

Investigating the Correlation of Demographic and Paraclinical Characteristics in Metastasis to Lymph nodes in Breast Cancer in Mazandaran Province

Fatemeh Zamaninasab¹
Afsaneh Fendereski^{2,3}
Gholamali Godazandeh⁴
Jamshid Yazdani-Charati^{5,3}

¹ MSc Student in Biostatistics, Faculty of Health, Mazandaran University of Medical Sciences, Sari, Iran

² Assistant Professor, Department of Biostatistics, Health Sciences Research Center, Addiction Institute, Mazandaran University of Medical Sciences, Sari, Iran

³ Faculty of Health, Mazandaran University of Medical Sciences, Sari, Iran

⁴ Professor, Department of General Surgery, Faculty of Medicine, Mazandaran University of Medical Sciences, Sari, Iran

⁵ Professor, Department of Biostatistics, Health Sciences Research Center, Addiction Institute, Mazandaran University of Medical Sciences, Sari, Iran

(Received July 20, 2023; Accepted October 24, 2023)

Abstract

Background and purpose: The status of axillary lymph nodes (ALN) is the most important prognostic factor in patients with primary breast cancer. The aim of this study is to measure the relationship between the factors influencing metastasis to the lymph nodes in patients with breast cancer.

Materials and methods: The present study is a retrospective cohort study. The research data were extracted from telephone interviews and medical records of 241 breast cancer patients in Mazandaran Medical Sciences Hospitals. For data analysis, the chi-square test and multiple logistic regression model using principal component analysis were used to determine the relationship between factors affecting metastasis to lymph nodes. All statistical analyses were performed using Stata software version 17, R and SPSS software version 26.

Results: The average age of patients at diagnosis is 52.03 ± 10.932 years test revealed a significant correlation between the markers ER, PR, P53, Her2 and Ki67 ($P < 0.05$). After removing collinearity using principal component analysis, the variables age, tumor size, skin involvement, grade, tubule formation and nuclear pleomorphism were selected as input variables for multiple regression. The variable grade was significant with a P-value of 0.012.

Conclusion: It is very important to identify the factors that may be effective in predicting response to treatment or disease progression. In short, there is a significant correlation between the markers ER, PR, P53, Her2 and Ki67, and the variable grade is an important factor for lymph node metastasis.

Keywords: Breast Cancer, Logistic Regression, Lymph Node Metastasis

J Mazandaran Univ Med Sci 2023; 33 (Supple 2): 181-191 (Persian).

Corresponding Author: Jamshid Yazdani-Charati Faculty of Health, Mazandaran University of Medical Sciences, Sari, Iran. (E-mail: jamshid.charati@gmail.com) and **Gholamali Godazandeh** Faculty of Medicine, Mazandaran University of Medical Sciences, Sari, Iran. (E-mail: godazandeh.gh@gmail.com)

بررسی رابطه بین ویژگی‌های جمعیت‌شناسی و پاراکلینیک با متاستاز به غدد لنفاوی در بیماران سرطان پستان استان مازندران

فاطمه زمانی نسب^۱
افسانه فندرسکی^{۳و۲}
غلامعلی گدازنده^۴
جمشید یزدانی چراتی^{۳و۵}

چکیده

سابقه و هدف: وضعیت غدد لنفاوی زیر بغل (ALN Axillary Lymph Node) مهم‌ترین عامل پیش‌آگهی در بیماران مبتلا به سرطان پستان اولیه است. هدف این مطالعه سنجش ارتباط بین عوامل مؤثر بر متاستاز به غدد لنفاوی در بیماران با سرطان پستان انجام شده است.

مواد و روش‌ها: پژوهش حاضر، مطالعه کوهورت گذشته‌نگر است. داده‌های تحقیق از مصاحبه تلفنی و بررسی پرونده‌های پزشکی ۲۴۱ نفر از مبتلایان به سرطان پستان بیمارستان‌های علوم پزشکی مازندران استخراج شده است. روش تحلیل داده‌ها با استفاده از آزمون کای دو و برازش مدل رگرسیون لجستیک چندگانه با استفاده از تحلیل مؤلفه‌های اصلی با هدف شناسایی ارتباط بین عوامل مؤثر بر متاستاز به غدد لنفاوی در بیماران در نرم‌افزارهای STATA نسخه ۱۷، R و SPSS نسخه ۲۶ انجام شد.

یافته‌ها: میانگین سن تشخیص بیماران ۵۲/۰۳±۱۰/۹۳۲ سال است. طبق آزمون کای دو بین مارکرهای ER، PR، P53، Ki67 و Her2 ارتباط معنی‌داری یافت شده است ($P < 0/05$) پس از رفع هم‌خطی با استفاده از تحلیل مؤلفه‌های اصلی، متغیرهای سن، سایز تومور، درگیری پوست، Grade، Tubule formation و Nuclear pleomorphism به‌عنوان متغیرهای ورودی رگرسیون چندگانه انتخاب شدند و متغیر Grade با $P = 0/012$ معنی‌دار شد.

استنتاج: شناسایی عواملی که می‌توانند در پیشگویی پاسخ به درمان و یا پیشرفت بیماری اثرگذار باشند، بسیار مهم است. متغیر Grade یک عامل اثرگذار بر متاستاز به غدد لنفاوی است. هم‌چنین بین مارکرهای ER، PR، P53، Her2 و Ki67 نیز ارتباط معنی‌دار موجود است.

واژه‌های کلیدی: سرطان پستان، رگرسیون لجستیک، متاستاز به غدد لنفاوی

مقدمه

امروزه سرطان به یک بیماری رایج تبدیل شده است. رشد نامنظم در یکی از سلول‌های تشکیل‌دهنده بدن که می‌تواند زندگی افراد را تحت تأثیر قرار دهد (۱). انسان را سرطان می‌نامند. یکی از مسائل مهم در انواع

E-mail: jamshid.charati@gmail.com

E-mail: godazandeh.gh@gmail.com

مؤلف مسئول: جمشید یزدانی چراتی - ساری: کیلومتر ۱۷ جاده فرح‌آباد، مجتمع دانشگاهی، پیامبر اعظم، دانشکده بهداشت

و غلامعلی گدازنده - ساری: کیلومتر ۱۷ جاده فرح‌آباد، مجتمع دانشگاهی، پیامبر اعظم، دانشکده پزشکی

۱. دانشجوی کارشناسی ارشد آمار زیستی، دانشکده بهداشت، دانشگاه علوم پزشکی مازندران، ساری، ایران

