌درمانی افرادی که رژیم Tax، Ac*4&Tax*4 و EC&TAX را دریافت کردند، با توجه به اینکه همگی به نوعی در رژیم تاکسوتر جای دارند و جنس ادم را تغییر می‌دهند با هم ترکیب شدند و باقی رژیم‌ها نیز در یک کلاس قرار داده شدند. همچنین افرادی که هیچ رژیم شیمی‌درمانی‌ای دریافت نکردند نیز در یک کلاس قرار گرفتند.

علاوه بر این، از ترکیب دو متغیر دست غالب و دست درگیر، متغیر جدیدی با دو کلاس یک و دو ساخته شد



شکل ۱- تشخیص رکوردهای پرت



شکل ۲- توزیع داده‌های اصلی برای متغیر هدف

مدل برگزیده شد. در این مدل، حساسیت به هرس کردن ۷۰ درصد در نظر گرفته شد و حداقل تعداد رکوردها برای هر زیرشاخه از درخت ۲۰ رکورد تعیین شد. با توجه به اینکه تقسیم‌بندی داده‌ها به داده‌های آزمون و آموزش بر اساس یک عدد تصادفی انجام می‌پذیرد؛ مدل C5.0 برگزیده شده با تغییر عدد تصادفی مربوط به تقسیم‌بندی داده‌ها، چندین بار اجرا شد و برترین مدل برای تحلیل انتخاب گردید.

در الگوریتم Chaid داده‌های آموزش در جریان تعریف شده به عنوان ورودی به گره Chaid وارد شد و پارامترهای مدل به گونه‌ای تنظیم شد تا بهترین نتایج حاصل شود. مدل Chaid با ۶ لایه و مقدار هزینه ۱/۳ به عنوان هزینه خطای سیستم در تشخیص نادرست فرد مبتلا به لنفادم به عنوان فرد مبتلا نشده برگزیده شد. با تغییر عدد تصادفی تقسیم‌بندی داده‌ها در چندین اجرا، برترین مدل جهت تحلیل انتخاب شد.

علاوه بر درخت‌های تصمیم C5.0 و Chaid، مدل C&RT نیز بر روی داده‌های آموزش اجرا گردید و برای دستیابی به مدلی با بالاترین حساسیت و دقت، پارامترهای مدل تنظیم شد. مدل C&RT با ۶ لایه و مقدار هزینه ۱/۳۴ به عنوان هزینه خطای سیستم در تشخیص نادرست فرد مبتلا به لنفادم به عنوان فرد مبتلا نشده در نظر گرفته شد. در نهایت با تغییر عدد تصادفی تقسیم بندی داده‌ها، مدل بارها اجرا شد و مدل با بالاترین حساسیت انتخاب گردید.

از مجموعه درختان تصمیم، مدل دیگری که بر روی داده‌های آموزش اجرا گردید، مدل Quest است. مدل با پارامترهای مختلف و تغییر در عدد تصادفی تقسیم بندی داده‌ها، بارها اجرا گردید و در نهایت مدل با شش لایه و هزینه ۱/۶ برای خطای سیستم در تشخیص نادرست فرد مبتلا به لنفادم به عنوان فرد مبتلا نشده اجرا شد. و مدلی که توانست بیش‌ترین حساسیت را ایجاد کند، جهت تحلیل انتخاب شد.

معمولاً مجموعه‌های قوانین، مهم‌ترین اطلاعات مرتبط با درخت تصمیم را در بردارند. در گزارش مدل تولید شده توسط مدل درخت تصمیم، درخت به زیرشاخه‌هایی شکسته می‌شود و این روند تناوبی ادامه می‌یابد تا در نهایت داده‌های هر گره در یک دسته قرار داشته باشند. برای بیان قوانین استخراج شده، مسیر

در این مجموعه داده، داده‌های مربوط به کلاس صفر یا همان کلاس بیماری که به لنفادم مبتلا نشده‌اند حدود ۲۵ درصد از کل داده‌ها (۲۲۶ نفر) و کلاس یک یا همان کلاس افراد مبتلا به لنفادم حدود ۷۵ درصد از کل داده‌ها (۷۰۷ نفر) است. با توجه به نامتوازن بودن داده‌ها از روش سطح داده (Data Level) شامل بیش نمونه‌برداری کلاس اقلیت (Random minority oversampling) استفاده شد. به این صورت که از کلاس اقلیت (بیماری که به لنفادم مبتلا نشدند) به صورت تصادفی داده‌هایی انتخاب و به مجموعه داده اضافه شد تا اندازه داده‌های کلاس اقلیت با اندازه داده‌های کلاس اکثریت (بیماران مبتلا به لنفادم) برابر شود (شکل ۲).

