

تاریخ دریافت مقاله: ۹۶/۰۳/۱۱
تاریخ پذیرش مقاله: ۹۶/۰۵/۱۰

مروری بر روش‌های پیش‌بینی بقا در همبودی سرطان با استفاده از تکنیک‌های یادگیری ماشین

مجید نقی زاده

کارشناس ارشد، مهندسی کامپیوتر (نرم‌افزار)، دانشکده فنی مهندسی، دانشگاه آزاد اسلامی واحد اصفهان، ایران.
پست الکترونیکی: majid.naghizadeh@gmail.com

نرگس حبیبی*

استادیار، مهندسی کامپیوتر (هوش مصنوعی)، دانشکده فنی مهندسی، دانشگاه آزاد اسلامی واحد اصفهان، ایران.
پست الکترونیکی: n.habibi@khuisf.ac.ir

فرساد زمانی

استادیار، مهندسی کامپیوتر (هوش مصنوعی)، دانشکده فنی مهندسی دانشگاه آزاد اسلامی واحد اصفهان، ایران.
پست الکترونیکی: f.zamani@khuisf.ac.ir

چکیده

از سرطان‌های شایع را تشکیل می‌دهند و به همین دلیل، پیش‌گیری و تشخیص این نوع سرطان‌ها در افراد مبتلا بسیار حائز اهمیت است. یافته‌های اخیر نشان می‌دهد که برخی از بیماران در طول بیماری ممکن است دچار همبودی سرطان، به معنای دارا بودن دو یا چند بیماری سرطان به صورت همزمان، باشند. این بیماران، نسبت به سایر بیمارانی که تنها به یک نوع سرطان مبتلا هستند، از شانس کمتری برای بقا برخوردارند. در این مقاله، به مروری بر تحقیقات انجام شده در زمینه پیش‌بینی میزان بقا در همبودی سرطان، با استفاده از تکنیک‌های یادگیری ماشین پرداخته شده است.

واژه‌های کلیدی: تصمیم‌گیری پزشکی، همبودی

سرطان، یادگیری ماشین، پیش‌بینی بقا، دادگان SEER^۲.

سرطان یک مشکل رایج در کلیه کشورهای جهان است. مطابق اطلاعات و آمار مرکز تحقیق سرطان بریتانیا^۱، در سال ۲۰۱۲ تعداد موارد جدید ابتلا به سرطان در جهان بیش از ۱۴/۱ میلیون نفر بوده که بیش از ۸/۲ میلیون مورد آن منجر به فوت شده است. همچنین مطابق اطلاعات و آمار انستیتوی ملی سرطان^۲ تنها در ایالات متحده آمریکا تعداد موارد جدید ابتلا به سرطان در سال ۲۰۱۶ بیش از ۱ میلیون و ششصد هزار نفر بوده است که از این میان بیش از ۳۵ درصد موارد فوت خواهند شد. سرطان سینه و دستگاه تناسلی (در خانم‌ها) و پروستات (در آقایان) در حدود ۳۰ درصد

* نویسنده مسئول

1- www.cancerresearchuk.org
2- www.cancer.gov

3- Surveillance, Epidemiology, and End Results Program

مانند [SVM[۷-۳], RF^۴[۸, ۷], ANN[۱۰-۸, ۵, ۴], ۹, ۶], KNN^۵[۵], DT[۱۲, ۹, ۸, ۶-۴], J48^۶[۴], LR^۷[۱۱-۸], MLP^۸ و PSO^۹ اشاره نمود که هر یک از آن‌ها مزایا و معایب مختص خود را در عمل داشته‌اند.

پس از اعمال روش‌های گوناگون یادگیری ماشین بر روی داده‌های سرطانی مختلف، و انجام رده بندی و پیش‌بینی‌های مورد نظر، لازم است نتایج بر اساس معیارهای خاصی مورد سنجش و ارزیابی قرار گیرد که از آن جمله می‌توان به معیارهایی مانند صحت^{۱۰}، فراخوانی^{۱۱} و دقت^{۱۲} اشاره نمود.

بر اساس بررسی‌های انجام شده، مقاله حاضر اولین مقاله مروری در حوزه پیش‌بینی بقا در همبودی سرطان در سطح بین‌المللی است. به‌طور کلی بحث همبودی بیماری‌ها مبحثی است که در حوزه انفورماتیک پزشکی به‌طور کافی به آن پرداخته نشده است.

تحقیقات بررسی شده در این مقاله توسط موتور جستجوی گوگل اسکولار گردآوری شده اند. کلمات کلیدی مورد استفاده در جستجو عبارت بودند از، cancer، comorbidity و survival. به منظور انجام یک پژوهش جامع، محدوده زمانی برای جستجو در نظر گرفته نشده است.

بحث بررسی بقا در همبودی سرطان برای اولین بار در تحقیق آقای ذوالبنین و همکاران در سال ۲۰۱۵ مطرح گردید [۸].

۲- مرور تحقیقات انجام شده

در این بخش ابتدا مروری بر مفاهیم یادگیری ماشین و داده‌کاوی صورت می‌گیرد. سپس نمونه‌هایی از کاربرد تکنیک‌های یادگیری ماشین در زمینه تشخیص سرطان، پیش‌بینی بقا، و به‌ویژه همبودی سرطان ارائه می‌شوند.

- 4- Random Forest
- 5- MultiLayer Perceptron
- 6- Logistic Regression
- 7- A kind of decision tree
- 8- k-Nearest Neighbors algorithm
- 9- Particle Swarm Optimization
- 10- Precision/Sensitivity
- 11- Recall/specificity
- 12- Accuracy

سرطان نامی است که به مجموعه بیماری‌هایی اطلاق می‌شود که از تکثیر مهار نشده سلول‌ها پدید می‌آیند. در یک جاندار سالم، همیشه بین میزان تقسیم سلول و مرگ طبیعی سلولی یک تعادل وجود دارد. سلول‌های سرطانی از سازوکارهای عادی تقسیم و رشد سلول‌ها جدا می‌افتند. علت دقیق این پدیده نامشخص است ولی احتمال دارد عوامل ژنتیکی یا مواردی که موجب اختلال در فعالیت سلول‌ها می‌شوند در هسته سلول اشکال وارد کنند، مانند مواد رادیو اکتیو، مواد شیمیایی و سمی یا تابش بیش از حد اشعه‌هایی مانند نور آفتاب.

همبودی یعنی هر موجودیت بالینی اضافه و مجزا که از ابتدا وجود داشته و یا در حین دوره درمانی بیماری مورد نظر بیمار، رخ می‌دهد [۱].

همبودی سرطان زمانی اتفاق می‌افتد که یک فرد دارای دو نوع یا بیشتر سرطان باشد. در طب، همبودی حضور یک یا چند بیماری یا اختلال است که همزمان (همزمان و یا موازی) با یک بیماری اولیه یا اختلال رخ می‌دهد که ممکن است جسمی، رفتاری یا روانی باشد. در پزشکی، همبودی، اثر تمام بیماری‌های دیگر یک بیمار که ممکن است از بیماری اولیه بهره گیرد، توصیف می‌شود [۲].

همبودی در میان بیماران سرطانی شایع است و با افزایش سن، احتمال ابتلا افزایش می‌یابد. همبودی در مرحله تشخیص و درمان افراد مبتلا به سرطان، به‌طور بالقوه توسعه و بهبود یافته است. با وجود رابطه نزدیک بین همبودی و سرطان، اتفاق نظر کمی در مورد چگونگی تفسیر یا مدیریت همبودی در زمینه سرطان وجود دارد. از جمله مشکلات مربوط به همبودی، محاسبه احتمال زنده ماندن بیمار برای پزشکان می‌باشد [۲].

