

خلاصه پروپوزال طرح تحقیقاتی

عنوان طرح: پیش بینی احتمال بقا در بیماران مبتلا به سرطان سینه با استفاده از تکنیک های داده کاوی

کلید واژه ها: سرطان سینه، احتمال بقا، پیش بینی، داده کاوی

Keywords: Breast Cancer, Survival, Prediction, Datamining

مجری اصلی:

محمد ستاری

لاین تحقیقاتی مجری اصلی:

داده کاوی در حوزه سلامت

ضرورت اجرای طرح به صورت خلاصه

سرطان سینه یکی از شایعترین انواع سرطانها در میان جوامع انسانی، به ویژه بانوان می باشد. این بیماری سالانه باعث ابتلا و مرگ تعداد زیادی از بیماران می شود. پیشگیری از ابتلای افراد به این بیماری و به حداقل رساندن عوارض ناشی از آن بسیار ضروری به نظر می رسد. شناخت الگو و روند ابتلای افراد به این بیماری و برآورد میزان بقای آنها از شاخصهای تاثیر گذار در فرآیند تصمیم گیری و برنامه ریزی اقدامات لازم برای کنترل و پیشگیری از این بیماری در سطوح ملی، استانی و محلی می باشد. برآورد احتمال بقا، منافع و تسهیلات مهمی در اختیار ارائه دهندگان مراقبت به این بیماران و تصمیم گیرندگان قرار می دهد، از جمله شناسایی روشهای مناسب غربالگری و پیشگیری از وقوع بیماری، بهبود مراقبت از عوارض دراز مدت درمان، حفظ کیفیت زندگی آنان و تصمیم گیریهای پیچیده تر در درمان بیماران. با توجه به اهمیت پیش بینی احتمال بقا در مدیریت بیماری سرطان سینه، محققین قصد دارند در این پژوهش احتمال بقای بیماران مبتلا به سرطان سینه را با استفاده از تکنیک های داده کاوی مورد پیش بینی قرار دهند.

روش شناسی پژوهش (به صورت خلاصه)

این مطالعه از نظر هدف کاربردی - توسعه ای می باشد که به شیوه ی مقطعی انجام خواهد شد. جامعه آماری شامل تمامی بیماران مبتلا به سرطان سینه در سطح شهر اصفهان که اطلاعات مربوط به بیماری آنها در بیمارستان سید الشهدای اصفهان موجود می باشد، است. زنانی که در فاصله ی بین سال های ۱۳۹۱ تا ۱۳۹۶ به سرطان سینه مبتلا شده اند و اطلاعات آنها در این بیمارستان موجود است در مطالعه وارد شدند.

جمع آوری داده ها در دو مرحله ی انتخاب زیر مجموعه ای از صفات و پاک سازی داده انجام خواهد شد. پس از جمع آوری اطلاعات مورد نیاز، با استفاده از نرم افزار Rapidminer داده ها دسته بندی خواهند شد. تکنیک های مورد استفاده در این مطالعه شامل درخت تصمیم، ماشین بردار پشتیبان و شبکه های عصبی خواهد بود. هر یک از تکنیک ها بر روی مجموعه ی داده ای آموزشی پیاده سازی می شوند. سپس الگوی میزان بقا استخراج شده و بر روی مجموعه ی داده ای تست مورد آزمایش و ارزیابی قرار می گیرد.

نام و نام خانوادگی مجری اصلی:

محمد ستاری

۱- نشانی (مجری):

محل کار:

اصفهان - خیابان هزار جریب - دانشگاه علوم پزشکی اصفهان - دانشکده مدیریت و اطلاع رسانی پزشکی - طبقه ی دوم - اطاق شماره ی ۲۰۶

تلفن: ۰۳۱۳۷۹۲۵۱۵۴

فاکس: -

پست الکترونیک:

msattarimng.mui@gmail.com

نام مؤسسه بانکی و شماره حساب مجری (مجری اصلی طرح):

شغل و سمت فعلی مجری اصلی طرح : استادیار گروه فناوری اطلاعات سلامت - دانشکده مدیریت و اطلاع رسانی پزشکی - دانشگاه علوم پزشکی اصفهان

سازمان متبوع :

محل اجرای طرح :

درجات علمی و سوابق تحصیلی مجری اصلی طرح:

کشور	دانشگاه	رشته تحصیلی و تخصصی	درجه تحصیلی	سال دریافت
ایران	اصفهان	مهندسی کامپیوتر - نرم افزار	دکتری	۱۳۹۷

مشخصات سایر مجریان و یا همکاران اصلی:

نام و نام خانوادگی	شغل	لاین تحقیقاتی	درجه علمی	نوع همکاری	امضای همکار
حسین باقریان	هیات علمی	مدیریت اطلاعات در نظام سلامت	استادیار	مجری دوم	
مهران شریفی اصفهانی	هیات علمی	سرطان سینه و لوسمی	استادیار	همکار	

هدف کلی

پیش بینی احتمال بقا در بیماران مبتلا به سرطان سینه با استفاده از تکنیک های داده کاوی

اهداف اختصاصی

- انتخاب عوامل اصلی تاثیر گذار بر بقای بیماران مبتلا به سرطان سینه
- پاکسازی داده ی بیماران مبتلا به سرطان سینه
- پیش بینی احتمال بقا بیماران مبتلا به سرطان سینه با استفاده از تکنیک درخت تصمیم
- پیش بینی احتمال بقا بیماران مبتلا به سرطان سینه با استفاده از تکنیک شبکه ی عصبی عمیق
- پیش بینی احتمال بقا بیماران مبتلا به سرطان سینه با استفاده از تکنیک ماشین بردار پشتیبان
- تعیین ترکیب بهینه تکنیک های بکار رفته در مراحل مختلف داده کاوی جهت پیش بینی احتمال بقای بیماران مبتلا به سرطان سینه

سوالات پژوهش:

- عوامل اصلی تاثیر گذار بر بقای بیماران مبتلا به سرطان سینه شامل چه مواردی هستند؟
- پاکسازی داده های بیماران مبتلا به سرطان سینه چگونه انجام می شود؟
- احتمال بقا بیماران مبتلا به سرطان سینه با استفاده از تکنیک درخت تصمیم چه میزان می باشد؟
- احتمال بقا بیماران مبتلا به سرطان سینه با استفاده از تکنیک شبکه ی عصبی چه میزان می باشد؟
- احتمال بقا بیماران مبتلا به سرطان سینه با استفاده از تکنیک ماشین بردار پشتیبان چه میزان می باشد؟
- ترکیب بهینه تکنیک های بکار رفته در مراحل مختلف داده کاوی جهت پیش بینی احتمال بقای بیماران مبتلا به سرطان سینه چیست؟

اطلاعات مربوط به متن کامل طرح پژوهشی

عنوان طرح پژوهشی به فارسی:

پیش بینی احتمال بقا در بیماران مبتلا به سرطان سینه با استفاده از تکنیک های داده کاوی

عنوان طرح پژوهشی به انگلیسی:

Survival prediction for patients with breast cancer by datamining techniques

نوع مطالعه:

بنیادی □ کاربردی ■ بنیادی - کاربردی □

مقدمه (بیان مساله - ضرورت انجام پژوهش)

سرطان سینه نوعی تومور بدخیم است که طی آن سلولهای بافت سینه به علت اختلالات ژنتیکی مانند جهش، افزایش کروموزومی، حذف، بازسازی، جابه جاشدگی و تکرارشدگی کروموزومی بدون هیچ کنترلی شروع به تکثیر شدن می کنند(۱) در حال حاضر سرطان سینه، شایعترین سرطان در میان زنان در سراسر جهان می باشد (۷-۲). به گونه ای که ۲۴/۲ درصد سرطانهایی که در زنان در سال ۲۰۱۸ شناسایی شده، سرطان سینه می باشد(۹-۸). در ایران این میزان ۲۷ درصد می باشد. میزان بروز سرطان سینه در زنان ایرانی ۳۱ در ۱۰۰ هزار نفر می باشد (۱۰). همچنین این سرطان در میان زنان در سطح جهانی، رتبه نخست مرگ و میر ناشی از سرطان را به خود اختصاص داده است.

طبق گزارش سازمان جهانی بهداشت در سال ۲۰۱۸ در حدود ۱۵ درصد مرگهای ناشی از سرطان در زنان مربوط به سرطان سینه بوده است(۹). سن مرگ و میر ناشی از سرطان پستان در ایران بین ۴۰ تا ۴۹ سالگی است، درحالی که این سن در کشورهای پیشرفته بین ۵۵ تا ۶۰ است؛ بنابراین در ایران حداقل ده سال پایین تر از کشورهای پیشرفته است و با توجه به نقش محوری زنان در این سن در خانواده و جامعه، مرگ و میر و ناتوانی ناشی از بیماری صدمات جبران ناپذیری به جامعه و خانواده وارد می کند(۱۱-۱۰).

تحلیل بقا رویکردی مهم در تعیین عوامل مرتبط با بقای بیماری است و روشهای تحلیل بقا با نوعی از داده ها سر و کار دارند که در آنها زمان انتظار تا رخداد یک حادثه را اندازه میگیرند(۱۳). بررسی میزان بقا در بیماران مبتلا به سرطان سینه، به عنوان شاخصی جهت ارزیابی و ارتقا روشهای مختلف درمانی محسوب می شود(۱۴). احتمال بقا بیماران مبتلا به سرطان سینه با توجه به ویژگیهای بالینی افراد متفاوت می باشد. در مقایسه با سایر انواع سرطانها احتمال بقا بیماران مبتلا به سرطان ریه بالاتر بوده به ویژه اگر این بیماری زود تشخیص داده شود(۴). بهبود میزان بقا این بیماری مسائل جدیدی از جمله ابداع روشهای مناسبتر جهت غربالگری و پیشگیری از وقوع بیماری و تشخیص به موقع عود و متاستاز، بهبود مراقبت از عوارض دراز مدت درمان و حفظ کیفیت زندگی آنان و تصمیم گیریهای پیچیده تر در درمان بیماران را در پیش روی پزشکان و محققین قرار میدهد (۱۵).

در سالهای اخیر رویکردهای داده کاوی مانند کشف دانش جدید از پایگاه های داده به عنوان ابزار تحقیقاتی مناسب برای پژوهشگران حوزه پزشکی تبدیل شده است و محققین از تکنیک های مختلف برای پیش بینی بقا یا عود مجدد سرطان ها استفاده می کنند(۱۶). داده کاوی از روش هایی است که به منظور یافتن الگو از درون پایگاه های بزرگ داده و نیز پیش بینی نتایج سرطان ها با استفاده از الگوریتم های پیش بینی کننده بکار می رود. با کمک داده کاوی می توان الگوهایی که به سختی قابل تشخیص هستند را از پایگاه های داده

استخراج نمود. مطالعات مختلفی از تکنیک های داده کاوی برای پیش بینی احتمال بقا در بیماران مبتلا به سرطان استفاده کرده اند (۲۰-۱۷).