یافته‌ها

در نهایت، پس از پیش‌پردازش و انتخاب مشخصه‌ها، ۹۳۳ برای دو گروه افراد مبتلا به لنفادم با فراوانی ۷۰۷ نفر (۷۵/۷۸ درصد) و افراد غیرمبتلا به لنفادم با فراوانی ۲۲۶ نفر (۲۴/۲۲ درصد) شامل ۲۵ متغیر مستقل به عنوان ورودی مدل برای متغیر وابسته لنفادم به عنوان خروجی مدل انتخاب شدند.

طبقه‌بندی (Classification) و پیش‌بینی (Prediction) دو نوع عملیات برای تحلیل داده‌ها و استخراج مدل به منظور توصیف دسته‌های مهم داده‌ها، فهم و پیش‌بینی رفتار آینده آن‌ها می‌باشند. بنابراین داده‌ها را می‌توان به دو قسمت تقسیم کرد: اول آن دسته که مدل بر اساس آن‌ها ساخته می‌شود و دوم گروهی که برای ارزیابی مدل استفاده می‌شود. داده‌های گروه دوم که دسته آن‌ها مشخص است را به مدل داده و خروجی مدل را با دسته‌های مشخص توسط مدل مقایسه کرده و دقت کل مدل استخراج می‌شود (۲۶). بدین منظور، ۷۰ درصد داده‌ها به صورت کاملاً تصادفی به عنوان داده‌های آموزش انتخاب گردید و مدل‌سازی بر روی این داده‌ها انجام گرفت و در نهایت پس از اجرای مدل، داده‌های باقی مانده به عنوان داده‌های آزمون برای ارزیابی مدل استفاده شد.

برای آموزش درخت تصمیم C5.0 متغیرهای پژوهش در جریان تعریف شده به عنوان ورودی به گره C5.0 وارد شد. پس از چندین بار اجرای مدل C5.0، بهترین

پیش‌گو در صورت معنی‌دار بودن به ترتیب وارد مدل می‌شوند و متغیرهایی که با متغیر هدف ارتباطی نداشته باشند، از تحلیل حذف می‌شوند و در آخرین مرحله متغیرهای موثر بر متغیر هدف باقی می‌مانند. نکته دیگر اینکه برای متغیرهای پیش‌گو، کلاس اول از هر متغیر، کلاس اصلی برای مقایسه با سایر کلاس‌های همان متغیر در نظر گرفته شد.

از میان متغیرهای باقی مانده، چنانچه مقادیر هر متغیر در سطح خطای کوچکتر از $0/05$ معنی‌دار باشد، در آن صورت نتیجه گرفته می‌شود که وجود آن متغیر در مدل مفید و اثر آن معنی‌دار است و می‌توان نسبت بخت‌ها را برای آن تحلیل کرد. از طرفی در میان مجموعه متغیرهای معنی‌دار باقیمانده در مدل، متغیر متاستاز بیش‌ترین توانایی را در پیش‌بینی متغیر پاسخ لنفادم دارد (جدول ۲).

میزان صحت یک روش بر روی مجموعه داده‌های آموزشی، درصد مشاهداتی از مجموعه آموزشی است که

ریشه تا برگ پیمایش می‌شود و قوانین به صورت شرطی بیان می‌شود. معیار اطمینان (Confidence) قوانین ایجاد شده برای کلاس یک (بیماران مبتلا به لنفادم) 75 درصد و برای کلاس صفر (بیماران مبتلا نشده به لنفادم) 25 درصد تعیین شد.

با توجه به گستردگی تعداد برگ‌ها و اندازه درخت‌های تصمیم، از آوردن نمودار فرم درختی و ذکر کامل فرم اگر آنگاه (If-Then) خودداری شده است و تنها مهم‌ترین مجموعه قوانین ایجاد شده توسط الگوریتم‌های درخت که معیار اطمینان برای کلاس یک آن‌ها بیش‌تر از 75 درصد و برای کلاس صفر آن بیش‌تر از 25 درصد است آورده شده است (جدول ۱).

همچنین مدل رگرسیون لجستیک برای دسته بندی داده‌ها بر اساس ورودی‌های تعیین شده در مرحله قبل جهت پیش‌بینی متغیر وابسته لنفادم استفاده شد. بر این اساس با به کارگیری روش گام‌به‌گام پیش‌رو (Forwards Stepwise) در هر مرحله متغیرهای