در زمینه تخمین و برآورد شانس بقا در بیماران مبتلا به همبودی سرطان از تکنیک‌های گوناگون یادگیری ماشین استفاده شده است که از آن جمله می‌توان به تکنیک‌هایی

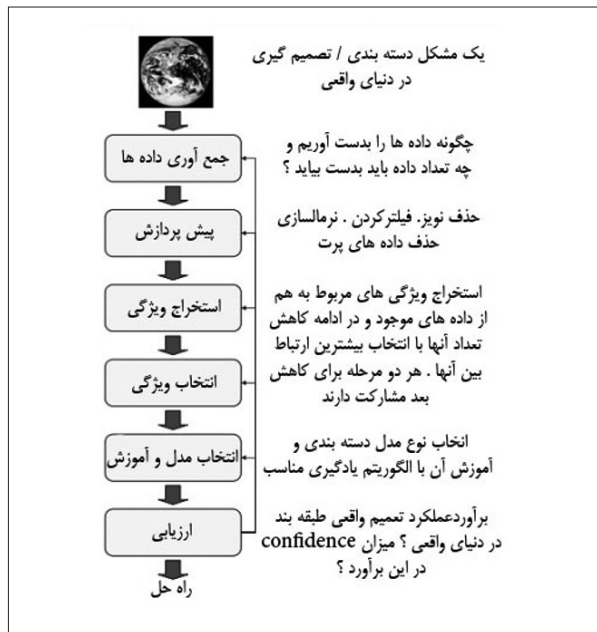
۲-۱ یادگیری ماشین و داده کاوی

یادگیری ماشین را به عنوان مجموعه‌ای از روش‌هایی تعریف می‌کنیم که به‌طور خودکار می‌توانند الگوهای موجود در داده‌ها را تشخیص داده و سپس با استفاده از الگوهای مخفی برای پیش‌بینی داده‌های آینده و یا انجام انواع دیگری از تصمیم‌گیری در شرایط عدم اطمینان استفاده کنند. آموزش بدان معنی است که از مشاهدات دانش نوظهوری تولید شود که این دانش برای دستیابی به اهداف از پیش تعریف شده است. داده خود به تنهایی، نوعی دانش است. اما برای برخی از برنامه‌ها و برای درک بهتر بشر، مجموعه داده‌های بزرگ نمی‌توانند به‌طور مستقیم به فرم خام خود کاربردی باشند. یادگیری از داده بدان معنی است که دانش جدید چکیده شده‌ای از مقدار زیادی اطلاعات، استخراج شود [۱۳].

الگوریتم‌های یادگیری ماشین به‌طور کلی به سه دسته یادگیری با ناظر، یادگیری بدون ناظر و یادگیری تقویت شده طبقه‌بندی می‌شوند. از جمله معروف‌ترین الگوریتم‌های یادگیری ماشین می‌توان به LR [۸-۱۱]، درخت تصمیم [۴-۶، ۸، ۹، ۱۲]، SVM [۳-۷]، KNN [۴]، Naive Bayes [۵] و جنگل تصادفی [۷، ۸] اشاره نمود.

داده‌کاوی فرایندی است که فرد را قادر می‌سازد که بینش اساسی و دانش از داده‌های عظیم به دست آورد و از آن در تصمیم‌گیری در فعالیت‌های تجاری مهم استفاده کند [۱۴]. در تعریفی دیگر اصطلاح داده‌کاوی به فرایند نیمه‌خودکار تجزیه و تحلیل پایگاه داده‌های بزرگ به منظور یافتن الگوهای مفید اطلاق می‌شود [۱۵]. به‌طور خلاصه داده‌کاوی یعنی کاوش در یک مجموعه داده بزرگ به منظور خارج کردن دانش و اطلاعات قابل توجه از بین آن‌ها. از جمله الگوریتم‌های پرکاربرد داده‌کاوی می‌توان به AdaBoost، PageRank، EM، Apriori، C4.5 و CART اشاره نمود.

منظور از یادگیری ترکیبی، ترکیب نتایج دو یا



شکل ۱: مراحل حل مسئله با استفاده از تکنیک‌های یادگیری ماشین [۱۸]

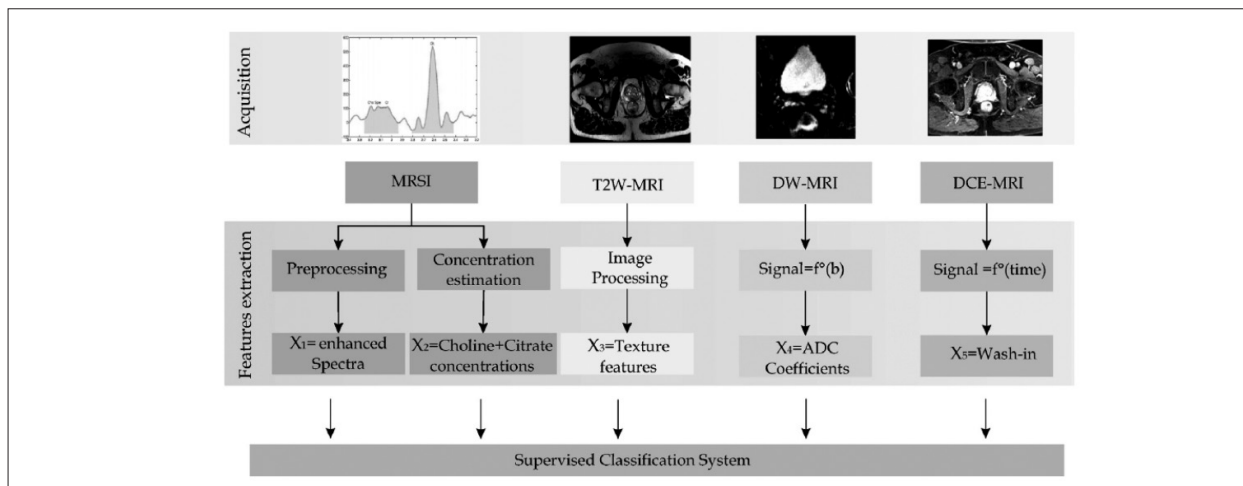
چند الگوریتم جهت ترکیب خروجی‌ها و ارائه یک پاسخ بهینه است. از یادگیری ترکیبی در بسیاری از کاربردهای مبتنی بر یادگیری ماشین استفاده می‌شود [۱۶]. تاکنون تکنیک‌های گوناگونی برای یادگیری ترکیبی پیشنهاد شده است که از آن جمله می‌توان به تکنیک‌های Boosting و Bootstrap aggregating (bagging) اشاره نمود.

در هنگام حل بسیاری از مشکلات در یادگیری ماشین، ویژگی‌های ورودی بسیار زیادی وجود دارد. با این حال همه آن ویژگی‌ها برای حل مشکل مناسب نیستند و در بسیاری از موارد انتخاب ویژگی‌های بی‌ربط باعث از بین رفتن کارایی مدل آموزشی می‌شود. انتخاب ویژگی یکی از مراحل حساس و مهم در ایجاد یک مدل آموزشی مناسب می‌باشد [۱۷].

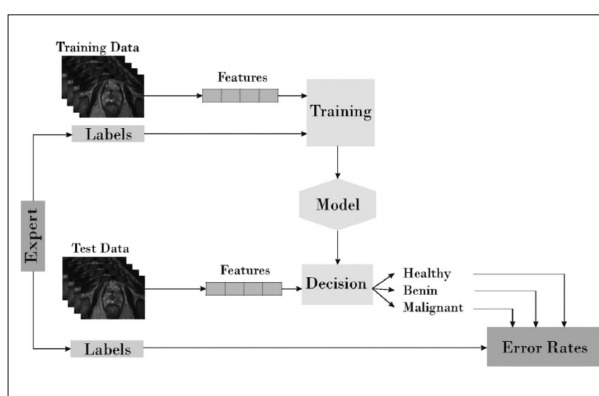
شکل (۱) مراحل حل یک مسئله با استفاده از تکنیک‌های یادگیری ماشین را نشان می‌دهد [۱۸].

۲-۲ کاربرد یادگیری ماشین در زمینه بیماری سرطان

در سال‌های اخیر استفاده از تکنیک‌های مختلف یادگیری ماشین، الگوریتم‌های انتخاب ویژگی و الگوریتم‌های



شکل ۲: فلوجارت اجمالی از استخراج ویژگی و ترکیب آن‌ها در سیستم طبقه‌بندی تحت نظارت در روش ترکیبی و همکاران [۷]



شکل ۳: طرح کلی یک سیستم طبقه‌بندی تحت نظارت SVM یا RF [۷]

خوش‌خیم، بدخیم و سالم دسته‌بندی کردند. پژوهشگران از برداری شامل ۱۸۲ ویژگی برای هر بیمار به‌عنوان ورودی به الگوریتم‌های مورد بررسی خود، استفاده نمودند. آن‌ها همچنین برای حل مشکل دودویی بودن دسته‌بندی SVM، از روش یکی در برابر همگی^{۱۶} استفاده نمودند. فلوجارت اجمالی از استخراج ویژگی و ترکیب آن‌ها در سیستم طبقه‌بندی تحت نظارت این پژوهشگران، در شکل (۲) نمایش داده شده است.