گلی و همکاران در مطالعه ای با بهره گیری از مدل رگرسیون بردار پشتیبان^۱ به پیش بینی احتمال بقای بیماران مبتلا به سرطان سینه در شهر همدان پرداختند (۱۷). مومنیان و همکاران در مطالعه خود از تکنیک درخت تصمیم و تحلیل رگرسیون لجستیک برای پیش بینی احتمال بقا در بیماران مبتلا به سرطان سینه در تهران استفاده کردند (۴). دِلن^۲ و همکاران در مطالعه خود از دو تکنیک درخت تصمیم و شبکه هوش مصنوعی به همراه تکنیک آماری رگرسیون لجستیک برای پیش بینی احتمال بقای بیماران مبتلا به سرطان سینه استفاده کردند (۲۱). لیو^۳ در مطالعه خود بر پایه داده های نامتعادل از درخت تصمیم و الگوریتم بگینگ برای پیش بینی احتمال بقا در مبتلایان به سرطان سینه استفاده کردند (۲۲). با توجه به مطالبی که در خصوص اهمیت پیش بینی بقا ارائه شد و با در نظر گرفتن این نکات که بسیاری از مطالعات برای پیش بینی احتمال بقای بیماران مبتلا به سرطان سینه از یک یا نهایتاً دو تکنیک داده کاوی بصورت مستقل یا در مقایسه با یکدیگر استفاده کرده اند و با ذکر این نکته که تا کنون مطالعه ای در این خصوص بروی بیماران مبتلا در سطح شهر اصفهان انجام نشده است، پژوهشگران قصد دارند در این مطالعه از سه تکنیک داده کاوی برای پیش بینی احتمال بقای بیماران مبتلا به سرطان سینه استفاده کنند.

ضرورت انجام پژوهش:

بر اساس آمارهای سازمان جهانی بهداشت، از هر ۸ تا ۱۰ زن، یک نفر به سرطان پستان مبتلا می گردد (۲۳). این بیماری هر ساله باعث مرگ و میر فراوانی در بین زنان و مردان می شود. با وجود پیشرفت های بسیاری که در زمینه تشخیص زودهنگام و درمان مناسب این بیماری صورت گرفته است، هنوز هم سرطان سینه سردهسته علل مرگ به علت سرطان در بین زنان می باشد (۲۴). میزان بقا پس از تشخیص، درمان و بررسی عوامل خطر ساز در بیماران سرطانی یکی از شاخصهای مهم در کنترل بیماری و ارزیابی روشهای درمانی مورد استفاده است (۲۵).

پیش بینی احتمال بقای بیماران پس از ابتلا به سرطان سینه زمینه مساعدی برای تصمیم گیری در خصوص برنامه ریزی و سرمایه گذاری بروی روش های درمانی و جراحی و صرف هزینه برای درمان بیماران مبتلا را فراهم می آورد. مدیران بخش سلامت با در نظر گرفتن احتمال بقای بیماران مبتلا با سهولت بیشتری می توانند نسبت به اولویت بندی برنامه های کنترل و پیشگیری از این بیماری بیندیشند. از سوی دیگر تقویت عواملی که منجر به افزایش احتمال بقای بیماران مبتلا به سرطان سینه می شوند در صورتی ایجاد می شود که تحلیل درستی از میزان بقای بیماران و عوامل تاثیر گذار بروی آن وجود داشته باشد. با در نظر گرفتن اهمیت پیش بینی احتمال بقای بیماران مبتلا به سرطان سینه و با توجه به اهمیت شناسایی عوامل تاثیر گذار بر احتمال بقای این بیماران، پژوهشگران قصد دارند در این مطالعه احتمال بقای بیماران مبتلا به سرطان ریه با استفاده از تکنیک های مختلف داده کاوی را مورد بررسی قرار دهند.

پیشینه های پژوهش

مروری بر مطالعات انجام شده در داخل ایران

1 - Support Vector Regression (SVR) Model

2 - Delen

3 - Liu

صالحی و همکاران (۱۳۹۲) در مطالعه ای با عنوان " مقایسه شبکه ی عصبی و تکنیک رگرسیون کاکس برای پیش بینی میزان بقای در بیماران مبتلا به سرطان سینه " از تکنیک داده کاوی شبکه ی عصبی و تکنیک آماری رگرسیون کاکس برای پیش بینی میزان بقا در بیماران مبتلا به سرطان سینه استفاده کردند. آنها یک مدل پیش بینی براساس دو تکنیک داده کاوی ارائه دادند. مجموعه ی داده ای مورد استفاده در پژوهش براساس داده های جمع آوری شده از بیمارستان فیاض بخش تهران بین سال های ۲۰۰۵ تا ۲۰۱۲ بود. این مجموعه شامل ۳۴۴ بیمار بود. نتایج نشان داد که بین دو مدل شبکه ی عصبی و رگرسیون، شبکه ی عصبی عملکرد بهتری داشته است (۲۶).

افشار و همکاران (۱۳۹۴) در مطالعه ای با عنوان " پیش بینی میزان بقای سرطان سینه با استفاده از اکتشاف دانش در پایگاه داده ها " از تکنیک های بیزین، ماشین بردار پشتیبان، تشخیص تعاملی خودکار مجذور کای برای پیش بینی میزان بقا در بیماران مبتلا به سرطان سینه استفاده کردند. این تکنیک ها بر روی داده هایی که در بازه ی زمانی ۱۹۹۹ تا ۲۰۰۴ از پایگاه داده ی جهانی برنامه ی نظارت، اپیدمیولوژی و نتایج نهایی در کشور آمریکا جمع آوری شده بود، پیاده سازی شدند. داده های جمع آوری شده شامل ۲۲۷۶۳ رکورد و ۶۷ متغیر بود. نتایج نشان دهنده ی دقت بیشتر روش ماشین بردار پشتیبان نسبت به دیگر تکنیک ها بود (۲۷).

خلخالی و همکاران (۱۳۹۵) در مطالعه ای با عنوان " بکار بردن تکنیک های داده کاوی برای اکتشاف الگوی پنهان درباره میزان بقای بیماران مبتلا به سرطان سینه در یک مطالعه ی کوهورت ایرانیان " از تکنیک های الگوریتم CART از تکنیک های مبتنی بر درخت تصمیم برای پیش بینی میزان بقا در بیماران مبتلا به سرطان سینه استفاده کردند. مجموعه ی داده ای مورد استفاده در پژوهش براساس داده های کوهورت جمع آوری شده از مرکز تحقیقاتی و رفتاری امید که از ۲۰۰۷ تا ۲۰۱۰ را در بر می گیرد، بود. این مجموعه شامل ۵۶۹ بیمار بوده و از این مجموعه، ۹ فاکتور ریسک انتخاب شد. نتایج نشان دهنده ی دقت بالای روش پیشنهادی ایشان بود (۲۸).