جدول ۱- مهم‌ترین قوانین ایجاد شده توسط الگوریتم‌های درخت تصمیم

ردیف	قوانین	اطمینان (درصد)
۱	اگر بیمار متاستاز نداشته باشد، نسبت تعداد غدد لنفاوی درگیر به تعداد غدد لنفاوی خارج شده کم‌تر مساوی $0/222$ باشد و تعداد دوره‌های شیمی‌درمانی کم‌تر مساوی ۴ باشد آنگاه به لنفادم مبتلا نخواهد شد.	۶۱/۵۴
۲	اگر تعداد غدد لنفاوی درگیر کم‌تر مساوی یک باشد، بیمار احساس سنگینی نداشته باشد و شاخص توده بدنی کم‌تر از 25 داشته باشد آنگاه به لنفادم مبتلا نخواهد شد.	۷۰
۳	اگر تعداد غدد لنفاوی درگیر کم‌تر مساوی یک باشد، اندازه تومور بیش‌تر مساوی $4/5$ باشد و بیمار یا احساس سنگینی در حد متوسط یا کم داشته باشد آنگاه به لنفادم مبتلا نخواهد شد.	۱۰۰
۴	اگر نسبت تعداد غدد لنفاوی درگیر به تعداد غدد لنفاوی خارج شده کم‌تر مساوی $0/18$ باشد، بیمار احساس سنگینی کمی داشته باشد و در بازه سنی ۴۰ تا ۵۰ سال باشد، آنگاه به لنفادم مبتلا نخواهد شد.	۵۷/۱۴
۵	اگر $0/354 < \text{نسبت تعداد غدد لنفاوی درگیر به تعداد غدد لنفاوی خارج شده}$ $0/18 <$ باشد، احساس سنگینی نداشته باشد و در بازه سنی ۴۰ تا ۵۰ سال باشد، آنگاه به لنفادم مبتلا نخواهد شد.	۵۰
۶	اگر نسبت تعداد غدد لنفاوی درگیر به تعداد غدد لنفاوی خارج شده بیش‌تر از $0/354$ باشد، بیمار احساس سنگینی داشته باشد و تحصیلات کم‌تر از دانشگاهی داشته باشد، آنگاه به لنفادم مبتلا خواهد شد.	۸۱/۸۲
۷	اگر تعداد غدد لنفاوی درگیر بیش‌تر مساوی ۲ باشد، تعداد غدد لنفاوی خارج شده بیش‌تر مساوی ۱۱ باشد، بیمار احساس سنگینی در حد متوسط یا شدید داشته باشد، شاخص توده بدنی کم‌تر از ۳۰ باشد، بیمار خانه‌دار یا معلم باشد و عمل جراحی از نوع MRM باشد، آنگاه به لنفادم مبتلا خواهد شد.	۱۰۰
۸	اگر تعداد غدد لنفاوی درگیر بیش‌تر مساوی ۲ باشد، شاخص توده بدنی بیمار بیش‌تر مساوی ۳۰ باشد، بیمار خانه‌دار یا معلم باشد و عمل جراحی از نوع MRM باشد، آنگاه به لنفادم مبتلا خواهد شد.	۸۹/۴۷
۹	اگر تعداد غدد لنفاوی درگیر بیش‌تر از ۴ و کم‌تر مساوی ۷ باشد و بیمار احساس سنگینی شدید داشته باشد آنگاه به لنفادم مبتلا خواهد شد.	۸۶/۶۷
۱۰	اگر تعداد غدد لنفاوی درگیر بیش‌تر از ۷ باشد و شاخص توده بدنی بیمار بیش‌تر از ۳۰ باشد آنگاه به لنفادم مبتلا خواهد شد.	۱۰۰
۱۱	اگر بیمار متاستاز نداشته باشد، نسبت غدد لنفاوی درگیر به خارج شده بیش‌تر از $0/222$ باشد و بیمار احساس سنگینی شدید داشته باشد آنگاه به لنفادم مبتلا خواهد شد.	۸۸/۴۶

جدول ۲- نسبت شانس برای متغیرهای موثر بر متغیر پاسخ لنفادم

نسبت شانس	p	ضریب تاثیر رگرسیونی	متغیر پیش‌گو
۳/۴۶۵	۰/۰۰۰	۱/۲۴۳	پرتودرمانی
	۰/۰۴۲		مرحله بیماری
۱/۸۹۶	۰/۰۱۲	۰/۶۴۰	مرحله دو بیماری
۱/۷۱۶	۰/۱۳۵	۰/۵۴۰	مرحله سه بیماری
۷/۹۱۳	۰/۰۰۰	۲/۰۶۹	متاستاز داشته باشد
۰/۵۲۶	۰/۰۰۰	-۰/۶۴۰	عمل جراحی از نوع حفظ پستان باشد
	۰/۰۰۱		نوع رژیم شیمی‌درمانی
۰/۷۰۳	۰/۳۴۲	-۰/۳۵۳	تاکسوتر
۱/۳۲۲	۰/۴۵۸	۰/۲۷۹	غیره
۰/۹۲۶	۰/۶۷۵	-۰/۰۷۷	هورمون‌درمانی شده باشد
۱/۳۶۱	۰/۰۵۴	۰/۳۰۸	بیماری همراه
	۰/۰۳۸		فعالیت فیزیکی
۰/۲۲۷	۰/۰۳۱	-۱/۴۸۲	متوسط
۰/۶۵۵	۰/۶۷۶	-۰/۴۲۳	زیاد
	۰/۰۰۱		درد
۱/۰۱۳	۰/۹۲۵	۰/۰۱۶	کم
۰/۳۷۰	۰/۰۰۰	-۰/۹۹۴	متوسط
۰/۵۲۶	۰/۰۳۶	-۰/۶۲۴	زیاد
	۰/۰۰۰		احساس سنگینی
۱/۳۴۵	۰/۰۸۴	۰/۲۹۶	کم
۲/۰۳۵	۰/۰۰۳	۰/۷۱۱	متوسط
۶/۰۵۴	۰/۰۰۰	۱/۸۰۱	زیاد
۳/۵۰۸	۰/۰۰۰	۱/۲۵۵	نسبت تعداد غدد لنفاوی درگیر به تعداد غدد لنفاوی خارج‌شده
	۰/۰۰۰		شاخص توده بدنی
۱/۳۶۵	۰/۰۹۴	۰/۳۱۱	۳۰-۲۵
۲/۰۰۳	۰/۰۰۰	۱/۱۲۵	>۳۰
۰/۷۱۴	۰/۰۳۲	-۰/۳۳۷	دست غالب و درگیر هم‌جهت نباشند
	۰/۰۱۸		سطح تحصیلات
۰/۴۰۵	۰/۰۴۱	-۰/۹۰۵	دبستان
۰/۳۹۳	۰/۰۲۹	-۰/۹۳۵	دیپلم
۰/۲۸۸	۰/۰۰۴	-۱/۲۴۳	دانشگاهی