۲. استادیار، گروه آمار زیستی و اپیدمیولوژی، مرکز تحقیقات علوم بهداشتی، دانشگاه علوم پزشکی مازندران، ساری، ایران

۳. دانشکده بهداشت، دانشگاه علوم پزشکی مازندران، ساری، ایران

۴. استاد، گروه جراحی، دانشکده پزشکی، دانشگاه علوم پزشکی مازندران، ساری، ایران

۵. استاد، گروه آمار زیستی و اپیدمیولوژی، مرکز تحقیقات علوم بهداشتی، دانشگاه علوم پزشکی مازندران، ساری، ایران

تاریخ دریافت: ۱۴۰۲/۴/۲۹ تاریخ ارجاع جهت اصلاحات: ۱۴۰۲/۵/۱۸ تاریخ تصویب: ۱۴۰۲/۸/۲۰

سرطان، انتشار سلول سرطانی به بافت‌های دیگر (متاستاز) است. هر ساله هزاران زن قربانی سرطان پستان می‌شوند (۳،۲). سرطان پستان شایع‌ترین نوع سرطان در بین زنان، دومین سرطان شایع در جهان پس از سرطان ریه و دلیل اصلی مرگ و میر زنان در تمامی کشورهای توسعه یافته و توسعه نیافته است (۴-۶). خطر ابتلا در طول زندگی ۱۲ درصد است (۶). طبق برآوردهای سازمان جهانی بهداشت (World Health Organization: WHO)، سرطان پستان تقریباً یک مورد از هر شش مرگ در سراسر جهان را شامل می‌شود، این سرطان بیش از ۲۴ درصد از کل سرطان‌ها را در ایران تشکیل می‌دهد، با نرخ شیوع ۳۴-۲۴/۸ در هر ۱۰۰۰۰۰ زن و نرخ مرگ و میر کم‌تر از ۱۰/۲ در هر ۱۰۰۰۰۰ نفر در سال ۲۰۱۸. میزان مرگ و میر ناشی از سرطان پستان در ۳۰ سال گذشته در ایران تغییری نکرده است. نرخ استاندارد شده برای مرگ و میر ۱۴/۲ در ۱۰۰۰۰۰ با میانگین سنی ۴۹/۸۴ سال بود. شایع‌ترین سرطان در ایران کارسینوم مجرای مهاجم است (۷). سرطان پستان یکی از انواع سرطانی است که تشخیص زودهنگام آن از اهمیت ویژه‌ای برخوردار است. هر چه سرطان پستان زودتر تشخیص داده شود، شانس درمان بیمار بیش‌تر است و می‌توان با ارائه مراقبت‌های به‌موقع امکان بقا را بهبود بخشید (۸). وضعیت غدد لنفاوی زیر بغل (Axillary Lymph Node: ALN) مهم‌ترین عامل پیش‌آگهی در بیماران مبتلا به سرطان پستان اولیه است (۱). شناخت عوامل پیش‌آگهی می‌تواند به انتخاب مناسب‌ترین روش درمانی کمک کند (۹). این مقاله منحصرأ به سرطان پستان، شایع‌ترین مشکل برای بانوان در سراسر جهان، می‌پردازد. هدف از این مطالعه تعیین عوامل مرتبط با متاستاز غدد لنفاوی زیر بغل است.

مواد و روش‌ها

پژوهش حاضر از نوع مطالعه کوهورت در بین بیماران سرطان پستان در استان مازندران انجام شده

است. از آنجایی که بیش‌تر بیماران برای جراحی به بیمارستان‌های آموزشی علوم پزشکی شهر ساری مراجعه می‌کنند؛ داده‌های تحقیق از بررسی پرونده‌های پزشکی ۲۴۱ نفر از مبتلایان به سرطان پستان که در فاصله بین سال‌های ۱۳۹۵ تا ۱۴۰۱ به بیمارستان‌های علوم پزشکی مازندران در شهر ساری مراجعه نموده‌اند، استخراج شده است و پرونده‌هایی که اطلاعات ناکافی داشتند از مطالعه خارج گردید و تنها پرونده‌هایی که اطلاعات کافی یا به نسبت کافی داشتند در مطالعه باقی ماندند. سپس برای تکمیل فرایند جمع‌آوری داده‌ها از طریق تماس تلفنی نسبت به گرفتن دیگر اطلاعات بیماران اقدام شد. در پایان برای آن اطلاعات گمشده‌ای (missing value) که طی فرایند پرونده خوانی و تماس تلفنی یافت نشدند و امکان دسترسی به آن‌ها وجود نداشت از جانهی استفاده شده است. معیار ورود به مطالعه، دارا بودن تومور بدخیم سرطان پستان بوده است. داده‌ها با استفاده از چک‌لیستی که به تأیید جراحان قفسه سینه و آنکولوژیست‌های سرطان پستان رسیده است، جمع‌آوری شده است. هم‌چنین روایی این چک‌لیست توسط متخصصان مربوطه مورد بررسی قرار گرفته و از پیشنهاد‌های آنان در رفع نقایص و بهبود و ارتقا کیفیت چک‌لیست استفاده شده است. این چک‌لیست توانایی لازم برای رسیدن به اهداف مقاله پیش‌رو را دارا است. داده‌های تحقیق شامل متغیرهای سن، وضعیت تأهل، گیرنده‌ی استروژن (ER)، گیرنده‌ی پروژسترون (PR)، سایز تومور، مکان تومور، stage، P53، Ki67، HER2، درگیری پوست، DCIS margin (Ductal carcinoma in situ)، Nuclear pleomorphism، Tubule formation، Grade و Mitosis به‌عنوان متغیرهای مستقل و زمینه‌ای و متاستاز به غدد لنفاوی به‌عنوان متغیر پاسخ در چک‌لیست وارد شد. حجم نمونه در این پژوهش حداقل ۲۰۰ (پنج تا ده برابر تعداد متغیرها) در نظر گرفته شده است (۱۰). برای سنجش ارتباط بین عوامل مؤثر بر متاستاز به غدد لنفاوی از آزمون کای دو (Chi-squared test) استفاده شده

ابعاد یک مجموعه داده است. PCA با ایجاد ترکیب خطی بین متغیرهای هم‌بسته و ساخت متغیر ترکیبی جدید، مشکل هم‌خطی را رفع می‌کند (۱۴).