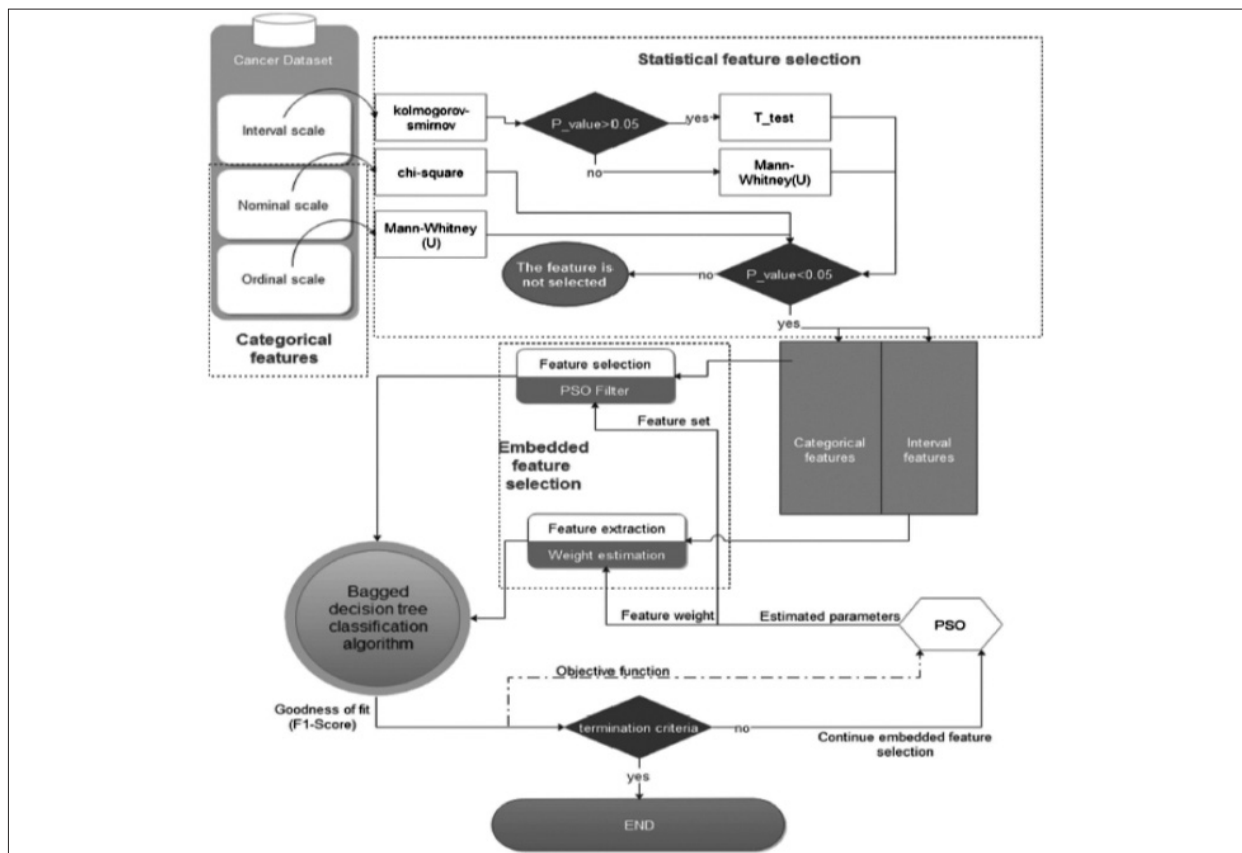
این پژوهشگران با به‌کارگیری طبقه‌بند SVM، نرخ خطای سراسری^{۱۷} ۱٪ و حساسیت^{۱۸} ۹۹/۱٪ را به دست آوردند. همچنین برای مقایسه، با استفاده از الگوریتم RF به نرخ خطای ۱۸/۲٪ و حساسیت ۷۲٪ دست یافتند. در نهایت

یادگیری ترکیبی برای کمک به حل مسائل بیوانفورماتیک بسیار رایج گردیده و کمک شایان توجهی نیز به علم پزشکی نموده است.

تریگویی و تعدادی از پژوهشگران^{۱۳} [۷] در سال ۲۰۱۷، با توجه به نقش موثر^{۱۴} MRI (تصویرسازی تشدید مغناطیسی: روشی است که می‌توان با کمک گرفتن از آن تصاویر بسیار دقیق و واضحی از اندام‌های درون بدن به دست آورد.) از نوع چند پارامتر و^{۱۵} MRSI (تصویربرداری اسپکتروسکوپی رزونانس مغناطیسی: یک روش تصویربرداری ویژه که به وسیله آن می‌توان بدون استفاده از سوزن و صدمات ناشی از آن به محتوای داخلی ضایعات در داخل بدن پی برد. در این روش پیچیده عکسبرداری، پس از تفسیر کامپیوتری یافته‌ها، پزشکان متخصص این قادر به ارزیابی ضایعه خواهند بود. به این روش، نمونه‌برداری غیرتهاجمی نیز گفته می‌شود.) در تشخیص و جایگزینی بافت‌های سرطانی، روشی جدید بر مبنای یادگیری ماشینی، ارائه کردند. سرطان پروستات به‌عنوان سومین و ششمین علت مرگ و میر مردان در کشورهای توسعه یافته و در حال توسعه شناخته شده است. در این مقاله پژوهشگران، با استفاده از طبقه‌بند SVM و RF براساس طرح طبقه‌بندی نظارت شده، با تحقیق به روی ۳۴ بیمار با میانگین سنی ۶۴ سال، آن‌ها را به سه رده

16- one-versus-all
17- Global error rate
18- Sensitivity

13- Trigui R, Mitéran J, Walker P, Sellami L, Hamida AB
14- Magnetic Resonance Imaging
15- Magnetic Resonance Spectroscopic Imaging



شکل ۴: ساختار سیستم پیش آگهی پیشنهاد شده HPBCR توسط محبیان و همکاران [۶]

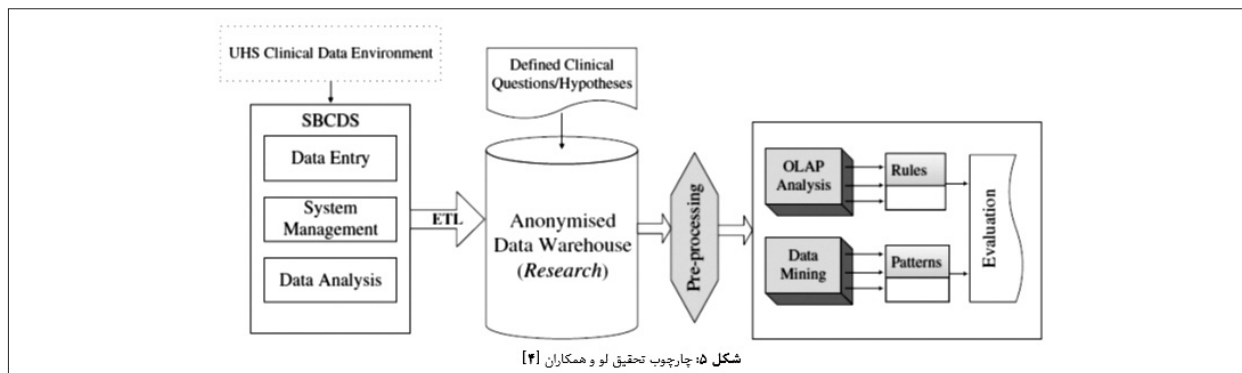
به عنوان ورودی سیستم طبقه‌بندی با یادگیری گروهی BDT^{۲۱} بهبود یافتند. در ادامه ترکیبی مناسب از ویژگی‌های طبقه‌بندی شده، انتخاب شده و همچنین اهمیت ویژگی‌های انتخاب شده با سه پارامتر فاصله، اندازه‌گیری و مقیاس، توسط الگوریتم PSO ارزیابی و شناسایی شد. همچنین پژوهشگران در این مقاله از ۳ طبقه‌بند SVM، ANN و DT برای مقایسه نتایج آن‌ها با نتایج خود، استفاده کردند. ویژگی‌هایی که برای پیش‌بینی عود توسط این پژوهشگران استفاده شد شامل سن تشخیص، اندازه تومور، سن شروع قاعدگی، تعداد حاملگی و تعداد گره‌های درگیر و جدا شده به عنوان ویژگی‌های بازه‌ای و وضعیت گیرنده استروژن و پروژسترون، نسبت درگیری غدد لنفاوی، تعداد غدد لنفاوی درگیر زیر بغل، سابقه خانوادگی سرطان و استفاده از هورمون درمانی و پرتودرمانی به عنوان ویژگی‌های اسمی و همچنین درجه تومور به عنوان یک ویژگی ترتیبی،