منتظری و همکاران (۱۳۹۵) در مطالعه ای با عنوان " مدل های یادگیری ماشین در پیش بینی میزان بقای بیماران مبتلا به سرطان سینه " از مدل های بیزین ساده، جنگل تصادفی درختان، ادراک چند سطحی، ادابوست، ماشین بردار پشتیبان و ۱- نزدیکترین همسایه برای پیش بینی میزان بقا در مبتلایان به سرطان سینه استفاده کردند. نتایج نشان می دهد که جنگل تصادفی درختان عملکرد بهتری نسبت به بقیه ی تکنیک ها داشته است. مجموعه ی داده ای مورد استفاده شامل ۸۷۶ زن و ۲۴ مرد و ۸ صفت است. (۲۹).

تاپاک و همکاران (۱۳۹۷) در مطالعه ای با عنوان " پیش بینی میزان بقای و متاستاز در بیماران مبتلا به سرطان با استفاده از دسته بندی های یادگیری ماشین " از تکنیک های شبکه بیزین، درخت تصمیم و رگرسیون لجستیک برای پیش بینی میزان بقا ۵ ساله در مبتلایان به سرطان سینه استفاده کردند. همچنین در این مدل از ترکیب تئوری مورچگان و انتخاب صفت مبتنی بر همبستگی جهت انتخاب صفات کلیدی استفاده شده است. علاوه بر این تکنیک های بیش نمونه برداری نوسانگر ذخیره سازی دیجیتال^۱ و بیش نمونه سازی ترکیبی اقلیت حاشیه ای^۲ استفاده شده است. این تکنیک ها بر روی داده هایی که در بازه ی زمانی ۲۰۰۴ تا ۲۰۰۷ از پایگاه داده ی جهانی برنامه ی نظارت، اپیدمیولوژی و نتایج نهایی در کشور آمریکا جمع آوری شده بود، پیاده سازی شد این مجموعه شامل ۲۷۰۹۸۹ رکورد و ۱۵۱ متغیر انتخاب شد. نتایج نشان دهنده ی عملکرد بهتر ترکیب تئوری مورچگان و انتخاب صفت مبتنی بر همبستگی و درخت تصمیم C4.5 به همراه تکنیک بیش نمونه برداری نوسانگر ذخیره سازی دیجیتال عملکرد بهتری خواهند داشت (۳۰).

میری رستمی و احمد زاده (۱۳۹۷) در مطالعه ای با عنوان " استخراج متغیرهای پیش بینی برای ایجاد مدل میزان بقای سرطان سینه با در نظر گرفتن مشکل عدم توازن در کلاس " از تکنیک های بیزین، ماشین بردار پشتیبان، جنگل تصادفی، رگرسیون لجستیک برای پیش بینی

¹ DSO (Digital Storage Oscillator)

² Borderline-SMOTE

میزان بقا ۵ ساله در مبتلایان به سرطان سینه در حالت عدم توازن در داده استفاده کردند. همچنین این پژوهش به دنبال تعیین ارتباط بین متغیر های پیش بینی و میزان بقا بود. مجموعه ی داده ای مورد استفاده در پژوهش براساس داده های کوهورت جمع آوری شده در دانشگاه شهید بهشتی در سال ۲۰۱۴ که از ۱۹۹۸ تا ۲۰۱۳ را در بر می گیرد، بود. این مجموعه شامل ۵۵۰ بیمار بوده و از این مجموعه، ۹ فاکتور ریسک انتخاب شد. نتایج نشان دهنده ی عملکرد بهتر تکنیک ماشین بردار پشتیبان نسبت به دیگر تکنیک ها بود (۳۱).

مروری بر مطالعات انجام شده در خارج از ایران

دلن و همکاران (۲۰۰۵) در مطالعه ای با عنوان "پیش بینی میزان بقای سرطان سینه با مقایسه سه تکنیک داده کاوی" از سه تکنیک درخت تصمیم، شبکه های عصبی، رگرسیون لجستیک برای پیش بینی میزان بقا در بیماران مبتلا به سرطان سینه استفاده کردند. این تکنیک ها بر روی داده هایی که در بازه ی زمانی ۱۹۷۳ تا ۲۰۰۰ از پایگاه داده ی جهانی برنامه ی نظارت، اپیدمیولوژی و نتایج نهایی ۱ در کشور آمریکا جمع آوری شده بود، پیاده سازی شدند. داده های جمع آوری شده شامل ۴۳۳.۲۷۲ رکورد و ۷۲ متغیر است. نتایج نشان دهنده ی عملکرد بهتر تکنیک درخت تصمیم نسبت به بقیه ی تکنیک های موجود بود (۲۱).

بالاچیا^۲ و گوون (۲۰۰۶) در مطالعه ای با عنوان "پیش بینی میزان بقای سرطان سینه با استفاده از تکنیک های داده کاوی" از سه تکنیک بیزین ساده، شبکه ی عصبی و C4.5 برای پیش بینی میزان بقا در بیماران مبتلا به سرطان سینه استفاده کردند. این تکنیک ها بر روی داده هایی که در بازه ی زمانی ۱۹۷۳ تا ۲۰۰۰ از پایگاه داده ی جهانی برنامه ی نظارت، اپیدمیولوژی و نتایج نهایی در کشور آمریکا جمع آوری شده بود، پیاده سازی شد داده های جمع آوری شده شامل ۴۳۳.۲۷۲ رکورد و ۷۲ متغیر است. نتایج نشان دهنده ی عملکرد بهتر C4.5 که بر مبنای درخت تصمیم است، نسبت به بقیه ی تکنیک ها است (۳۲).