نرم‌افزارهای مورد استفاده در این تحقیق استیتم (STATA)، SPSS و R هستند. تحلیل‌ها و آزمون‌های آماری در نرم‌افزار STATA نسخه ۱۷ و SPSS نسخه ۲۶ انجام شده است. از نرم‌افزار R برای انجام جانهای برای داده‌های گمشده (missing value) با استفاده از پکیج randomForest و دستور rfImpute استفاده شده است. به علت عملکرد مناسب روش جنگل تصادفی برای جانهای داده‌های گمشده، در این تحقیق مورد استفاده قرار گرفته است (۱۵). دستور rfImpute جانهای را با برازش جنگل تصادفی بر روی داده‌ها و به دست آوردن مقادیر پیش‌بینی برای داده‌های گمشده و تکرار این فرایند انجام می‌دهد. در مرحله اول مقدار میانگین، میانه یا مد را به عنوان مقدار اولیه شروع الگوریتم برای داده‌های گمشده جایگذاری کرده سپس با اعمال برازش جنگل تصادفی، مقادیر پیش‌بینی حاصل را برای داده‌های گمشده جایگذاری می‌کند و مجدد برازش جنگل تصادفی را انجام می‌دهد با تکرار این عمل به بهترین مقدار پیش‌بینی برای جانهای داده‌های گمشده می‌رسد. در بیش تر موارد با پنج تا شش تکرار برازش جنگل تصادفی بهترین مقدار برای داده‌های گمشده حاصل می‌شود. نکته‌ی مهم در این دستور این است که گمشدگی در متغیر پاسخ نباید وجود داشته باشد (۱۶). بر اساس آنچه گفته شد نخست برای سنجش ارتباط بین متغیرهای مستقل و زمینه‌ای با متغیر وابسته از آزمون‌های کای دو و t-test استفاده شده و هم‌چنین برای سنجش ارتباط دو به دو برخی متغیرهای مستقل، آزمون کای دو استفاده شد. هدف دیگر این تحقیق سنجش ارتباط هم‌زمان متغیرهای مستقل و زمینه‌ای با متغیر پاسخ بوده است. برای تحقق این هدف نخست با استفاده از تحلیل مؤلفه‌های اصلی مشکل هم‌خطی برخی متغیرها رفع شده و سپس با استفاده از رگرسیون لجستیک تک متغیره

است. برای سنجش ارتباط هم‌زمان متغیرهای مستقل و متغیر پاسخ از رگرسیون لجستیک چندگانه استفاده شده است. طبق تعریف سرفصل‌های موضوعی پزشکی (MeSH)، رگرسیون لجستیک به معنای مدلی است که ارتباط بین یک متغیر وابسته‌ی کیفی (که فقط می‌تواند مقادیر گسسته‌ی خاصی را بگیرد، مثل وجود یا عدم وجود یک بیماری) و یک یا چند متغیر مستقل را توصیف می‌کند. به رگرسیون لجستیک، مدل‌های لجیت (Logit Models) نیز گفته می‌شود. به‌طور معمول متغیر وابسته کیفی دو حالت دارد، مانند وجود یا عدم وجود سرطان که در این صورت مدل به کار رفته را مدل لجستیک دو حالت (Binary Logistic) می‌گویند (۱۱). رگرسیون لجستیک از خانواده‌ی GLM است. عبارت GLM مخفف Generalized Linear Models است که حالت تعمیم‌یافته‌ای از رگرسیون خطی معمولی است (۱۲).

همانند بسیاری از مدل‌های آماری، مدل رگرسیون لجستیک نیز دارای مفروضاتی است (۱۳):

- فرض اول استقلال خطاها است که به‌موجب آن تمام نتایج گروه نمونه از یکدیگر جدا هستند.
 - فرض دوم خطی بودن در لجیت (Logit) برای هر متغیر مستقل پیوسته (مثلاً سن) است.
 - فرض سوم عدم وجود هم‌خطی چندگانه در بین متغیرهای مستقل است. یک مدل رگرسیون لجستیک با متغیرهای مستقل بسیار هم‌بسته معمولاً منجر به خطاهای استاندارد (Standard Errors) بزرگ برای ضرایب رگرسیونی این متغیرها می‌شود.
 - فرض چهارم فقدان داده‌های پرت و تأثیرگذار است؛ اگر داده‌های پرت بیش از حد وجود داشته باشد، دقت کلی مدل ممکن است به خطر بیفتد.
- به علت عدم برقراری فرض سوم، لازم است ابتدا مشکل هم‌خطی موجود در متغیرهای این تحقیق رفع شود. به کارگیری تحلیل مؤلفه اصلی یا PCA از جمله راه‌های رفع مشکل هم‌خطی در متغیرها است. تحلیل مؤلفه‌های اصلی در واقع یک روش آماری برای کاهش

بهترین متغیرها برای ورود به رگرسیون لجستیک چندگانه انتخاب شدند و رگرسیون لجستیک چندگانه بر روی آنان برازش داده شد.

طرح تحقیقاتی حاضر در کمیته اخلاق در پژوهش (کد اخلاق: IR.MAZUMS.REC.1401.14995) مطرح گردید و با شماره مورد تصویب قرار گرفت.

یافته‌ها

داده‌های این مطالعه برگرفته از بررسی پرونده‌ی ۲۴۱ بیمار با سرطان پستان از بیمارستان‌های علوم پزشکی شهر ساری است. همان‌طور که در نمودار ۱ مشاهده می‌شود، از میان ۲۴۱ پرونده‌ی بررسی شده، ۱۸ پرونده به دلایلی چون عدم درج اطلاعات لازمه، گمشدگی بالای اطلاعات و غیره از مطالعه خارج گردید و حدود ۲۲۳ پرونده در این مطالعه مورد ارزیابی قرار گرفته است. تمامی داده‌ها متعلق به بانوان است. میانگین سن تشخیص مبتلایان ۱۰/۹۳۲±۵۲/۰۳ سال است. از میان ۲۲۳ نفر از مبتلایان به سرطان پستان، ۱۱۱ نفر متاستاز به غدد لنفاوی دارند و ۱۱۲ نفر متاستاز به غدد لنفاوی ندارند. آزمون مقایسه‌ی میانگین سن تشخیص سرطان پستان در دو گروه متاستاز به غدد لنفاوی و عدم متاستاز به غدد لنفاوی در سطح معنی‌داری ۰/۹۵ معنی‌دار نشد (sig=۰/۱۹۵). بنابراین تفاوت معناداری در میانگین سنی در دو گروه متاستاز به غدد لنفاوی و عدم متاستاز به غدد لنفاوی وجود ندارد. اطلاعات توصیفی متغیرها در جدول شماره ۱ نشان داده شده است، با انجام آزمون کای دو برای متغیرهای کیفی و t-test برای متغیرهای کمی ارتباط بین تک تک متغیرها با متغیر پاسخ سنجیده شده است و P-value تک تک آزمون‌ها در سطح معنی‌داری ۰/۹۵ در جدول شماره ۱ گزارش شده است. براساس نتایج حاصل از جدول شماره ۱ تنها بین متاستاز به غدد لنفاوی و متغیرهای Grade و Tubule formation از نظر آماری ارتباط معنی‌داری وجود دارد. ارتباط دو به دوی برخی از متغیرهای مستقل و