نیز با نمایش نتایج دسته‌بندی با فرم نقشه کد رنگی (یک ابزار تشخیص کامپیوتری است که می‌تواند در ارزیابی نتایج کمک کند). علاوه بر ارزیابی نتایج، یک برآورد از شکل تومور و حجم آن ارائه کردند. طرح کلی یک سیستم طبقه‌بندی تحت نظارت SVM یا RF برای به دست آوردن نرخ خطا در شکل (۳) نمایان شده است.

محبیان و همکاران [۶] در سال ۲۰۱۶، از یک روش ترکیبی برای پیش‌بینی ۵ ساله عود سرطان سینه، با استفاده از بهینه‌سازی آموزشی گروهی با نام HPBCR^{۲۰} ارائه کردند. پیش‌بینی عود سرطان سینه می‌تواند برای ارائه طرح درمان مناسب، مورد استفاده متخصصان باشد. این الگوریتم روی ۵۷۹ بیمار مبتلا به سرطان سینه با شیوع عود ۳/۱۹٪ برای تجزیه و تحلیل مورد بررسی واقع شده است. ویژگی‌های تشخیصی، با استفاده از روش‌های

انتخابی، انتخاب شده و سپس با استفاده از الگوریتم PSO
 19- Mohebian MR, Marateb HR, Mansourian M, Mañanas MA, Mokarian F
 20- A Hybrid Computer-aided-diagnosis System for Prediction of Breast Cancer Recurrence

21- Bagged Decision Tree



شکل ۵: چارچوب تحقیق لو و همکاران [۴]

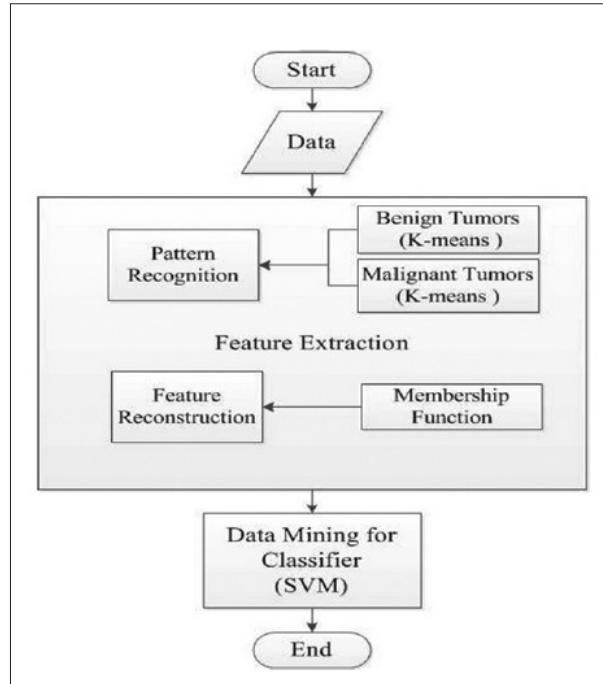
شکل ۵: چارچوب تحقیق لو و همکاران [۴]

شده، استفاده کردند. همچنین پنج مدل طبقه‌بندی با هدف طبقه‌بندی بیماران بر اساس ویژگی‌های انتخاب شده در مورد آن‌ها استفاده شده و مورد ارزیابی واقع شده است. مدل‌هایی شامل Naïve Bayes، ANN، MLP، SVM، LR و J48. این تحقیق، فرایند داده‌کاوی و استفاده از تکنیک‌های مختلف آن را برای تبدیل مقادیر زیادی از داده‌های بیماران به اطلاعات مفید و الگوهای ارزشمند برای کمک به درک نتایج سرطان را نمایش می‌دهد. روندنمای روش پیشنهادی این پژوهشگران در شکل (۵) نمایش داده شده است.

با توجه به این‌که فیلتر کردن تمام اطلاعات و ویژگی‌های مربوط به تشخیص بیماری بالینی یک کار چالش برانگیز و وقت‌گیر است، ژنگ و همکاران [۳] در سال ۲۰۱۴، روشی جدید برای استخراج ویژگی برای تشخیص سرطان سینه بر اساس ویژگی‌های استخراج شده تومور، ارائه نمودند. استخراج ویژگی‌ها و انتخاب آن‌ها در کیفیت طبقه‌بندی‌هایی که از طریق روش‌های داده‌کاوی ایجاد می‌شود، بسیار حیاتی هستند. در این مقاله پژوهشگران با استفاده از ترکیبی از الگوریتم‌های SVM و خوشه‌بندی K-Means، الگوریتم K-SVM را برای استخراج اطلاعات مفید و یافتن الگوهای پنهان برای تشخیص تومورهای خوش‌خیم از بدخیم معرفی کردند. در ابتدا الگوریتم K-means برای تشخیص الگوهای مخفی تومورهای خوش‌خیم و بدخیم به صورت جداگانه استفاده شده، عضویت هر تومور به این الگوها محاسبه و به‌عنوان یک ویژگی جدید در مدل

می‌باشند. به‌طور خلاصه در این روش ویژگی‌های ورودی ابتدا توسط تکنیک انتخاب ویژگی آماری انتخاب می‌شوند. در ادامه ویژگی‌های انتخاب شده توسط درخت تصمیم Bagged برای ساختن طبقه‌بند استفاده شده و مجموعه ویژگی‌های بهینه به دست آمده و وزن ویژگی‌ها با استفاده از الگوریتم PSO در طول یادگیری تخمین زده می‌شود. در نهایت الگوریتم زمانی متوقف می‌شود که عملکرد تابع هدف، بهبود قابل توجهی نداشته یا به حداکثر تعداد تکرارها (تعداد ۱۰۰ عدد در مقاله مورد اشاره) رسیده باشد. حداقل دقت به دست آمده از این روش ۹۵٪ می‌باشد که دقت قابل قبولی نسبت به سایر روش‌های بحث شده در آن مقاله می‌باشد. طرح کلی پیاده‌سازی شده توسط پژوهشگران این تحقیق در شکل (۴) نمایش داده شده است.

لو و همکاران [۴] در سال ۲۰۱۵ مطالعه موردی استفاده از تکنیک‌های داده‌کاوی در تجزیه و تحلیل، تشخیص و درمان رویدادهای مربوط به بیماری سرطان سینه را ارائه نمودند. پژوهشگران در این مقاله داده‌های مربوط به بیش از ۱۶۰۰۰ بیمار را از پایگاه داده سرطان بدون ترتیبی مربوط به دانشکده بیمارستانی ساوتهمپتون با نام SBCDS^{۳۳} جمع‌آوری و پردازش کردند. سپس چندین روش داده‌کاوی را با استفاده از نرم‌افزار WEKA بر روی داده‌ها اجرا کردند. به‌طور خاص پژوهشگران از الگوهای ترتیبی برای کشف الگوهای تکرار شونده از اطلاعات بیماران بر اساس گروه‌های زندگی و بیماران فوت



شکل ۶: چارچوب کلی طرح تحقیقی ژنگ و همکاران [۳]

آموزش تلقی می‌گردد. بعد از آن از یک ماشین بردار پشتیبان برای به دست آوردن یک طبقه‌بندی جدید برای متمایز ساختن این تومور ورودی به سیستم، استفاده شده است. بر اساس اعتبار سنجی متقاطع 10-Fold، روش پیشنهادی دقت را به ۹۷/۳۸٪ افزایش داد. همچنین زمانی که این مدل بر روی داده‌های سرطان سینه ویسکانسین (WDBC) دانشگاه کالیفرنیا آزمایش شد، شش ویژگی از ۳۲ ویژگی اصلی برای مرحله آموزش استخراج شدند که نتایج نه تنها نشان دهنده قابلیت روش پیشنهادی در تشخیص سرطان سینه بوده، بلکه موجب صرفه‌جویی در زمان در طول فاز آموزش نیز شده است. چارچوب کلی طرح این پژوهشگران در شکل (۶) قابل مشاهده است.