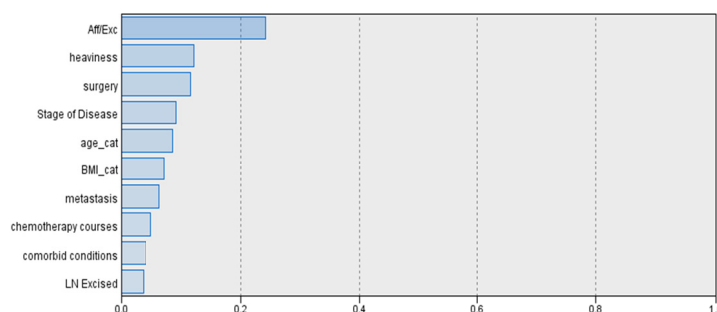
معیارهای دیگری را برای این دسته از مسایل که اصولاً در آنها با دو نوع خطا مواجه هستیم، باید در نظر گرفت. خطای اول تشخیص نادرست فرد مبتلا به لنفادم به عنوان فرد مبتلا نشده و خطای دوم تشخیص اشتباه فرد مبتلا نشده به عنوان فرد مبتلا به لنفادم است. از بین این دو، اشتباه اول هزینه بالاتری دارد و باعث عدم درمان فرد مبتلا به لنفادم می‌شود. بر این اساس معیار مناسب برای تشخیص میزان خطای سیستم در تشخیص نادرست فرد مبتلا به لنفادم به عنوان فرد مبتلا نشده استفاده از معیار حساسیت است. در جدول ۳ شاخص‌های ارزیابی الگوریتم‌های درختان تصمیم و

به درستی توسط روش مورد استفاده، دسته‌بندی شده‌اند. رای بررسی صحت مدل از ماتریس درهم‌ریختگی (Confusion matrix) استفاده شد. جهت ارزیابی مدل‌ها می‌توان از شاخص‌های حساسیت (Sensitivity)، شفافیت یا ویژگی (Specificity)، دقت (Accuracy)، ارزش اخباری مثبت (Precision) و ارزش اخباری منفی (Recall) استفاده کرد (۲۶).

نکته مورد بحث در این مطالعه مربوط به معیارهای ارزیابی است؛ از این منظر باید گفت معیار دقت در پیش‌بینی عملکرد طبقه‌بندی‌ها با توجه به هدف پژوهش، معیار مناسبی محسوب نمی‌شود. در این راستا،

جدول ۳- مقایسه نتایج حاصل از الگوریتم‌های داده کاوی بر حسب درصد

Quest	درخت‌های تصمیم			رگرسیون لجستیک	معیارها
	C&RT	Chaid	C5.0		
۷۲/۶۴	۷۱/۹۲	۷۴/۴۱	۷۹/۳۳	۷۷/۸۳	حساسیت
۵۲/۲۴	۵۴/۹۳	۵۱/۵۶	۵۲/۷۸	۵۴/۵۵	ویژگی
۸۲/۰۲	۸۲/۰۲	۸۳/۵۱	۸۲/۹۱	۸۱/۸۷	ارزش اخباری مثبت
۶۱/۱۱	۵۹/۳۸	۶۲/۰۷	۵۳/۰۹	۵۱/۷۲	ارزش اخباری منفی
۶۷/۵۴	۶۷/۵۲	۶۹/۰۹	۷۲/۵۰	۷۱/۴۳	دقت



شکل ۳- رتبه‌بندی عوامل تاثیرگذار بر پیش‌بینی ابتلا به لنفادم با اجرای مدل درخت تصمیم C5.0