زمینه‌ای در سطح معنی‌داری ۰/۹۵ با استفاده از آزمون کای دو سنجیده شده است و نتایج آن در جدول شماره ۲ قابل مشاهده است. طبق نتایج جدول شماره ۲ بین متغیرهایی که P-Value آنان کم‌تر از ۰/۰۵ است از نظر آماری ارتباط معنی‌داری وجود دارد؛ بنابراین مارکر ER با متغیرهای Grade، مارکر PR، مارکر P53، مارکر Ki67، Nuclearpleomorphism و Mitosis ارتباط معنی‌داری دارد اما با متغیرهای درگیری پوست، مارکر Her2، درگیری حاشیه، DCIS و Tubule formation ارتباط معنی‌داری ندارد. مارکر PR نیز علاوه بر موارد گفته شده با متغیرهای مارکر P53، مارکر Ki67، Mitosis و Nuclear pleomorphism، Grade، Her2، درگیری حاشیه، DCIS و Tubule formation ارتباط معنی‌داری یافت نشد. مارکر P53 نیز به جز روابط گفته شده با متغیرهای Her2، درگیری حاشیه، Ki67، Grade، Nuclear pleomorphism و Mitosis نیز ارتباط معنی‌داری دارد اما با متغیرهای درگیری پوست، CIS و Tubule formation ارتباط معنی‌داری یافت نشد. مارکر Ki67 علاوه بر موارد گفته شده با دو متغیر DCIS و Mitosis ارتباط معنی‌داری دارد. علاوه بر روابط گفته شده برای متغیر Her2؛ بین این متغیر و متغیرهای درگیری حاشیه، Grade، Nuclear pleomorphism و Mitosis نیز ارتباط معنی‌داری مشاهده شد. متغیر Mitosis نیز ارتباط معنی‌داری دارد. درگیری پوست علاوه بر موارد گفته شده با متغیرهای درگیری حاشیه، DCIS، Nuclear pleomorphism و Mitosis از نظر آماری ارتباط معنی‌داری دارد. متغیر DCIS علاوه بر روابط گفته شده با متغیرهای Grade، Mitosis و Nuclear pleomorphism، Tubule formation نیز ارتباط معنی‌داری دارد.

برای برقراری فرض سوم مدل رگرسیون لجستیک (عدم وجود هم‌خطی) به علت وجود متغیرهای هم‌نوع (subtype) مارکر ER، مارکر PR، مارکر P53، مارکر Ki67 و Her2 (۱۵) در بین متغیرهای تحقیق برای رفع

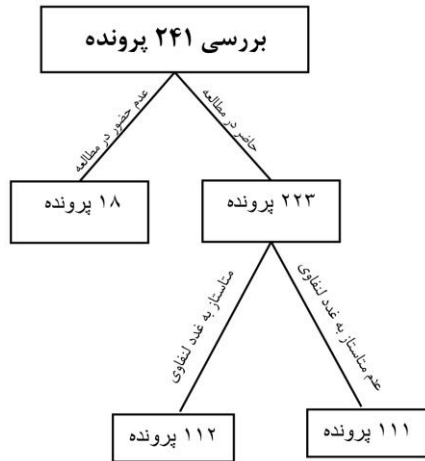
جدول شماره ۱: اطلاعات توصیف و مقایسه ویژگی‌های جمعیت شناختی و بالینی در دو گروه بیماران سرطان پستان با و بدون متاستاز به غدد لنفاوی