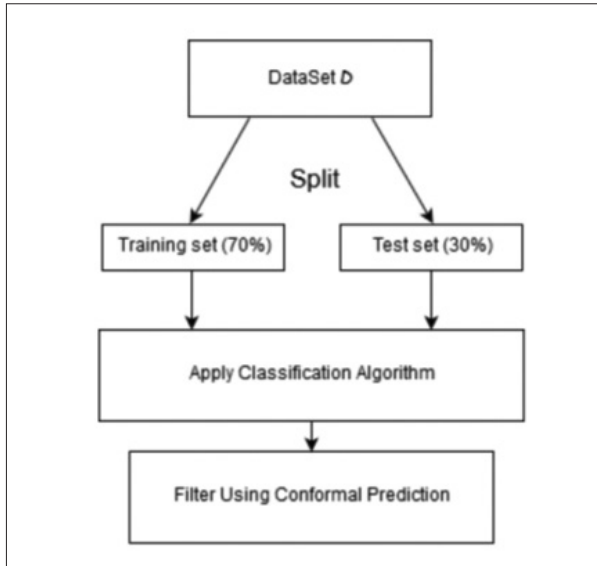
هارتای و همکاران [۱۱] در سال ۲۰۱۵، روند نرخ مرگ و میر در اثر سرطان دستگاه تناسلی زنان را مورد بررسی قرار دادند. داده‌های این پژوهش مربوط به تعداد مرگ ناشی از سرطان که در کشور مجارستان ثبت شده، می‌باشد. این پژوهشگران از روش رگرسیون استفاده کرده و با توجه به طبقه‌بندی‌هایی که انجام دادند به این نتیجه

رسیدند که اکثر مرگ و میرهای ناشی از سرطان دستگاه تناسلی زنان در زمستان رخ داده است. پژوهشگران در این مطالعه نشان دادند که یک اثر فصلی قابل توجهی در اندام تناسلی زنان و مرگ و میر ناشی از آن وجود دارد.

چاراسیا و پال [۱۰] در سال ۲۰۱۴، با استفاده از تکنیک‌های داده‌کاوی روشی جدید برای پیش‌بینی و حل مشکل ادامه زندگی (بقا) در افراد مبتلا به سرطان سینه ارائه دادند. سرطان سینه یکی از مرگبارترین و شایع‌ترین بیماری‌ها است (بیش از ۱/۶٪). این سرطان علت اصلی مرگ و میر زنان در سراسر جهان می‌باشد. این پژوهشگران عوامل خطرناک ابتلا به سرطان سینه را به دو دسته عوامل تغییرپذیر و عوامل تغییرناپذیر تقسیم کردند. عوامل خطر تغییرناپذیر شامل سن، جنسیت، تعداد بستگان درجه اول مبتلا به سرطان سینه، سابقه قاعدگی و سن یائسگی می‌باشد. در حالی که عوامل خطر تغییرپذیر شامل سن اولین تولد فرزند، تعداد فرزندان، مدت تغذیه با شیر مادر، میزان مصرف الکل، رژیم غذایی و تعداد سقط جنین می‌باشد. پژوهشگران در این مقاله یک سیستم تشخیص برای سرطان سینه براساس DT (REP TREE) و ANN (Network RBF) و Simple Logistic پیشنهاد کرده‌اند. در مرحله آزمایش اولیه از روش 10-Fold برای اعتبارسنجی استفاده شده و عملکرد سیستم در پایگاه داده مرکز پزشکی دانشگاه لیوبلیانا^{۲۸} یوگسلاوی، انستیتوی انکولوژی، مورد بررسی قرار گرفته است. نرخ طبقه‌بندی صحیح سیستم پیشنهادی ۷۴/۵٪ است. این پژوهشگران نشان دادند که با استفاده از منطق لجستیک ساده می‌توان ابعاد ویژگی‌ها را کاهش داد و از الگوریتم‌های ساده داده‌کاوی REP TREE و Network RBF می‌توان برای به دست آوردن سیستم تشخیص خودکار سریع برای بیماری‌های فراگیر استفاده کرد. در این مقاله بهترین الگوریتم بر اساس اطلاعات بیمار، طبقه‌بند منطق لجستیک ساده با دقت ۷۴/۴۷٪ و زمان کل ساخت مدل ۰/۶۲ ثانیه

27- Chaurasia V, Pal S
28- Ljubljana

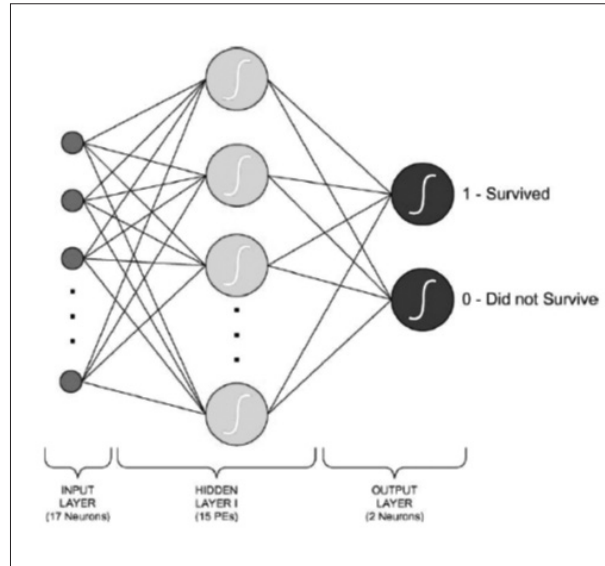
25- Wisconsin Diagnostic Breast Cancer
26- Hartai M, Nyári TA, McNally RJ



شکل ۱۱: فلوجارت روش پیشنهادی النمر و همکاران [۵]

الگوریتم درخت تصمیم و روش bagging برای پیش‌بینی بقای بیماران مبتلا به آن در طی ۵ سال ارائه کردند. پس از پیش‌پردازش داده‌ها از مجموعه داده‌های سرطان پستان SEER، بدیهی است که طبقه‌بندی توزیع داده‌ها نامتوازن است. پس به منظور کاهش معیار عملکرد مدل‌های ناشی از داده‌های نامتوازن نمونه برداری انجام شده است. همچنین از الگوریتم Bagging برای ساخت یک مدل درخت تصمیم‌گیری برای پیش‌بینی سرطان سینه استفاده شده است. عملکرد مدل‌ها در حالتی که توزیع داده‌ها تقریباً برابر است، در بهترین حالت خود قرار دارد. عملکرد مدل‌ها توسط AUC^{30} (معیار مهمی که برای تعیین میزان کارایی یک دسته‌بند استفاده می‌شود) تحت منحنی ROC^{31} (این نمودار روشی برای بررسی کارایی دسته‌بندها می‌باشد)، ارزیابی شدند که با ضریب زیر نمونه‌برداری ۱۵٪، AUC مدل 0.7484 است اما پس از استفاده از الگوریتم Bagging، AUC مدل به 0.7678 افزایش یافته است.

دلن و همکاران [۹]^{۳۲} در سال ۲۰۰۵، تعداد ۲۰۲،۹۳۲ پرونده بیماران مبتلا به سرطان سینه را از پایگاه داده SEER که با الگوریتم‌های طبقه‌بندی به دو گروه «ادامه



شکل ۱۰: نمایش گرافیکی مدل ANN MLP دلن و همکاران [۹]

است. این پژوهشگران نشان دادند که در میان الگوریتم یادگیری ماشین، تابع طبقه‌بندی منطق لجستیک ساده توانایی بهبود قابل توجه روش‌های طبقه‌بندی متعارف مورد استفاده در مطالعه را دارد. همچنین نشان دادند که مهم‌ترین ویژگی‌ها در بقای بیماران مبتلا به سرطان سینه، درجه بدخیمی، اندازه تومور، سن بیمار، یائسگی، سرطان در هر دو سینه و برخی دیگر از ویژگی‌ها می‌باشند که این ویژگی‌ها با استفاده از سه آزمون برای ارزیابی متغیرهای ورودی، شامل آزمون‌های Chi-square، Gain و آزمون Ratio Gain مورد بحث و بررسی قرار گرفتند.