دوره‌های شیمی‌درمانی، بیماری همراه و تعداد غدد لنفاوی خارج شده مهم‌ترین عوامل بروز لنفادم هستند. میتیلی و همکاران نیز با گردآوری اطلاعات مربوط به ۳۰۳ بیمار و با به‌کارگیری الگوریتم‌های درختان تصمیم، ماشین بردار پشتیبان و رگرسیون لجستیک به مدلی جهت پیش‌بینی بیماری قلبی دست یافتند (۲۷). دلن و همکاران با به‌کارگیری الگوریتم‌های شبکه‌های عصبی، درخت تصمیم C5.0 و رگرسیون لجستیک به پیش‌بینی بروز سرطان پستان پرداختند. در این مطالعه الگوریتم درخت تصمیم C5.0 با دقت ۹۳/۶٪ برترین مدل شناخته شد و الگوریتم‌های شبکه عصبی و رگرسیون لجستیک به ترتیب با دقت ۹۱/۲٪ و ۸۹/۲٪ در رتبه‌های بعدی قرار گرفتند (۲۸). شریف‌خانی و همکاران اطلاعات مربوط به ۶۷۰ بیمار را به منظور پیش‌بینی ابتلا به پوکی استخوان بررسی کردند. دقت مدل‌های ساخته شده با استفاده از الگوریتم‌های C5.0، CHAID و شبکه عصبی مصنوعی با یکدیگر مقایسه شده‌اند و نتایج نشان داد الگوریتم شبکه عصبی مصنوعی به نسبت الگوریتم درخت تصمیم دقت بیشتری دارد (۲۹). عامری و همکاران با استفاده از الگوریتم‌های درخت تصمیم C5.0 و شبکه عصبی به پیش‌بینی عوارض بیماران دیابتی بر اساس علائم

رگرسیون لجستیک، ارائه شده است. عوامل نسبت تعداد غدد لنفاوی درگیر به تعداد غدد لنفاوی خارج‌شده، احساس سنگینی، نوع عمل جراحی، مرحله بیماری، سن، شاخص توده بدنی، متاستاز، تعداد دوره‌های شیمی‌درمانی، بیماری همراه و تعداد غدد لنفاوی خارج شده به ترتیب موثرترین عوامل در پیش‌بینی ابتلا به لنفادم هستند.

با توجه به توضیحاتی که داده شد، مدل درخت تصمیم C5.0 با حساسیت ۷۹/۳۳ درصد و دقت ۷۲/۵۰ درصد به عنوان برترین مدل ارائه شده برای پیش‌بینی ابتلا به لنفادم برگزیده شد (شکل ۳).

بحث و نتیجه‌گیری

در این پژوهش، الگوریتم‌های درخت تصمیم C5.0، Quest، C&RT، Chaid و الگوریتم رگرسیون لجستیک به منظور پیش‌بینی احتمال ابتلا به لنفادم بررسی شدند. یافته‌ها نشان می‌دهد الگوریتم درخت تصمیم C5.0 با بالاترین حساسیت و دقت، مدل برتر برای پیش‌بینی ابتلا به لنفادم است. همچنین متغیرهای نسبت تعداد غدد لنفاوی درگیر به تعداد غدد لنفاوی خارج شده، احساس سنگینی، نوع عمل جراحی، مرحله بیماری، سن، شاخص توده بدنی، متاستاز، تعداد

برابر می‌کند. لذا توصیه می‌شود آموزش‌های لازم در خصوص توجه به احساس سنگینی به بیماران داده شود.

مرحله بیماری نیز از عوامل مهم ابتلا به لنفادم به دست آمد که در مطالعات حقیقت و همکاران، دیو و همکاران و همچنین پارک لی و همکاران نیز نتایج مشابه بود (۱۸، ۳۴، ۳۶). همچنین نتایج الگوریتم رگرسیون لجستیک نشان داد در صورتی که بیمار در مرحله دو بیماری باشد به نسبت اینکه در مرحله یک باشد، شانس ابتلا به لنفادم حدوداً دو برابر می‌شود.

همان‌طور که گفته شد اکثر بیماران مبتلا در این مطالعه، سن بالای ۵۰ سال دارند. این موضوع نشان دهنده این است که با افزایش سن، شانس ابتلا به لنفادم بالا می‌رود که این یافته با مطالعه همتی و همکاران همخوانی دارد (۱۷). همچنین در مطالعه مک و همکاران به ازای یک سال افزایش سن، شانس ابتلا به لنفادم در حدود ۱/۰۵ برابر می‌شود (۳۷).

در این مطالعه تعداد دوره‌های شیمی‌درمانی نیز به عنوان یکی از عوامل خطر لنفادم شناخته شد که با مطالعه ایگر و همکاران مطابقت دارد (۶).