نام متغیر	متاستاز به غدد لنفاوی ندارد		متاستاز به غدد لنفاوی دارد		سطح معنی داری
	انحراف معیار ± میانگین	دامنه	انحراف معیار ± میانگین	دامنه	
سن*	۵۱٫۳±۱۱٫۵۳	۸۶-۲۲	۵۳٫۱۷±۹٫۸	۳۳-۷۴	۰/۱۹۵
اندازه تومور*	۲٫۳±۱/۱۲	۰-۵	۲/۵±۱/۲۵	۱-۸	۰/۱۸۲
متغیر	متاستاز به غدد لنفاوی ندارد	متاستاز به غدد لنفاوی دارد	متغیر	متاستاز به غدد لنفاوی دارد	سطح معنی داری
وضعیت تأهل	مجرد متاهل	۴۱(۵۲/۶) ۷۰(۴۸/۳)	۳۷(۴۷/۴) ۷۵(۵۱/۷)	۷۸(۳۵) ۱۴۵(۶۵)	۰/۵۴۱
مکان تومور	۰(چپ) ۱(راست) ۲(هر دو)	۴۸(۵۲/۲) ۶۱(۴۷/۷) ۲(۶۶/۷)	۴۴(۴۷/۸) ۶۷(۵۲/۳) ۱(۳۳/۳)	۹۲(۴۱/۳) ۱۲۸(۵۷/۴) ۳(۱/۳)	۰/۶۷۶
Stage	I II III	۶۱(۵۳) ۴۵(۴۷/۴) ۵(۳۸/۵)	۵۴(۴۷) ۵۰(۵۲/۶) ۸(۶۱/۵)	۱۱۵(۵۱/۶) ۹۵(۴۲/۶) ۱۳(۵/۸)	۰/۵۰۲
ER	۰(منفی) ۱(مثبت)	۲۵(۴۵/۵) ۸۶(۵۱/۲)	۳۰(۵۴/۵) ۸۲(۴۸/۸)	۵۵(۲۴/۷) ۱۶۸(۷۵/۳)	۰/۶۶۰
PR	۰(منفی) ۱(مثبت)	۳۲(۴۵/۱) ۷۹(۵۲)	۳۹(۵۴/۹) ۷۳(۴۸)	۷۱(۳۱/۸) ۱۵۲(۶۸/۳)	۰/۳۳۷
P53	۰(منفی) ۱(مثبت)	۶۲(۴۶/۶) ۴۹(۵۴/۴)	۷۱(۵۳/۴) ۴۱(۴۵/۶)	۱۳۳(۵۹/۶) ۹۰(۴۰/۴)	۰/۲۵۱
Ki67	۰(منفی) ۱(مثبت)	۳۳(۵۵/۹) ۷۸(۴۷/۶)	۲۶(۴۴/۱) ۸۶(۵۲/۴)	۵۹(۲۶/۵) ۱۶۴(۷۳/۵)	۰/۲۷۰
Her2	۰(منفی) ۱(مثبت)	۷۰(۵۱/۹) ۴۱(۴۶/۶)	۶۵(۴۸/۱) ۴۷(۵۲/۴)	۱۳۵(۶۰/۵) ۸۸(۳۹/۵)	۰/۴۴۲
درگیری پوست	۰(عدم درگیری) ۱(درگیری)	۹۷(۴۸) ۱۴(۶۶/۷)	۱۰۵(۵۲) ۳۷(۳۳/۳)	۲۰۲(۹۰/۶) ۲۱(۷/۹)	۰/۱۰۴
درگیری حاشیه	۰(عدم درگیری) ۱(درگیری)	۹۷(۴۸/۷) ۱۴(۵۸/۳)	۱۰۲(۵۱/۳) ۱۰(۴۱/۷)	۱۹۹(۸۹/۲) ۲۴(۱۰/۸)	۰/۳۷۵
DCIS	۰(absent) ۱(present)	۳۶(۴۷/۴) ۳۸(۵۱/۴)	۴۰(۵۲/۶) ۳۶(۴۸/۶)	۷۶(۳۴/۱) ۷۴(۳۳/۲)	۰/۸۷۲
Grade	I II III	۳۵(۶۲/۵) ۲۴(۴۱/۹) ۲۲(۵۷/۹)	۲۱(۳۷/۵) ۷۵(۵۸/۱) ۱۶(۴۱/۱)	۵۶(۲۵/۱) ۱۲۹(۵۷/۸) ۳۸(۱۷)	۰/۰۰۲
Tubule formation	I II III	۱۸(۵۶/۳) ۲۴(۶۶/۷) ۶۹(۴۴/۵)	۱۴(۴۲/۸) ۱۲(۳۳/۳) ۸۶(۵۵/۵)	۳۲(۱۴/۳) ۳۶(۱۶/۱) ۱۵۵(۶۹/۵)	۰/۰۴۲
Nuclear pleomorphism	I II III	۲۲(۵۷/۹) ۲۵(۴۶/۱) ۲۴(۵۴/۵)	۱۶(۴۲/۱) ۷۶(۵۳/۹) ۲۰(۴۵/۵)	۳۸(۱۷) ۱۴۱(۶۳/۲) ۴۴(۱۹/۷)	۰/۳۳۹
Mitosis	I II III	۲۴(۵۰/۳) ۱۹(۴۵/۲) ۱۸(۵۲/۹)	۳۳(۴۹/۷) ۲۳(۵۴/۸) ۱۶(۴۷/۱)	۱۴۷(۶۵/۹) ۴۲(۱۸/۸) ۳۴(۱۵/۲)	۰/۸۷۸

*: آزمون t_test استفاده شده است.

جدول شماره ۲: سطح معنی داری حاصل از آزمون کای دو بین متغیرهای کیفی در بیماران با سرطان پستان

Mitosis	Nuclear pleomorphism	Tubule formation	Grade	DCIS	درگیری حاشیه	درگیری پوست	Her2	Ki67	P53	PR	ER
											ER
											PR
										۰/۰۰	۰/۰۰
									۰/۰۰۲	۰/۰۰	۰/۰۰۱
								۰/۰۶	۰/۰۰۷	۰/۰۰۳	۰/۰۰۹
							۰/۰۸۲	۰/۰۴۵	۰/۰۱۹	۰/۰۰۷	۰/۰۹۲۴
						۰/۰۰	۰/۰۴۵	۰/۰۱۹	۰/۰۴۵	۰/۰۵۸۸	درگیری پوست
					۰/۳۴	۰/۰۳۱	۰/۰۳	۰/۰۳۷	۰/۵۵	۰/۳۲	۰/۷۱۹
				۰/۰۰	۰/۰۲	۰/۰۱	۰/۰۰۱	۰/۴۲	۰/۰۰۲	۰/۰۰	۰/۰۰
			۰/۰۰	۰/۰۰	۰/۰۸	۰/۴۳	۰/۰۰۳	۰/۰۰۱	۰/۰۰۶	۰/۳۵	۰/۸۷۸
		۰/۰۰	۰/۰۰	۰/۰۰	۰/۰۰۱	۰/۰۱۳	۰/۰۰۲	۰/۰۰۹	۰/۰۰	۰/۰۰	۰/۰۰
	۰/۰۰	۰/۱۳	۰/۰۰	۰/۰۰	۰/۰۰	۰/۰۰۷	۰/۰۰۷	۰/۰۰	۰/۰۰	۰/۰۰	۰/۰۰
۰/۷۷	۰/۳۴	۰/۰۴	۰/۰۲	۰/۸۷	۰/۳۷	۰/۱	۰/۴۴	۰/۲۷	۰/۲۵	۰/۳۳۷	۰/۴۶



فلوچارت شماره ۱: فلوچارت بررسی پرونده‌های بیمارستانی

در ادامه تحلیل با دستیابی به دو متغیر ترکیبی جدید حاصل از تحلیل مؤلفه‌های اصلی (شامل متغیرهای مارکر ER، مارکر PR، مارکر P53، مارکر Ki67 و Her2) و رفع مشکل هم‌خطی رگرسیون لجستیک تک متغیر مطابق جدول شماره ۳ بر روی داده‌های جدید برازش داده شده است. طبق جدول شماره ۳ در تحلیل تک متغیره، متغیرهای سن، سایز تومور، درگیری پوست، Nuclear pleomorphism و Tubule formation، Grade در سطح ۰/۷ معنی دار شدند و به‌عنوان متغیرهای ورودی برای رگرسیون لجستیک چندگانه انتخاب می‌شوند، سایر متغیرها تأثیر معنی‌داری بر متاستاز به غدد لنفاوی نداشتند. پس از برازش رگرسیون لجستیک چندگانه با متغیرهای ورودی نام برده شده نتایج جدول شماره ۴ حاصل می‌شوند. طبق جدول شماره ۴ در تحلیل رگرسیون لجستیک چندگانه با متغیرهایی که در تحلیل تک متغیره معنی دار شده بودند نتایج به شرح زیر است:

به ازای هر سال افزایش سن، شانس متاستاز به غدد لنفاوی ۱/۰۲ برابر (دو درصد بیش‌تر) می‌شود. به ازای هر واحد افزایش در سایز تومور، شانس متاستاز به غدد لنفاوی ۱/۱۴ برابر (۱۴ درصد بیش‌تر) می‌شود. شانس متاستاز به غدد لنفاوی در بیمارانی که درگیری پوست داشتند ۰/۵۱ برابر (نصف) بیمارانی است که درگیری پوست نداشتند.