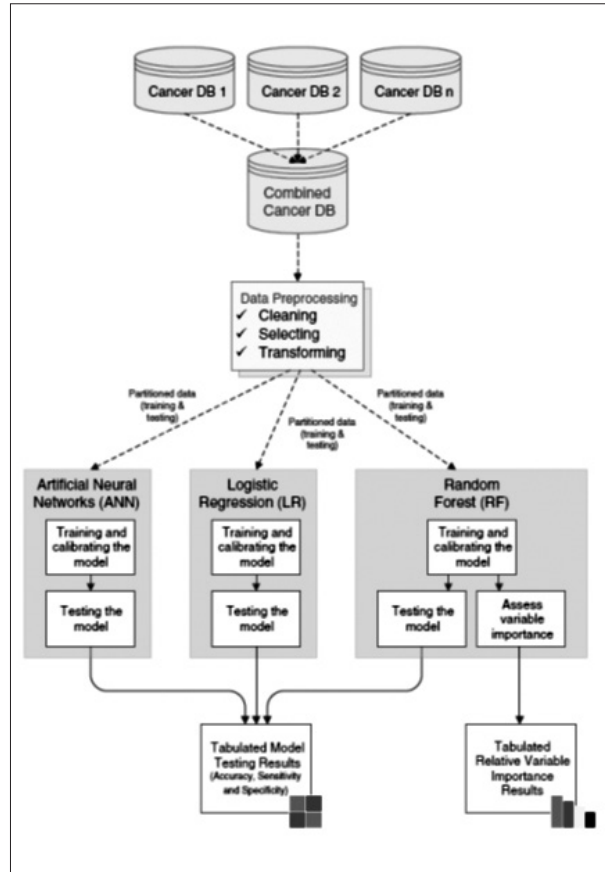
یاکین و همکاران [۱۲]^{۲۹} نیز در سال ۲۰۰۹، روشی جدید بر روی داده‌های نامتوازن سرطان سینه SEER (برنامه ملی مؤسسه سرطان، اطلاعاتی را در مورد آمار سرطان در تلاش برای کاهش بار سرطان در میان جمعیت ایالات متحده، فراهم می‌کند. SEER توسط برنامه پژوهش و نظارت، پشتیبانی می‌شود. این برنامه رهبری ملی را در علم نظارت بر سرطان و همچنین ابزارهای تحلیلی و متدولوژی در جمع‌آوری، تحلیل، تفسیر و انتشار آمار قابل اعتماد مبتنی بر جمعیت، بر عهده دارد.) با استفاده از

30- Area Under Curve

31- Receiver Operating Characteristic

32- Delen D, Walker G, Kadam A

29- Ya-Qin L, Cheng W, Lu Z



شکل ۱۲: روش تحقیق ذوالبنین و همکاران [۸]

بقا، (۹۳،۲۷۳) و «فوت شده» (۱۰۹،۶۵۹) تقسیم شدند با الگوریتم‌های ANN، درخت تصمیم C5 و روش آماری منطق لجستیک ساده برای توسعه مدل‌های پیش‌بینی با استفاده از مجموعه داده‌های بزرگ، مورد بررسی قرار دادند. در انتها نیز از روش 10-Fold برای اندازه‌گیری برآورد غیرمنطقی از سه مدل پیش‌بینی برای مقایسه عملکرد استفاده کرده‌اند. نتایج حاصل از پیش‌بینی بقا در محدوده دقت ۸۹/۲٪ تا ۹۳/۶٪ بود که در این بین درخت تصمیم بهترین نتایج را نسبت به ادبیات مورد بررسی در مقاله و منطق لجستیک ساده بدترین نتایج را در بین ۳ الگوریتم مورد بررسی برای پژوهشگران به دست آورد. نمایش گرافیکی مدل پیشنهادی این پژوهشگران در شکل (۱۰) نمایش داده شده است.

النمر و همکاران [۵] در سال ۲۰۱۶، روشی دیگر برای پیش‌بینی بقای افراد مبتلا به سرطان سینه بر روی

داده‌های SEER ارائه کردند. پژوهشگران در این مقاله، الگوریتم پیش‌بینی کننده متعارف CP^{۲۴} (الگوریتمی که برای محاسبه سطح اطمینان هر پیش‌بینی با استفاده از پیش‌بینی‌های قبلی استفاده می‌شود و می‌توان آن را با هر تکنیک طبقه‌بندی مورد استفاده قرار داد.) را برای حذف پیش‌بینی‌های غیر قابل اعتماد و بهبود نتایج طبقه‌بندی چهار الگوریتم یادگیری ماشین KNN، DT، ANN و SVM اعمال نمودند. استفاده از روش پیشنهادی، تعداد پیش‌بینی‌های کاذب را کاهش و حساسیت، دقت و صحت نتایج را بهبود داده است. روندنمای روش پیشنهادی این پژوهشگران در شکل (۱۱) به نمایش گذاشته شده است.

۳-۲ کاربرد یادگیری ماشین در زمینه همبودی سرطان

ذوالبنین و همکاران [۸] در سال ۲۰۱۵، روشی جدید برای پیش‌بینی بقای کلی در همبودی سرطان با استفاده از چهار تکنیک داده‌کاوی ANN، RF، DT و LR ارائه کردند. سرطان و دیگر بیماری‌های مزمن، بخش قابل توجهی از هزینه‌های مراقبت بهداشتی را در ایالات متحده به خود اختصاص داده‌اند. پژوهشگران در این پژوهش، با استفاده از داده‌های سرطانی SEER به ایجاد دو مجموعه داده پرداختند. مجموعه داده اول برای پیش‌بینی نرخ بقا در همبودی سرطان سینه و دستگاه تناسلی در زنان و مجموعه داده دوم برای تشخیص نرخ بقا در همبودی سرطان دستگاه تناسلی و سیستم ادراری در مردان بوده است. پژوهشگران پس از پاکسازی داده‌ها با حذف داده‌های اضافی و پرت و نرمال‌سازی آن‌ها در مرحله پیش‌پردازش، با انتخاب ویژگی‌های احتمالی موثر در مرگ یا زنده ماندن بیماران به طریقی ساده، برای مدل‌سازی از چند روش مختلف یادگیری ماشین استفاده کردند. از آن میان الگوریتم RF که الگوریتمی باتکنیک جمعی می‌باشد، بهترین نتایج دقت در پیش‌بینی بقای بیماران مبتلا به همبودی را که معادل ۷۷/۸٪ است، به ارمغان آورده است.

34- Conformal Prediction

35- Zolbanin HM, Delen D, Zadeh AH

33- Alnemer LM, Rajab L, Aljarah I

پژوهشگران با توجه به استفاده از نرم‌افزار پر قدرت SAS^{۳۸} در تحلیل و پردازش داده‌ها و نتیجه‌گیری از آن‌ها با تکنیک‌های گوناگون، در این مقاله به شرح جزئیات مراحل کار به صورت مفصل نپرداخته‌اند و تنها با توجه به مقایسه نتایج به این نکته اشاره نموده‌اند که داشتن اطلاعات بیشتر در مورد بیماری‌های همزمان بیماران، می‌تواند قدرت پیش‌بینی بقای آن‌ها را به طرز چشم‌گیری بهبود ببخشد و به پزشکان در تشخیص و تصمیم‌گیری بهتر و حتی درمان کمک کند. بنابراین با استفاده از وضعیت همبودی بیماران می‌توان هزینه‌های درمان را به‌طور چشمگیری پایین آورد و مراقبت‌های بهداشتی مربوط به چالش‌های اقتصادی را کاهش داد. روش تحقیق این پژوهشگران در شکل (۱۲) نمایان شده است.