اندو^۳ و همکاران (۲۰۰۸) در مطالعه ای با عنوان "مقایسه هفت تکنیک داده کاوی برای پیش بینی میزان بقای سرطان سینه" از تکنیک های مبتنی بر بیزین، درخت تصمیم، شبکه های عصبی و رگرسیون لجستیک برای پیش بینی میزان بقا در بیماران مبتلا به سرطان سینه استفاده کردند. این تکنیک ها بر روی داده هایی که در بازه ی زمانی ۱۹۹۲ تا ۱۹۹۷ از پایگاه داده ی جهانی برنامه ی نظارت، اپیدمیولوژی و نتایج نهایی در کشور آمریکا جمع آوری شده بود، پیاده سازی شدند. داده های جمع آوری شده شامل ۳۷.۲۵۶ رکورد و ۶۷ متغیر است. نتایج نشان دهنده ی دقت بیشتر روش رگرسیون لجستیک نسبت به دیگر تکنیک ها است (۳۳).

خان^۴ و همکاران (۲۰۰۸) در مطالعه ای با عنوان "درخت تصمیم وزن دار فازی برای پیش بینی میزان بقای سرطان سینه" از تکنیک های درخت تصمیم وزن دار فازی و درخت تصمیم برای پیش بینی میزان بقا در بیماران مبتلا به سرطان سینه استفاده کردند. این تکنیک ها بر روی داده هایی که در بازه ی زمانی ۱۹۷۳ تا ۲۰۰۳ از پایگاه داده ی جهانی برنامه ی نظارت، اپیدمیولوژی و نتایج نهایی در کشور آمریکا جمع آوری شده بود، پیاده سازی شدند. داده های جمع آوری شده شامل ۵۰۵۳۶۷ و ۸۶ متغیر است. نتایج نشان می دهد که رویکرد درخت تصمیم وزن دار فازی عملکرد بهتری نسبت به درخت تصمیم دارد (۳۴).

لیو و همکاران (۲۰۰۹) در مطالعه ای با عنوان "مدل پیش بینی براساس درخت تصمیم برای پیش بینی میزان بقای سرطان سینه در حالت عدم توازن داده" از تکنیک درخت تصمیم برای پیش بینی میزان بقا در بیماران مبتلا به سرطان سینه در حالت عدم توازن داده استفاده کردند. این تکنیک بر روی داده هایی که در بازه ی زمانی ۱۹۷۳ تا ۲۰۰۴ از سه مجموعه ی داده ای موجود در پایگاه داده ی جهانی برنامه

1- SEER (Surveillance, Epidemiology, and End Results Program)

2 - Bellaachia

3 - Endo

4 - Khan

ی نظارت، اپیدمیولوژی و نتایج نهایی در کشور آمریکا جمع آوری شده بود پیاده سازی شد. سه مجموعه ی داده ای در هم ادغام شده و در نهایت یک مجموعه ی داده ای با ۷۷۹۹۹۹ رکورد و ۱۱۵ متغیر ایجاد شد. بررسی های پژوهشگر نشان داده وقتی داده ها عدم توازن دارند، تکنیک های داده کاوی دسته بندی به خوبی عمل نمی کنند. لذا از روش بگینگ برای بهبود نتایج استفاده کرد و موفق شد کارایی مدل پیشنهادی را افزایش دهد (۲۲).

ثونکام^۱ و همکاران (۲۰۰۹) در مطالعه ای با عنوان "به سمت پیش بینی میزان بقای سرطان سینه با بهبود فضای داده ی آموزشی" اثر عمل پیش پردازشی بر روی تکنیک های داده کاوی جهت پیش بینی میزان بقا در بیماران مبتلا به سرطان سینه بررسی کرد. پژوهشگر سعی در بهبود داده های قبل از پردازش توسط تکنیک های داده کاوی داشتند. در واقع با انجام یک عمل پیش پردازشی به دنبال مقابله با عدم توازن در داده ها بودند. این عمل پیش پردازشی شامل تشخیص داده های پرت و تکنیک های بیش نمونه برداری بود و سپس چهار تکنیک ادابوست، بگینگ، درخت تصمیم C4.5 و ماشین بردار پشتیبان برای تعیین میزان بقا در بیماران مبتلا به سرطان سینه استفاده شد. داده ها از یک بیمارستان در تایلند از سال ۱۹۹۰ تا ۲۰۰۱ جمع آوری شده و شامل ۳۳۷۸ بیمار است نتایج نشان می دهد که پیش پردازش اعمال شده باعث بهبود عملکرد تکنیک های داده کاوی شده است (۳۵).

کیم^۲ و همکاران (۲۰۱۳) در مطالعه ای با عنوان "پیش بینی میزان بقای سرطان سینه با استفاده از داده های بیمار برجسب دار، بدون برجسب و شبه برجسب" از تکنیک های نیمه اتوماتیک، شبکه ی عصبی و ماشین بردار پشتیبان برای پیش بینی میزان بقا در بیماران مبتلا به سرطان سینه استفاده کرد. با توجه به اینکه جمع آوری داده های بدون برجسب راحت تر است، پیشنهاد استفاده از تکنیک های نیمه اتوماتیک را برای تعیین میزان بقا در بیماران مبتلا به سرطان سینه ارائه شد. این تکنیک عل بر روی داده هایی که در بازه ی زمانی ۱۹۷۳ تا ۲۰۰۳ از پایگاه داده ی جهانی برنامه ی نظارت، اپیدمیولوژی و نتایج نهایی در کشور آمریکا جمع آوری شده بود، پیاده سازی شدند. داده های جمع آوری شده شامل ۵۰۵۳۶۷ و ۸۶ متغیر است. نتایج نشان دهنده ی عملکرد مناسب این تکنیک نسبت به شبکه ی عصبی و ماشین بردار پشتیبان بود (۳۶).