از جمله عوامل موثر دیگر بروز لنفادم، نوع عمل جراحی است که در پایان نامه دکتری اسدی و پژوهش همتی و همکاران جز عوامل خطر لنفادم معرفی شده‌است (۱۷، ۳۸). در الگوریتم رگرسیون لجستیک، عمل جراحی حفظ پستان در مقابل عمل جراحی از نوع ماستکتومی شانس ابتلا به لنفادم را در حدود ۵۰ درصد کاهش می‌دهد.

در مطالعه حاضر، متاستاز با بروز لنفادم در ارتباط است و با توجه به الگوریتم رگرسیون لجستیک، از جمله عوامل با اهمیت برای پیش‌بینی ابتلا به لنفادم محسوب می‌شود. در مطالعه ین و همکاران نیز متاستاز از عوامل خطر لنفادم شناخته شده است (۳۹). همچنین این متغیر شانس ابتلا به لنفادم را ۸ برابر می‌کند.

این مطالعه نشان داد تعداد غدد لنفاوی خارج شده نیز با بروز لنفادم ارتباط داشت. در مطالعه میسکی و همکاران تعداد بیش‌تر از ۱۰ غده لنفاوی خارج شده عامل خطر لنفادم شمرده شد (۳۳).

نسبت تعداد غدد لنفاوی درگیر به تعداد غدد لنفاوی

مشاهده شده در آن‌ها پرداختند. دقت درخت تصمیم C5.0 و شبکه عصبی به ترتیب ۸۹/۷۴٪ و ۵۱/۲۸٪ به دست آمد. همچنین مهم‌ترین عوامل تأثیرگذار نیز توسط مدل برتر شناسایی شدند (۳۰).

بر اساس مطالعه حاضر، چاقی بر اساس شاخص توده بدنی و میانگین غدد لنفاوی درگیر از عوامل موثر بر پیش‌بینی ابتلا به لنفادم می‌باشند که این نتیجه با یافته‌های مطالعه حقیقت و همکاران مطابقت دارد (۱۸). شاخص توده بدنی بالا را می‌توان با توجه به اثر فشاری بافت چربی مجاور عروق و غدد لنفاوی در افراد چاق توجه کرد. با توجه به نتایج فوق، به نظر می‌آید که رژیم‌های کنترل وزن باید در افراد در معرض خطر بالای لنفادم مورد توجه قرار گیرد. در مطالعات کلارک و همکاران، همتی و همکاران، رزند و همکاران و لی و همکاران نیز چاقی از عوامل خطر لنفادم است (۴، ۱۷، ۳۱، ۳۲). همچنین از نتایج الگوریتم رگرسیون لجستیک حاصل شد که برای بیماران با شاخص توده بدنی بالای ۳۰ به نسبت بیماران با شاخص توده بدنی کم‌تر از ۲۵، شانس ابتلا به لنفادم ۲ برابر است.

وجود بیماری‌های همراه نظیر پرفشاری خون و یا دیابت عامل خطر دیگری برای بروز لنفادم است که در مطالعه میسکی و همکاران، پرفشاری خون و در مطالعه دیو و همکاران وجود بیماری‌های همراه عامل خطر لنفادم معرفی شده‌اند (۳۳، ۳۴). این متغیر در نتایج حاصل از رگرسیون لجستیک در این پژوهش، معنی‌دار نشد.

در خصوص رابطه بین احساس سنگینی و بروز لنفادم نتایج این پژوهش با نتایج پژوهش آرمر و همکاران تطابق داشت. مطابق مطالعه آرمر و همکاران، احساس سنگینی به عنوان عامل خطر لنفادم شناخته شده‌است. در این مطالعه بیان شده است که با توجه به اینکه تغییرات در حجم اندام درگیر، فاقد دقت مطلوب برای تشخیص لنفادم می‌باشد بنابراین باید دیگر نشانه‌ها نظیر احساس سنگینی نیز در بیماران بررسی شود (۱۶). در واقع اکثر بیمارانی که احساس سنگینی در عضو مبتلا دارند، آن را نادیده می‌گیرند. همچنین با بررسی نتایج الگوریتم رگرسیون لجستیک، احساس سنگینی در حد متوسط شانس ابتلا به لنفادم را ۲ برابر می‌کند و احساس سنگینی زیاد شانس ابتلا را ۶

5. Chen, C, Crooks, S, Keeley, V, Mortimer, P, Sitzia, J, Todd, J. BLS Clinical Definitions. Sevenoaks. Br Lymphol Soc; 2001.

6. Eyigör, S, Cinar, E, Caramat, I, Unlu, BK. Factors influencing response to lymphedema treatment in patients with breast cancer-related lymphedema. Support Care Cancer; 2015. 23(9):2705-2710.

7. Salonen P, Rantanen A, Kellokumpu-Lehtinen, PL, Huhtala H, Kaunonen M. The quality of life and social support in significant others of patients with breast cancer—a longitudinal study. Eur J Cancer Care; 2014. 23(2):274-283.