آلبرتسن و همکاران^{۳۷} [۱۴] در سال ۲۰۱۱، احتمال بقای بیماران دارای همبودی سرطان را بر اساس سن بیمار، مرحله تومور و درجه تومور، در مردانی با سن ۶۶ سال و بالاتر و در طی ۱۰ سال بر روی حدود ۲۰,۰۰۰ بیمار بررسی نموده‌اند. همه این مردان مبتلا به پروستات موضعی تشخیص داده شده و در طی ۱۸۰ روز از تشخیص، مورد جراحی یا پرتودرمانی قرار نگرفته‌اند. همچنین علت مرگ هر بیمار در پایگاه داده SEER مورد بررسی قرار گرفته است. اکثر بیماران (۷۲٪) در گروه مورد مطالعه دارای همبودی معنی داری نبودند، ۱۸٪ همبودی معنی داری داشتند و ۱۰٪ دارای دو یا چند همبودی همزمان بودند. حدود دو سوم بیماران به‌عنوان بیماری دارای اختلال متوسطی بودند و حدود یک سوم دارای بیماری ضعیف. سن متوسط در تشخیص ۷۷ سال بوده و پیگیری متوسط برای جامعه مورد مطالعه ۶ سال. در پایان دوره مطالعه، ۶۰٪، ۵۱٪ و ۴۱٪ از مردان با صفر، یک یا دو و چندین بیماری مشترک هنوز زنده بودند و تنها ۵٪ و ۱۱٪ از مردان مبتلا به بیماری متوسط یا ضعیف به علت سرطان پروستات فوت کردند. بدین ترتیب پژوهشگران

نشان دادند که در ۱۰ سال اول پس از تشخیص، مردان مبتلا به سرطان پروستات نسبتاً ضعیف و ضعیف، بیشتر به خاطر ابتلای همزمان به سایر بیماری‌ها جان خود را از دست می‌دهند و نه از بیماری سرطان به تنهایی. نتایج به دست آمده توسط این پژوهشگران می‌تواند کمک شایان توجهی در تخمین میزان نرخ بقای بیماران دارای همبودی سرطان پروستات موضعی نموده که خطر مرگ در مردان دارای بیماری‌های متوسط همزمان با سرطان پروستات بسیار بیشتر از مردان دارای بیماری‌های ضعیف همزمان با سرطان پروستات می‌باشد.

فیری و همکاران^{۳۸} [۱۹] در سال ۲۰۰۹، برای نمایش ارتباط بین همبودی و بقای بیماران دارای سرطان مثانه از پایگاه داده سیستم‌تومی رادیکال (حذف کامل مثانه، گره‌های لنفاوی مجاور، بخشی از مجرا و اندام‌های مجاور که ممکن است شامل سلول‌های سرطانی باشد). آلبرتا^{۳۹}، که پایگاه داده‌ای چندرسانه‌ای و مستمرا در حال به روز رسانی است، استفاده نموده‌اند. این پایگاه داده حاوی داده‌های مربوط به تمام بیماران بالغ مبتلا به تشخیص اولیه سرطان مثانه است که در ماه اکتبر ۱۹۹۴ در ادمنتون کانادا تحت درمان با رادیکال سیتکتومی قرار گرفته‌اند. این مقاله یک تحلیل از بیماران پایگاه داده مذکور می‌باشد که از آوریل ۱۹۹۴ تا سپتامبر ۲۰۰۷ تحت درمان بودند. تجزیه و تحلیل رگرسیون متناسب کاکس (Cox) جهت تعیین ارتباط بین همبودی و میزان بقای کلی بیماران و همچنین ارتباط بین میزان بقا در بیماران دارای سرطان مثانه و همبودی مورد استفاده قرار گرفت. از بیماران پایگاه داده مورد بررسی ۱۶۰، ۲۲۵ و ۸۳ بیمار، دارای همبودی خفیف، متوسط و شدید بودند. پس از بررسی بیماران مذکور در نهایت به نقش انکارناپذیر همبودی در سرطان مثانه که باید در آزمایش‌های بالینی مد نظر قرار گیرند دست یافتند، چرا که پس از تجزیه و تحلیل رگرسیون متناسب با چند متغیر کاکس شامل سن، میزان مصرف دارو پس

38- Fairey AS, Jacobsen N-EB, Chetner MP, Mador DR, Metcalfe JB, Moore RB

39- AUI : The Alberta Urology Institute

36- Statistical Analysis Software

37- Albertsen PC, Moore DF, Shih W, Lin Y, Li H, Lu-Yao GL

جدول ۱: مقایسه تحقیقات انجام شده قبلی

General Subject	Specific Subject	Paper	Dataset	Modeling Technique	Results
Cancer Prediction	Prostate	[7] 2017	MRI/MRSI data base of 34 patients	SVM Classification RF Classification	error rate 1% sensitivity 99.1% specificity 98.4%. error rate 18.2% sensitivity 72% specificity of 88%
	Breast	[6] 2016	16-year registry cohort database (1998-2014) on 1085 women who diagnosed with breast cancer in Isfahan Sayed-o-Shohada cancer research center	HPBCR classifiers SVM classifiers Decision tree classifiers MLP classifiers	Sensitivity 80% specificity 96.1% precision 89.2% accuracy 96.5% Sensitivity 73% specificity 85.3% precision 77.6% accuracy 82.1% Sensitivity 74% specificity 77% precision 77.1% accuracy 77.3% Sensitivity 67.5% specificity 84.3% precision 76% accuracy 84.2%
	Breast	[4] 2015	Southampton Breast Cancer Data System (SBCDS) for over 16,000 patients	Naive Bayes DT (J48) ANN SVM LR	N/A
	Breast	[3] 2014	Wisconsin Diagnostic Breast Cancer (WDBC)	KSVM (K-means + SVM) ACO-SVM (Prasad et al., 2010) GA-SVM (Prasad et al., 2010) PSO-SVM (Prasad et al., 2010)	accuracy 97.38% accuracy 95.96% accuracy 97.19% accuracy 97.19%

Cancer Survivability	Female Genital (ovary- cervix - uterus) Breast	[11] 2015	35-year period between 1979 and 2013 nationwide population register of the Hungarian Central Statistical Ofc	LR	N/A
	Breast	[10] 2014	University Medical Centre, Institute of Oncology, Ljubljana, Yugoslavia database	REP Tree ANN (RBF Network) LR REP Tree ANN (RBF Network) LR	No recurrence results Accuracy 71.32% Precision 73.7% Recall 92% Accuracy 73.77% Precision 77.4% Recall 88.6% Accuracy 74.47% Precision 76.2% Recall 92.5% recurrence results Accuracy 71.32% Precision 54.3% Recall 22.4% Accuracy 73.77% Precision 58.9% Recall 38.8% Accuracy 74.47% Precision 64.3% Recall 31.8%
	Breast	[12] 2009	SEER	C5 DT Bagging	Specificity 23.25 % Sensitivity 98.14 % Accuracy 88.05 % Specificity 78.59 % Sensitivity 74.96 % Accuracy 76.59 %
	Breast	[9] 2005	SEER	LR ANN(MLP) DT (C5)	Accuracy 89.20% Sensitivity 90.17% Specificity 87.86% Accuracy 91.21% Sensitivity 94.37% Specificity 87.48% Accuracy 93.62% Sensitivity 96.02% Specificity 90.66%

Cancer Co-morbidity	In General	[2] 2016	-	-	-
	Breast (Survivability)	[5] 2016	SEER	ANN DT KNN SVM	Accuracy 90.15% Sensitivity 66.86% Specificity 98.11% Precision 92.37% Accuracy 90.44% Sensitivity 71.51% Specificity 96.92% Precision 88.81% Accuracy 88.95% Sensitivity 65.99% Specificity 96.82% Precision 87.64% Accuracy 90.88% Sensitivity 71.51% Specificity 97.51% Precision 90.77%
	Breast + Female Genital Urinary + Male Genital	[8] 2015	SEER	RF(Breast-female genital) ANN(Breast-female genital) DT(Breast-female genital) LR(Breast-female genital) RF (Male genital-urinary) ANN (Male genital-urinary) DT (Male genital-urinary) LR (Male genital-urinary)	Accuracy 77.80% Sensitivity 64.43% Specificity 86.02% Accuracy 75.35% Sensitivity 67.54% Specificity 80.14% Accuracy 76.07% Sensitivity 66.10% Specificity 82.20% Accuracy 74.98% Sensitivity 67.30% Specificity 79.70% Accuracy 73.48% Sensitivity 53.48% Specificity 85.73% Accuracy 70.39% Sensitivity 56.43% Specificity 78.94% Accuracy 72.50% Sensitivity 54.34% Specificity 83.62% Accuracy 46.64% Sensitivity 63.09% Specificity 36.56%
	Prostate + other diseases	[20] 2011	SEER	Statistical Analysis	N/A
	Bladder + other diseases	[19] 2009	Alberta Urology Institute Radical Cystectomy	Cox Regression	N/A

از شیمی درمانی، حجم جراحی، وضعیت گره‌های لنفاوی پاتولوژیک، تعداد کل گره‌های لنفاوی حذف شده، وضعیت حاشیه جراحی و تهاجم لنفاوی، اثبات نمودند که افزایش همبودی به‌طور مستقل با افزایش خطر مرگ و میر کلی و مرگ و میر خاص سرطان مثانه پس از سیستکتومی رادیکال، همراه بوده است.