پارک^۳ و همکاران (۲۰۱۳) در مطالعه ای با عنوان "یک مدل پیش بینی قوی برای ارزیابی میزان بقای سرطان سینه" از تکنیک های شبکه ی عصبی، ماشین بردار پشتیبان و دسته بندی نیمه اتوماتیک برای پیش بینی میزان بقا در بیماران مبتلا به سرطان سینه استفاده کردند. این تکنیک ها بر روی داده هایی که در بازه ی زمانی ۱۹۷۳ تا ۲۰۰۳ از پایگاه داده ی جهانی برنامه ی نظارت، اپیدمیولوژی و نتایج نهایی در کشور آمریکا جمع آوری شده بود، پیاده سازی شدند. داده های جمع آوری شده شامل ۵۰۵۳۶۷ و ۸۶ متغیر است. بررسی ها نشان داد که دو تکنیک شبکه ی عصبی و ماشین بردار پشتیبان وابسته به مقادیر مختلف پارامتر ها هستند، این در حالی است که تکنیک دسته بندی نیمه اتوماتیک وابستگی به مقدار پارامتر ها نداشته و عملکرد نسبتاً قابل قبولی داشته است (۳۷).

راثوره^۴ و همکاران (۲۰۱۴) در مطالعه ای با عنوان "پیش بینی میزان بقای سرطان سینه با استفاده از رویکرد گروهی" از تکنیک های درخت تصمیم، بیزین و دسته بندی مبتنی بر انجمنی برای پیش بینی میزان بقا در بیماران مبتلا به سرطان سینه استفاده کرد. سپس یک سیستم رای گیری بر مبنای اجماع نظرات این تکنیک ها به همراه رویکرد پیشنهادی بر روی داده هایی که در بازه ی زمانی ۱۹۷۳ تا ۲۰۰۰ از پایگاه داده ی جهانی برنامه ی نظارت، اپیدمیولوژی و نتایج نهایی در کشور آمریکا جمع آوری شده بود، پیاده سازی شد. داده های

1 - Thongkam

2 - Kim

3 - Park

4 - Rathore

جمع آوری شده شامل ۴۳۳۲۷۲ و ۷۲ متغیر است. نتایج نشان می دهد که رویکرد پیشنهادی نسبت به بقیه ی تکنیک های کلاسه بندی بهتر عمل می کرد (۳۸).

وانگ^۱ و همکاران (۲۰۱۴) در مطالعه ای با عنوان " یک رویکرد ترکیبی با استفاده از تکنیک بیش نمونه سازی ترکیبی اقلیت^۲ و تئوری مورچگان برای پیش بینی میزان بقای ۵ ساله در بیماران مبتلا به سرطان سینه " از تکنیک های درخت تصمیم، k – نزدیکترین همسایه و رگرسیون لجستیک برای پیش بینی میزان بقا ۵ ساله در بیماران مبتلا به سرطان سینه استفاده کرد. همچنین در این مدل از تئوری مورچگان جهت انتخاب صفات کلیدی استفاده شده است. علاوه بر این تکنیک های بیش نمونه برداری بیش نمونه سازی ترکیبی اقلیت استفاده شده است. این تکنیک ها بر روی داده هایی که در بازه ی زمانی ۱۹۷۳ تا ۲۰۰۷ از پایگاه داده ی جهانی برنامه ی نظارت، اپیدمیولوژی و نتایج نهایی در کشور آمریکا جمع آوری شده بود، پیاده سازی شدند. این مجموعه شامل ۹۷۳۱۲۵ رکورد و ۱۱۸ متغیر انتخاب شد. نتایج نشان دهنده ی ترکیب تئوری مورچگان به همراه درخت تصمیم C5 و تکنیک بیش نمونه برداری بیش نمونه سازی ترکیبی اقلیت عملکرد بهتری خواهند داشت (۳۹).

جمع بندی پیشینه پژوهش

طبق بررسی های انجام شده در پژوهش های داخلی و خارجی، غالباً روش های دسته بندی به صورت مجزا بررسی شده اند و کمتر ترکیب ها آنها با دیگر تکنیک ها مورد نظر قرار گرفته است. در واقع بیشتر پژوهش ها به دنبال مقایسه عملکرد تکنیک های داده کاوی و انتخاب یکی از آنها بوده اند که در برخی مواقع، درخت تصمیم و در موارد دیگر شبکه ی عصبی و همچنین در برخی دیگر ماشین بردار پشتیبان عملکرد بهتری داشته است. همچنین تاثیر نوع مجموعه های داده ای و میزان توازن و عدم توازن داده ها کمتر در نظر گرفته شده است. این مسئله با توجه به مطالب فوق قابل توجه است چون نمی توان گفت که یک تکنیک برای همه ی مجموعه های داده ای مناسب است. غالباً داده هایی مورد بررسی و استفاده قرار گرفته در پژوهش های داخلی و خارجی که مربوط به پایگاه داده برنامه ی نظارت، اپیدمیولوژی و نتایج نهایی (SEER) در کشور آمریکا است. با توجه به موارد ذکر شده در خصوص مطالعات داخلی و خارجی و با در نظر گرفتن این نکته که تاکنون داده های مرتبط با بیماران مبتلا به سرطان سینه در سطح شهر اصفهان مورد بررسی قرار نگرفته است، پژوهشگران قصد دارند از طریق ترکیب بهینه تکنیک های بکار رفته در مراحل مختلف داده کاوی، میزان بقای بیماران مبتلا به سرطان سینه را به صورت دقیق و بهینه پیش بینی کنند.

تعریف واژه ها

سرطان سینه (Breast Cancer)

تعریف نظری: مجموعه بیماری هایی است که از تکثیر کنترل و مهار نشده سلول ها در بافت سینه ایجاد می شود (۴۰)
تعریف عملیاتی: در این پژوهش، بیمارانی که در شهر اصفهان، ابتلا به سرطانشان مورد تایید قرار گرفته است، در نظر گرفته شده اند.

داده کاوی (Data Mining)

تعریف نظری: استخراج الگوهای جذاب و اطلاعات نهان از حجم عظیمی از داده را داده کاوی می گویند (۴۱).
تعریف عملیاتی: مطابق تعریف نظری است.