هم‌خطی بین این متغیرها از روش تحلیل مؤلفه‌های اصلی استفاده شده و به جای پنج متغیر ذکر شده از دو متغیر ترکیبی حاصل از تحلیل مؤلفه‌های اصلی (PC_Scores) در برازش مدل رگرسیون لجستیک استفاده شده است.

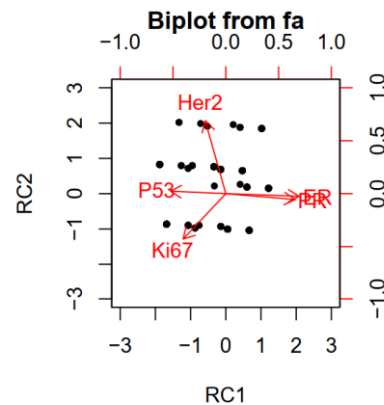
پس از به‌کارگیری تحلیل مؤلفه‌های اصلی به دلیل هم‌خطی، دو ترکیب خطی زیر حاصل شد:

$$PC1 = 0.869ER + 0.822PR - 0.655P53 - 0.484Ki67 - 0.275Her2$$

$$PC2 = 0.001ER - 0.039PR + 0.005P53 - 0.554Ki67 + 0.851Her2$$

نمودار biplot حاصل از تحلیل مؤلفه اصلی مطابق

نمودار شماره ۲ است.



تصویر شماره ۱: بای پلات روابط بین متغیرهای مارکر ER، مارکر PR، مارکر P53، مارکر Ki67 و Her2. مؤلفه اول: ۴۳/۴ درصد مؤلفه دوم: ۲۰/۷ درصد

طبق فلوچارت شماره ۱، دو متغیر مارکرهای ER و PR به علت نزدیکی دو فلش مربوطه‌شان در فلوچارت شماره ۱ با یکدیگر همبستگی بالایی دارند. متغیر مارکر P53 با دو متغیر مارکر ER و PR به علت تشکیل زاویه حدودی ۱۸۰ درجه بین فلش‌هایشان در نمودار یاد شده همبستگی منفی بالا دارند. سه متغیر مارکرهای ER، PR و P53 بیش‌ترین تأثیر را بر روی مؤلفه‌ی اول دارند. متغیر Her2 به علت تشکیل زاویه تقریباً ۹۰ درجه بین فلش‌های مربوطه‌شان در نمودار یاد شده، همبستگی کمی با دو متغیر مارکرهای ER و PR دارد.

بحث

در مطالعات درمان بیماری‌ها، شناسایی عوامل و ویژگی‌هایی که می‌توانند در پیشگویی پاسخ به درمان یا پیشرفت بیماری اثرگذار باشند بسیار مهم است؛ زیرا می‌توان با تجویز عواملی که دارای اثرات مثبت اکتسابی و قابل تغییر هستند به درمان بیماری‌ها کمک نمود. از طرف دیگر، کنترل عوامل خطر قابل پیشگیری نیز می‌توانند در جلوگیری از پیشرفت بیماری مؤثر باشند. برای شناسایی رایج‌ترین روش‌ها استفاده از آمار کلاسیک است.

در این مطالعه با استفاده از آزمون کای دو ارتباط بین متغیرها ارزیابی شده است. به‌طور خلاصه متغیر Grade یک عامل مؤثر در متاستاز به غدد لنفاوی است. هم‌چنین بین متغیرهای مارکر ER، مارکر PR، مارکر P53، مارکر Her2 و مارکر Ki67 ارتباط معناداری وجود دارد. در مطالعات پیشین نیز به بررسی ارتباط عوامل مؤثر بر متاستاز به غدد لنفاوی پرداخته شده است.

فنگک یان لی و همکارانش در مقاله خود با بررسی پرونده پزشکی ۱۱۳۱ بیمار سرطان پستان دریافتند Grade، وضعیت PR و ER و Her2 و P53 به‌طور معنی‌داری با بیان Ki67 ارتباط دارد؛ در مطالعه‌ی حاضر نیز ارتباط معنی‌داری بین متغیر Ki67 و مارکر ER، مارکر PR، مارکر P53 وجود دارد اما ارتباط معنی‌داری با Her2 و Grade مشاهده نشد (۱۷). ددی هرمانس و همکارانش در سال ۲۰۲۱ با انجام مطالعه‌ی مقطعی بر روی اطلاعات ۵۱ پرونده‌ی پزشکی بیماران سرطان پستان، دریافتند طبق نتایج آزمون کای دو بین Grade و متاستاز به غدد لنفاوی در سرطان پستان ارتباط معنی‌داری وجود دارد، در این مطالعه‌ی حاضر نیز متغیر Grade ارتباط معناداری با متاستاز به غدد لنفاوی دارد (۱۸). شیانگ کویی و همکارانش در سال ۲۰۲۰ با مطالعه بر روی ۲۸۹۶۶ برای بررسی وضعیت غدد لنفاوی بیمار با سرطان پستان با برازش مدل رگرسیون لجستیک تک متغیره و سپس برازش رگرسیون لجستیک چند متغیره دریافتند متغیرهای

شانس متاستاز به غدد لنفاوی در بیمارانی که در سطح ۲ Grade بیماری هستند ۲/۲۳ برابر بیماران سطح ۱ Grade است.

شانس متاستاز به غدد لنفاوی در بیمارانی که در سطح ۳ Grade بیماری هستند ۱/۲۱ برابر (۲۱ درصد بیش‌تر از) بیماران سطح ۱ Grade است. طبق تحلیل رگرسیون لجستیک چند متغیره، سطح ۲ متغیر Grade ارتباط معنی‌داری با متاستاز به غدد لنفاوی دارد بنابراین متغیر Grade یک عامل تأثیرگذار بر متاستاز به غدد لنفاوی است.