جدول (۱)، تحقیقات انجام شده در زمینه پیش‌بینی سرطان و به ویژه بررسی نرخ بقا بیماران مبتلا به همبودی سرطان را به‌طور خلاصه نشان می‌دهد.

بحث و نتیجه‌گیری

در میان الگوریتم‌های به کار رفته در مقالات بررسی شده، به ترتیب الگوریتم‌های ANN، DT و SVM از بیشترین تعداد کاربرد و SVM از بهترین نتایج برخوردار است. نتایج دقت SVM از ۸۵/۳ درصد در تشخیص سرطان سینه بر روی داده‌های سرطان بیمارستان سیدالشهدا اصفهان تا ۹۷/۳۸ درصد که روشی ترکیبی با K-mean بوده و در تشخیص سرطان سینه روی داده‌های WDBC استفاده شده، متغیر بوده است.

با توجه به مرور ادبیات تحقیق انجام شده به نظر می‌رسد می‌توان نتایج بهتری بر روی موضوع بررسی نرخ بقای بیماران مبتلا به همبودی سرطان به دست آورد. آقای ذوالبنین و همکارانش [۸] جهت تشخیص نرخ بقای سرطان‌های سینه، دستگاه تناسلی و سیستم ادراری در خانم‌ها و آقایان از چهار تکنیک داده‌کاوی ANN، RF، DT و LR استفاده نموده و بهترین نتایج را از الگوریتم جنگل تصادفی که نوعی الگوریتم ترکیبی است، به دست آورده‌اند. با توجه به این‌که در روند طی شده در این منبع از الگوریتم‌های انتخاب ویژگی که می‌توان با استفاده از آن‌ها یا ترکیبی از آن‌ها به نتایج بهتری دست یافت استفاده نشده لذا می‌توان با به کارگیری یک روش انتخاب ویژگی، ویژگی‌های بهینه‌ای از داده‌های مربوط به سرطان سینه، دستگاه تناسلی و سیستم ادراری را

گزینش نموده و سپس با به کارگیری روش‌های یادگیری ترکیبی، تشخیص نرخ بقای بیماران مبتلا به آن‌ها را که دارای همبودی نیز هستند، نسبت به روش مذکور، بهبود داد.

مراجع

- [1] D. Sarfati and J. Gurney, "What Is Comorbidity?," in: Koczwara B. (eds) Cancer and Chronic Conditions, ed: Springer, 03 September 2016, pp. 1-33.
- [2] Sarfati, Diana, Bogda Koczwara, and Christopher Jackson. "The impact of comorbidity on cancer and its treatment." CA: a cancer journal for clinicians, Volume 66, Issue 4 July/August 2016, pp. 337-350
- [3] B. Zheng, S. W. Yoon, and S. S. Lam, "Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms," Expert Systems with Applications, vol. 41, 2014, pp. 1476-1482.
- [4] J. Lu, A. Hales, D. Rew, M. Keech, C. Fröhlingsdorf, A. Mills-Mullett, et al., "Data mining techniques in health informatics: a case study from breast cancer research," in In: Renda M., Bursa M., Holzinger A., Khuri S. (eds) International Conference on Information Technology in Bio-and Medical Informatics, ITBAM 2015. Lecture Notes in Computer Science, vol 9267. Springer, Cham
- [5] L. M. Alnemer, L. Rajab, and I. Aljarah, "Conformal Prediction Technique to Predict Breast Cancer Survivability," International Journal of Advanced Science and Technology 96, 2016, pp 1-10.
- [6] M. R. Mohebian, H. R. Marateb, M. Mansourian, M. A. Mañanas, and F. Mokarian, "A Hybrid Computer-aided-diagnosis System for Prediction of Breast Cancer Recurrence (HPBCR) Using Optimized Ensemble Learning," Computational and Structural Biotechnology Journal, 15, 2017 Dec 31, pp.75-85.
- [7] R. Trigui, J. Mitéran, P. Walker, L. Sellami, and A. B. Hamida, "Automatic classification and localization of prostate cancer using multi-parametric MRI/MRS," Biomedical Signal Processing and Control, vol. 31, pp. 189-198, 2017.
- [8] H. M. Zolbanin, D. Delen, and A. H. Zadeh, "Predicting overall survivability in comorbidity of cancers: A data mining approach," Decision Support Systems, vol. 74, pp. 150-161, 2015.
- [9] D. Delen, G. Walker, and A. Kadam, "Predicting breast cancer survivability: a comparison of three data mining methods," Artificial intelligence in medicine, vol. 34, pp. 113-127, 2005.
- [10] V. Chaurasia and S. Pal, "Data mining techniques: To predict and resolve breast cancer survivability," International Journal of Computer Science and Mobile Computing, vol. 3, pp. 10-22, 2014.
- [11] M. Hartai, T. A. Nyári, and R. J. McNally, "Trends in mortality rates in female genital organs and breast cancers in Hungary between 1979 and 2013," European Journal of Obstetrics

2016. Lecture Notes in Computer Science, vol 10040. Springer, Cham 2016 Jul 29, pp. 578-588.

[17]S. Gu, R. Cheng, and Y. Jin, "Feature selection for high-dimensional classification using a competitive swarm optimizer," Soft Computing, pp. 1-12, 07 October 2016.

[18] M. Akay, Pattern Recognition in: Wiley encyclopedia of biomedical engineering: Wiley-Interscience, Lyapunov exponents 2006. New York: Wiley.

[19]A. S. Fairey, N.-E. B. Jacobsen, M. P. Chetner, D. R. Mador, J. B. Metcalfe, R. B. Moore, et al., "Associations between comorbidity, and overall survival and bladder cancer specific survival after radical cystectomy: results from the Alberta Urology Institute Radical Cystectomy database," The Journal of urology, vol. 182, pp. 85-93, 2009.

[20] P. C. Albertsen, D. F. Moore, W. Shih, Y. Lin, H. Li, and G. L. Lu-Yao, "Impact of comorbidity on survival among men with localized prostate cancer," Journal of Clinical Oncology, vol. 29, pp. 1335-1341, 2011.

& Gynecology and Reproductive Biology, vol. 194, pp. 168-172, 2015.

[12]L. Ya-Qin, W. Cheng, and Z. Lu, "Decision tree based predictive models for breast cancer survivability on imbalanced data," in ICBBE 2009 3rd International Conference on Bioinformatics and Biomedical Engineering, 2009, pp. 1-4, 2009.

[13]D. Oliva and E. Cuevas, "Advances and Applications of Optimised Algorithms in Image Processing," ed: Springer, 2017 Jan 2, Vol. 117,

[14]M. J. Zaki and W. Meira Jr, Data mining and analysis: fundamental concepts and algorithms: Cambridge University Press, 2014 May 12.

[15]D. T. Larose, Discovering knowledge in data: an introduction to data mining: John Wiley & Sons, 2014 Jun 2.

[16]L. Liu, B. Wang, B. Yu, and Q. Zhong, "A Novel Selective Ensemble Learning Based on K-means and Negative Correlation," in Sun X., Liu A., Chao HC., Bertino E. (eds) International Conference on Cloud Computing and Security, ICCCS

جدیدترین کتاب از انتشارات انجمن انفورماتیک ایران منتشر شد!

کار عمیق

برای تهیه کتاب با دفتر انجمن انفورماتیک ایران

تماس بگیرید ۶۶۴۱۲۸۶۱