¹ - Wang

² SMOTE (Synthetic Over-Sampling Technique)

پیش بینی میزان بقا (Survivality Prediction)

تعریف نظری: مدت زمان ادامه زندگی افراد براساس تعداد سال پس از مبتلا شدن به سرطان سینه را پیش بینی می کند (۱۹).
تعریف عملیاتی: در این پژوهش، پیش بینی میزان بقا در بیماران مبتلا به سرطان در شهر اصفهان مورد بررسی قرار گرفته است.

هدف کلی

پیش بینی احتمال بقا در بیماران مبتلا به سرطان سینه با استفاده از تکنیک های داده کاوی

اهداف اختصاصی

- ۱- انتخاب عوامل اصلی تاثیر گذار بر بقای بیماران مبتلا به سرطان سینه
- ۲- پاکسازی داده ی بیماران مبتلا به سرطان سینه
- ۳- پیش بینی احتمال بقا بیماران مبتلا به سرطان سینه با استفاده از تکنیک درخت تصمیم
- ۴- پیش بینی احتمال بقا بیماران مبتلا به سرطان سینه با استفاده از تکنیک شبکه ی عصبی
- ۵- پیش بینی احتمال بقا بیماران مبتلا به سرطان سینه با استفاده از تکنیک ماشین بردار پشتیبان
- ۶- تعیین ترکیب بهینه تکنیک های بکار رفته در مراحل مختلف داده کاوی جهت پیش بینی احتمال بقای بیماران مبتلا به سرطان سینه

سوالات پژوهش:

- ۱- عوامل اصلی تاثیر گذار بر بقای بیماران مبتلا به سرطان سینه شامل چه مواردی هستند؟
- ۲- پاکسازی داده های بیماران مبتلا به سرطان سینه چگونه انجام می شود؟
- ۳- احتمال بقا بیماران مبتلا به سرطان سینه با استفاده از تکنیک درخت تصمیم چه میزان می باشد؟
- ۴- احتمال بقا بیماران مبتلا به سرطان سینه با استفاده از تکنیک شبکه ی عصبی چه میزان می باشد؟
- ۵- احتمال بقا بیماران مبتلا به سرطان سینه با استفاده از تکنیک ماشین بردار پشتیبان چه میزان می باشد؟
- ۶- ترکیب بهینه تکنیک های بکار رفته در مراحل مختلف داده کاوی جهت پیش بینی احتمال بقای بیماران مبتلا به سرطان سینه چیست؟

روش شناسی پژوهش

نوع پژوهش: این مطالعه از نظر هدف کاربردی - توسعه ای می باشد که به شیوه ی مقطعی انجام خواهد شد.

جامعه آماری: تمامی بیماران مبتلا به سرطان سینه در سطح شهر اصفهان که اطلاعات مربوط به بیماری آنها در بیمارستان سید الشهدای اصفهان موجود می باشد.

معیارهای ورود و خروج:

معیار ورود به مطالعه:

۱) تمامی بیماران زن مبتلا به سرطان سینه در سطح شهر اصفهان که داده های مربوط به عوامل تاثیرگذار بر احتمال بقای آنها (استخراج شده براساس هدف ۱) در بیمارستان سیدالشهدا موجود باشد.

۲) زنانی که در فاصله ی بین سال های ۱۳۹۱ تا ۱۳۹۶ به سرطان سینه مبتلا شده اند و اطلاعات آنها در این بیمارستان موجود است.

حجم نمونه و روش نمونه گیر

در این مطالعه، نمونه گیری انجام نخواهد شد و اطلاعات مربوط به کلیه ی بیماران مبتلا به سرطان سینه (طبق معیارهای ورود) به روش تمام شماری مورد بررسی و استفاده قرار خواهد گرفت.

روش جمع آوری اطلاعات:

جمع آوری داده ها در دو مرحله انجام می شود :

۱) انتخاب زیر مجموعه ای از صفات

۲) پاک سازی داده ها

مجموعه ی داده ای مستخرج از اطلاعات بیماران در بیمارستان، اطلاعات را به صورت ماتریسی ذخیره می کند. مرحله ی اول، انتخاب یک زیر مجموعه از صفات است. با توجه به اینکه صفات زیادی در مجموعه ی داده ای وجود دارد و ممکن است خیلی از آنها مرتبط با هدف تحقیق نباشد، باید از بین آنها، یکسری از صفات انتخاب شوند که بیشترین ارتباط با متغیر هدف تحقیق دارند. به همین منظور، در ابتدا لیستی از عوامل تاثیرگذار بر احتمال بقای بیماران زن مبتلا به سرطان سینه با استفاده از بررسی متون استخراج خواهد شد سپس با بهره گیری از نظر یکی از متخصصین سرطان، عوامل اصلی تاثیرگذار انتخاب خواهند شد. در ادامه پس از اخذ مجوزهای لازم از مراجع ذی ربط در دانشگاه علوم پزشکی اصفهان، محققین به بیمارستان سیدالشهدای اصفهان مراجعه خواهند کرد و لیستی از پرونده های بیماران زن مبتلا به سرطان سینه در فاصله ی سال های ۱۳۹۱ تا ۱۳۹۶ را استخراج خواهند کرد. پس از بررسی پرونده ی این بیماران، داده های مرتبط با عوامل اصلی تاثیرگذار بر احتمال بقا استخراج خواهند شد. داده های اولیه بیماران زن که در فاصله ی سال های ۱۳۹۱ تا ۱۳۹۶ به سرطان سینه مبتلا شده اند از بیمارستان سیدالشهدای اصفهان جمع آوری می شود.