جدول شماره ۳: نتایج حاصل از برازش مدل رگرسیون لجستیک تک متغیره با متغیر پاسخ متاستاز به غدد لنفاوی در بیماران سرطان پستان

متغیرها	ضریب	خطای استاندارد	سطح معنی‌داری	OR	95% CI for OR
Age	۰/۱۶	۰/۱۲۶	۰/۱۹	۱/۰۱۶	(۰/۹-۱/۰۴)
Marital Status	۰/۱۷	۰/۲۸	۰/۵۴	۱/۱۹	(۰/۶۸-۲/۰۶)
Tumor location 2	-۰/۶۱	۱/۲۴	۰/۶۳	۰/۵۵	(۰/۰۵-۶/۲۳)
Tumor location 1	۰/۱۸	۰/۲۷	۰/۵۱	۱/۲	(۰/۰۷-۲/۰۵)
Stage 3	۰/۵۹	۰/۶	۰/۳۲	۱/۸۱	(۰/۵۶-۵/۸۵)
Stage 2	۰/۲۳	۰/۲۸	۰/۴۱	۱/۲۵۵	(۰/۰۷۳-۲/۱۶)
Size	۰/۱۵	۰/۱۲	۰/۱۸	۱/۱۷	(۰/۹۳-۱/۵)
Skin nipple	-۰/۷۷	۰/۴۸	۰/۱۱	۰/۴۶	(۰/۱۸-۱/۲)
Margin	-۰/۳۹	۰/۶۴	۰/۳۸	۰/۶۸	(۰/۲۹-۱/۶)
DCIS 2	-۰/۱۳	۰/۳۳	۰/۶۹	۰/۸۸	(۰/۵-۱/۶۶)
DCIS 1	-۰/۱۶	۰/۳۳	۰/۶۳	۰/۸۵	(۰/۴۵-۱/۶۲)
Grade 3	۰/۱۹	۰/۴۳	۰/۶۵	۱/۲۱	(۰/۵۲-۲/۸)
Grade 2	۰/۸۴	۰/۳۳	۰/۰۱	۲/۳۲	(۱/۲۲-۴/۴۱)
Tubule formation	۰/۶۷	۰/۳۹	۰/۳۳	۱/۶	(۰/۰۷۴-۳/۵)
Tubule formation	-۰/۴۴	۰/۵۰	۰/۳۸	۰/۴۴	(۰/۲۴-۱/۷۲)
Nuclear pleomorphism 3	۰/۱۴	۰/۴۵	۰/۷۶	۱/۱۵	(۰/۵-۲/۸)
Nuclear pleomorphism 2	۰/۶۷	۰/۳۷	۰/۲	۱/۶۱	(۰/۰۸-۳/۳۲)
Mitosis 3	-۰/۱	۰/۳۸	۰/۷۸	۰/۹	(۰/۴۳-۱/۹)
Mitosis 2	۰/۲	۰/۳۵	۰/۵۶	۱/۲۲	(۰/۶۲-۲/۴۴)
PC 1	-۰/۸۸	۰/۱۳	۰/۵۱	۰/۸۲	(۰/۰۷-۱/۲)
PC 2	۰/۱۳	۰/۱۳	۰/۹۲	۱/۰۱	(۰/۰۸-۱/۳۲)

سطح اطمینان ۷۰ درصد

کم‌ترین سطح هر متغیر به عنوان رفرنس انتخاب شده است

جدول شماره ۴: نتایج حاصل از برازش مدل رگرسیون لجستیک چندگانه با متغیر پاسخ متاستاز به غدد لنفاوی در بیماران سرطان پستان

متغیرها	ضریب	خطای استاندارد	سطح معنی‌داری	OR	95% CI for OR
Age	۰/۲	۰/۱۳	۰/۱۶	۱/۰۲	(۰/۹۹-۱/۰۴)
Size	۰/۱۳	۰/۱۲۱	۰/۲۹	۱/۱۴	(۰/۰۸۹-۱/۴۴)
Yes Skin nipple	-۰/۶۸	۰/۵	۰/۱۸	۰/۵۱	(۰/۱۹-۱/۳۵)
Grade 2	۰/۸۴	۰/۳۴	۰/۰۱۲	۲/۳۲	(۱/۲۰-۴/۴۶)
Grade 3	۰/۲	۰/۴۵	۰/۶۸	۱/۲۱	(۰/۵۱-۲/۸۹)

سطح اطمینان ۹۵ درصد

کم‌ترین سطح هر متغیر به عنوان رفرنس انتخاب شده است

مطالعه حاضر نیز ارتباط معنی داری بین مارکر Ki67، Her2 و متاستاز به غدد لنفاوی یافت نشد (۲۱). کانگ و جیانگ و همکارانش در سال ۲۰۲۲ با بررسی بر روی ۱۵۲۷ بیمار سرطان پستان دریافتند متغیر سن و اندازه تومور ارتباط معنی داری با متاستاز به غدد لنفاوی دارد اما در مطالعه حاضر ارتباط معنی داری بین متغیر سن و اندازه تومور با متاستاز به غدد لنفاوی مشاهده نشده است (۲۲). منجولا کاپوروشوتام و همکارانش در سال ۲۰۲۱ در مطالعه‌ی خود با بررسی ۱۰۰ مورد بیمار سرطان پستان، دریافتند که بین اندازه‌ی تومور، Grade و Stage ارتباط معناداری با متاستاز به غدد لنفاوی در سرطان پستان وجود دارد و افزایش سه عامل اندازه‌ی تومور، stage و Grade با افزایش متاستاز به غدد لنفاوی همراه است. در مطالعه حاضر تنها بین Grade و متاستاز به غدد لنفاوی ارتباط معنی داری مشاهده شد (۲۳). هادی و جمال در مقاله خود در سال ۲۰۱۶ با بررسی بر روی ۱۵۰ بیمار سرطان پستان طبق تحلیل رگرسیون تک متغیره، به این نتیجه رسیدند که متغیر اندازه تومور یکی از متغیرهای معنی دار است. در مطالعه حاضر نیز متغیر اندازه‌ی تومور یکی از متغیرهای معنی دار در رگرسیون لجستیک تک متغیره است (۲۴).