8. Paiva, DM, Rodrigues, VO, Cesca, MG, Palma, PV, Leite, IC. Prevalence of lymphedema in women undergoing treatment for breast cancer in a referral center in southeastern Brazil. BMC Womens Health; 2013. 13(1):6.

9. Vignes S, Porcher R, Champagne A, Dupuy A. Predictive factors of response to intensive decongestive physiotherapy in upper limb lymphedema after breast cancer treatment: a cohort study. Breast Cancer Res Treat; 2013. 98(1):1-6.

10. Das N, Baumgartner RN, Riley EC, Pinkston CM, Yang D, Baumgartner KB. Treatment-related risk factors for arm lymphedema among long-term breast cancer survivors. J Cancer Surviv; 2015. 9(3): 422-430.

11. Seifart U, Albert U, Heim M, Hübner J, Jungkunz W, Prokein R, et al. Lymphedema in patients with breast cancer—a consensus regarding diagnostics and therapy in patients with postoperative lymphedema after primary breast cancer. Die Rehabil; 2007. 46(6):340-348.

12. Dominick SA, Madlensky L, Natarajan L, Pierce JP. Risk factors associated with breast cancer-related lymphedema in the WHEL Study. J Cancer Surviv; 2013. 7(1):115-123.

13. Ahmed RL, Schmitz KH, Prizment AE, Folsom AR. Risk factors for lymphedema in breast cancer survivors, the Iowa Women's Health Study. Breast Cancer Res Treat; 2011. 130(3):981-991.

14. Wanchai A, Armer JM, Stewart BR, Lasinski BB. Breast cancer-related lymphedema: A literature review for clinical practice. Int J Nurs Stud; 2016. 3(2):202-207.

15. Hortobagyi GN. Treatment of breast cancer. N Engl J Med; 1998. 339(14):974-984.

16. Armer JM, Radina ME, Porock D, Culbertson SD. Predicting breast cancer-related lymphedema using self-reported symptoms. Nurs Res; 2003. 52(6): 370-379.

17. Hemmati S, Jabbari H, Akbari M, Tajvidi M, Rankoohi PA. Factors Associated with the Severity of Lymphedema after the Treatment of Invasive Breast Cancers. J Isfahan Med Sci; 2013. 30(211). (Persian)

18. Haghghat S, Akbari A, Ansari M, Homaei

خارج شده از جمله عوامل خطر دیگر برای پیش بینی ابتلا به لنف ادم می باشد که در این مطالعه مورد بررسی قرار گرفت. با توجه به نتایج الگوریتم رگرسیون لجستیک، با افزایش یک واحد از این نسبت، شانس ابتلا به لنف ادم حدوداً ۳/۵ برابر می شود.

همچنین به نظر می آید که رژیم های کنترل وزن باید در افراد در معرض خطر بالای لنف ادم مورد توجه قرار گیرد. به کارگیری سایر الگوریتم های پیش بینی کننده با دقت بالاتر و تعیین عوامل موثر در بروز لنف ادم برای بیماران سرطان پستان می تواند زمینه های دیگری از پژوهش ها را فراهم سازد.

همچنین می توان در مطالعات دیگری دقت این مدل را در داده های سایر مراکز درمانی و یا سایر مقاطع زمانی مورد ارزیابی قرار داد و پس از ایجاد تغییرات ضروری و رسیدن به سطح مطلوب دقت، در انتخاب مدلی جامع و مناسب جهت پیشگویی احتمال ابتلا به لنف ادم اقدام کرد.

در مطالعات رزند و همکاران و ساعدی همکاران نیز تعداد

دغدغه لنف ادم خارج شده را عامل خطر لنف ادم می دانند (۳۱ و ۳۵).

در مجموع، بر اساس نتایج حاصل از این مطالعه، توجه به احساس سنگینی بیمار در عضو مبتلا و استفاده کم تر از جراحی تهاجمی به عنوان استراتژی های پیش گیرانه لنف ادم توصیه می گردد.

References

1. Haddad P, Farzin M, Amouzegar-Hashemi F, Kalaghchi B, Babazadeh S, Mirzaei HR, et al. A multicentre cross-sectional study of arm lymphedema four or more years after breast cancer treatment in Iranian patients. Breast Cancer; 2010. 17(4):281-285.

2. Erickson VS, Pearson ML, Ganz PA, Adams J, Kahn KL. Arm edema in breast cancer patients. J National Cancer Inst; 2001. 93(2):96-111.

3. Petrek, JA, Pressman, PI, Smith, RA. Lymphedema: current issues in research and management. CA Cancer J Clin; 2000. 50(5):292-307.

4. Clark B, Sitzia J, Harlow W. Incidence and risk of arm oedema following treatment for breast cancer: a three-year follow-up study. QJM; 2005. 98(5):343-348.

- Shandiz F, Najafi M, Ebrahimi M, et al. Risk factors of lymph edema in breast cancer patients. *Int J Breast Cancer*; 2013.
19. Can AG, Ekşioğlu E, Bahtiyarca ZT, Çakıcı FA. Assessment of Risk Factors in Patients who presented to the Outpatient Clinic for Breast Cancer-Related Lymphedema. *J Breast Cancer*; 2016. 12(1):31.
20. Tsai RJ, Dennis LK, Lynch CF, Snetselaar LG, Zamba GK, Scott-Conner C. The risk of developing arm lymphedema among breast cancer survivors: a meta-analysis of treatment factors. *Ann Surg Oncol*; 2009. 16(7):1959-1972.
21. Meeske KA, Sullivan-Halley J, Smith AW, McTiernan A, Baumgartner KB, Harlan LC, Bernstein L. Risk factors for arm lymphedema following breast cancer diagnosis in Black women and White women. *Breast Cancer Res Treat*; 2009. 113(2):383-391.
22. Li X, Huang H, Lin Q, Yu Q, Zhou Y, Long W, et al. Validation of a breast cancer nomogram to predict lymphedema in a Chinese population. *J Surg Res*; 2017. 210:132-138.
23. Paskett ED, Naughton MJ, McCoy TP, Case LD, Abbott JM. The epidemiology of arm and hand swelling in premenopausal breast cancer survivors. *Cancer Epidem Biomar*; 2007. 16(4):775-782.
24. Chapman P, Clinton J, Kerber R, Khabaza T, Reinartz T, Shearer C, et al. CRISP-DM 1.0. Step-by-step data mining guide; 1999.
25. Specht MC, Miller CL, Russell TA, Horick N, Skolny MN, O'Toole JA, et al. Defining a threshold for intervention in breast cancer-related lymphedema: what level of arm volume increase predicts progression? *Breast Cancer Res Treat*; 2013. 140(3):485-494.
26. Ghazanfari M, Alizadeh S, Teimourpour B. Data mining and knowledge discovery. Tehran: Iran University of Science and Technology, Publishing Center. 2008. (Persian)
27. Mythili T, Mukherji D, Padalia N, Naidu A. A heart disease prediction model using SVM-Decision Trees-Logistic Regression (SDL). *Int J Comput Appl*; 2013. 68(16).
28. Delen D, Walker G, Kadam A. Predicting breast cancer survivability: a comparison of three data mining methods. *Artif Intell Med*; 2005. 34(2):113-127.
29. Sharifkhani M, Alizadeh S, Abbasi M, Ameri H. Providing a model for predicting the risk of osteoporosis using decision tree algorithms. *J Mazandaran Uni Med Sci*; 2014. 24(116):110-118.
30. Ameri H, Alizadeh S, Barzegari A. Knowledge Extraction of diabetics' data by decision tree method. *J Health Administ*; 2013. 16(53):58-72.
31. Rezende Lfd, Rocha AVR, Gomes CS. Risk factors for breast cancer related lymphedema. *J Vasc Brasil*; 2010. 9(4):233-238.
32. Li X, Huang H, Lin Q, Yu Q, Zhou Y, Long W, et al. Validation of a breast cancer nomogram to predict lymphedema in a Chinese population. *J Surg Res*; 2017. 210:132-138.
33. Meeske KA, Sullivan-Halley J, Smith AW, McTiernan A, Baumgartner KB, Harlan LC, et al. Risk factors for arm lymphedema following breast cancer diagnosis in Black women and White women. *Breast Cancer Res Treat*; 2009. 113(2):383-391.
34. Deo SV, Ray S, Rath G, Shukla N, Kar M, Asthana S, et al. Prevalence and risk factors for development of lymphedema following breast cancer treatment. *Indian J Cancer*; 2004. 41(1):8.
35. Saedi HS, Kermani AT, Aghaee MA, Mohagheghi T, Moghadam S. Study of Lymphedema Risk Factors among Patients with Invasive Breast Cancer after the Initial Treatment. *Reports Radiother Oncol*; 2013. 1(1).
36. Park JH, Lee WH, Chung HS. Incidence and risk factors of breast cancer lymphoedema. *J Clin Nurs*; 2008. 17(11):1450-1459.
37. Mak SS, Yeo W, Lee YMM KF, Tse KY, Tse SM, et al. Predictors of lymphedema in patients with breast cancer undergoing axillary lymph node dissection in Hong Kong. *Nurs Res*; 2008. 57(6):416-425.
38. Asadi S. Lymphedema Risk Factors in Breast Cancer Patients) PhD Thesis), Shahid Beheshti University of Medical Sciences, Faculty of Medicine. 2008. (Persian)
39. Yen TW, Fan X, Sparapani R, Laud PW, Walker AP, Nattinger AB. A contemporary, population-based study of lymphedema risk factors in older women with breast cancer. *Ann Surg Oncol*; 2009. 16(4):979-988.