سن، نژاد، Stage، محل اولیه تومور و زیرگروه بافت‌شناسی (شامل مارکرهای Ki67, P53, PR, ER و Her2) ارتباط معنی داری با وضعیت غدد لنفاوی دارند و تنها برای متغیر Grade ارتباط معنی داری مشاهده نشد. متغیر Stage دارای همبستگی مثبت با وضعیت غدد لنفاوی است؛ اما در مطالعه‌ی حاضر بین متغیر Grade و متاستاز به غدد لنفاوی ارتباط معنی داری یافت نشد (۱۹). هانگی چن و همکارانش در سال ۲۰۲۲ در مطالعه خود با مقایسه متغیرهای دارای تفاوت بین دو گروه با تجزیه و تحلیل تک متغیر و شناسایی عوامل مرتبط با متاستاز غدد لنفاوی زیر بغل با مدل رگرسیون لجستیک دریافتند غدد لنفاوی تهاجمی و سرطان در پستان چپ عوامل مؤثر بر متاستاز به غدد لنفاوی است هم‌چنین دریافتند مارکرهای Ki67، Her2، ER و PR طبق آزمون کای دو ارتباط معنی داری با درگیری غدد لنفاوی ندارد که مشابه نتایج حاصل از آزمون کای دو در مقاله حاضر است (۲۰).
مروه نیشا و همکارانش در مقاله‌ی خود در سال ۲۰۱۸ با بررسی ۷۵ بیمار سرطان پستان دریافتند ارتباط معنی داری بین Ki67، وضعیت یائسگی، Her2 و متاستاز به غدد لنفاوی وجود ندارد اما ارتباط معنی داری بین متغیر ki67، Grade و اندازه‌ی تومور وجود دارد در این

References

1. Yoshihara E, Smeets A, Laenen A, Reynders A, Soens J, Van Ongeval C, et al. Predictors of axillary lymph node metastases in early breast cancer and their applicability in clinical practice. *The Breast* 2013; 22(3): 61-357.
2. Sumbaly R, Vishnusri N, Jeyalatha S. Diagnosis of breast cancer using decision tree data mining technique. *International Journal of Computer Applications* 2014; (10)98.
3. Aydin Buyruk B, Kebapci N, Yorulmaz G, Buyruk A, Kebapci M. An Evaluation of Clinicopathological Factors Effective in the Development of Central and Lateral Lymph Node Metastasis in Papillary Thyroid Cancer. *J Natl Med Assoc* 2018; 110(4): 384-390.
4. Tarlan M, Khazaei S, Madani SH, Saleh E. Prognostic factors for cancer-specific survival in 220 patients with breast cancer: A single center experience. *Cancer Rep (Hoboken)* 2023; 6(1): e1675.
5. Dhimmar S, Nair A. Breast Cancer Detection Using Classification Algorithms. *IRGMETS* 2022; 4(7): 2582-5208.
6. Kanyılmaz G, Yavuz BB, Aktan M,

- Karaağaç M, Uyar M, Findik S. Prognostic Importance of Ki-67 in Breast Cancer and Its Relationship with Other Prognostic Factors. *Eur J Breast Health* 2019; 15(4): 256-261.
7. Nafissi N, Khayamzadeh M, Zeinali Z, Pazooki D, Hosseini M, Akbari ME. Epidemiology and histopathology of breast cancer in Iran versus other Middle Eastern countries. *Middle East Journal of Cancer* 2019; 9(3): 51-243.
 8. Keleş MK. Breast cancer prediction and detection using data mining classification algorithms: a comparative study. *Tehnički Vjesnik* 2019; 26(1): 149-155.
 9. Pati N, Panigrahi M, Patra KC. Evaluation of Different Paradigms of Machine Learning Classification for Detection of Breast Carcinoma. *Smart and Sustainable Technologies: Rural and Tribal Development Using IoT and Cloud Computing: Proceedings of ICSST 2021*: 349-356.
 10. Tinsley HE, Brown SD. Multivariate statistics and mathematical modeling. In *Handbook of applied multivariate statistics and mathematical modeling*. Academic Press 2000: 3-36.
 11. Tranmer M, Elliot M. Binary logistic regression. *Cathie Marsh for census and survey research*, paper 2008; 20.
 12. Agresti A. *Categorical data analysis*. John Wiley & Sons 2012 .
 13. Stoltzfus JC. *Logistic regression: a brief primer*. Academic emergency medicine 2011; 18(10): 1099-1104.
 14. Everitt B, Hothorn T. *An introduction to applied multivariate analysis with R*. Springer Science & Business Media 2011.
 15. Ishioka T. Imputation of missing values for unsupervised data using the proximity in random forests. In *International Conference on Mobile, Hybrid, and On-line Learning*. Nice 2013: 30-66.
 16. Sage AJ, Genschel U, Nettleton D. Random forest variable importance in the presence of missing data. *Random forest robustness, variable importance, and tree aggregation*. 2018: 37.
 17. Li FY, Wu SG, Zhou J, Sun JY, Lin Q, Lin HX, Guan XX, He ZY. Prognostic value of Ki-67 in breast cancer patients with positive axillary lymph nodes: a retrospective cohort study. *PLoS One* 2014; 9(2): e87264.
 18. Hermansyah D, Pricilia G, Azrah A, Rahayu Y, Paramita DA, Siregar ES. Correlation between Grading Histopathology and Sentinel Lymph Node Metastasis in Early Breast Cancer in University of Sumatera Utara Hospital. *Open Access Macedonian Journal of Medical Sciences* 2021; 9(B): 679-682.
 19. Cui X, Zhu H, Huang J. Nomogram for predicting lymph node involvement in triple-negative breast cancer. *Frontiers in Oncology* 2020; 10: 608334.
 20. Chen H, Meng X, Hao X, Li Q, Tian L, Qiu Y, Chen Y. Correlation Analysis of Pathological Features and Axillary Lymph Node Metastasis in Patients with Invasive Breast Cancer. *J Immunol Res* 2022; 2022: 7150304.
 21. Marwah N, Batra A, Marwah S, Gupta V, Shakya S, Sen R. Correlation of proliferative index with various clinicopathologic prognostic parameters in primary breast carcinoma: A study from North India. *J Cancer Res Ther* 2018; 14(3): 537-542.
 22. Jiang C, Xiu Y, Qiao K, Yu X, Zhang S, Huang Y. Prediction of lymph node metastasis in patients with breast invasive micropapillary carcinoma based on machine learning and SHapley Additive exPlanations framework. *Front Oncol* 2022; 12: 981059.
 23. Venkatesh PM. Association of Histopathological Parameters and Axillary Lymphnode Metastasis

- in Primary Breast Carcinoma. APJCC 2021; 6(4): 379-382.
24. Hadi NI, Jamal Q. Comparison of clinicopathological characteristics of lymph node positive and lymph node negative breast cancer. Pak J Med Sci 2016; 32(4): 863-868.