مرحله دوم، پاک سازی داده است. در این مرحله، تشخیص مقادیر از دست رفته و حذف یا اصلاح آن انجام می شود. مقادیر از دست رفته برای صفات، مقادیری هستند که مقدار آنها مشخص نیست. هم اکنون بسیاری از این داده ها به دلیل خطای کاربر در ورود داده و یا مشخص نشدن آن توسط پزشک و یا بیمار ایجاد می شوند که می توانند مشکلات زیادی در دقت و صحت روش استفاده شده بوجود آورند. می توان از رویکرد های مختلفی از قبیل محتمل ترین مقادیر یا محاسبه ی مقدار میانگین برای تعیین مقدار مقادیر از دست رفته استفاده کرد. یک رویکرد دیگر می تواند حذف رکورد متناظر از مجموعه ی داده ای باشد. با توجه به اینکه انتخاب یکی از این رویکردها، متناسب با داده ی مورد استفاده خواهد بود، در این پژوهش، یک یا ترکیبی از این تکنیک ها برای پاک سازی داده ها استفاده خواهد شد.

روش تجزیه و تحلیل داده ها:

پس از جمع آوری اطلاعات مورد نیاز، با استفاده از نرم افزار Rapidminer داده ها دسته بندی خواهند شد. پیش از دسته بندی داده ها، مجموعه ی داده ای مورد نظر به دو قسمت مجموعه ی داده ای آموزشی و مجموعه ی داده ای تست تقسیم خواهد شد. مجموعه ی داده ای آموزشی جهت استخراج الگوی مناسب برای تعیین میزان بقا براساس صفات مورد استفاده بکار گرفته می شود. سپس این الگو بر روی

مجموعه ی داده ای تست پیاده سازی می شود. نحوه تقسیم بندی مجموعه ی داده ای به دو قسمت مجموعه ی داده ای آموزشی و تست بدین شکل است که ۷۰ درصد مجموعه ی داده ای به عنوان مجموعه ی داده ای آموزشی و ۳۰ درصد آن به عنوان مجموعه ی داده ای تست استفاده می شود.

این نرم افزار از تکنیک های مختلفی برای دسته بندی اطلاعات استفاده می کند. تکنیک های مورد استفاده در این مطالعه شامل درخت تصمیم، ماشین بردار پشتیبان و شبکه های عصبی خواهد بود. تکنیک های ذکر شده از مهمترین و پرکاربردترین تکنیک های دسته بندی هستند که معمولاً مستقلاً بر روی مجموعه های داده ای استفاده شده و کمتر به ترکیب آنها با دیگر تکنیک های موجود در دنیای کامپیوتر توجه می شود. هر یک از تکنیک ها بر روی مجموعه ی داده ای آموزشی پیاده سازی می شوند. سپس الگوی میزان بقا استخراج شده و بر روی مجموعه ی داده ای تست مورد آزمایش و ارزیابی قرار می گیرد.

ارزیابی تکنیک های استفاده شده براساس مجموعه ی داده ای تست انجام می شود. بدین صورت که ابتدا یک ماتریس پیچیدگی محاسبه می شود که شامل معیار های مختلفی از قبیل صحیح مثبت^۱، صحیح منفی^۲، غلط مثبت^۳ و غلط منفی^۴ است. صحیح مثبت، تعداد رکوردهایی را مشخص می کند که دسته بند آنها را به طور صحیح جز کلاس نهایی قرار داده است. صحیح منفی، رکوردهایی را مشخص می کند که دسته بند آنها را به طور صحیح جز کلاس نهایی قرار نداده است. غلط مثبت، رکوردهایی را مشخص می کند که دسته بند آنها را به اشتباه جز کلاس نهایی قرار داده است. غلط منفی، رکوردهایی را مشخص می کند که دسته بند آنها را به اشتباه جز کلاس نهایی قرار نداده است. سپس براساس این ماتریس، معیارهای برای کارایی تکنیک داده کاوی محاسبه می شود. یکی از این معیار ها صحت است که صحت براساس فرمول ذیل محاسبه می شود :

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

در واقع صحت میزان تشخیص درست را نسبت به کل تشخیص ها می سنجد. در صورت نیاز، سایر معیارها از قبیل دقت و حساسیت نیز مورد بررسی قرار خواهد گرفت.

ملاحظات اخلاقی:

- محترم شمردن قوانین و مقررات خاص بیمارستان های مورد پژوهش در هنگام جمع آوری داده
- تضمین محرمانه بودن کلیه ی اطلاعات و نظرات افراد
- بی طرفی و پرهیز از گرایش های خاص توسط پژوهشگر

محدودیت ها و مشکلات اجرایی پژوهش :

به دلیل عدم ثبت داده های مورد نیاز به شیوه الکترونیک، فرآیند بازیابی و استفاده از پرونده های پزشکی بسیار وقت گیر و پرهزینه خواهد بود. محققین به منظور رفع این چالش در صورت امکان از نیروهای داوطلب این مرکز برای جمع آوری داده استفاده خواهند کرد.

¹ TP : True Positive

² TN : True Negative

³ FP : False Positive

⁴ FN : False Negative

روش اندازه گیری	واحد	نوع متغیر				نام متغیر
		کیفی	کمی			
			اسمی /رتبه ای	پیوسته		
تکنیک داده کاوی				گسسته	وابسته	میزان بقا
نظر متخصص و بررسی متون		اسمی / رتبه ای	پیوسته	گسسته	مستقل	عوامل تاثیر گذار بر بقا

- 10-Zendehdel K. Cancer Statistics in IR Iran in 2018. Basic & Clinical Cancer Research. 2019 Jan 2;11(1).
- 11- Akbari ME, Khayamzadeh M, Khoshnevis S, Nafisi N, Akbari A. Five and ten years survival in breast cancer patients mastectomies vs. breast conserving surgeries personal experience. Iranian Journal of Cancer Prevention 2008;1(2):53-6.
- ۱۲-دهقان، پروانه؛ مقربی، مانده؛ ذباح، ایمان؛ لایقی، کامران؛ ماروسی، علی. مدلسازی بیماری سرطان پستان با استفاده از روشهای مبتنی بر داده کاوی. مجله انفورماتیک سلامت و زیست پزشکی، مرکز تحقیقات انفورماتیک پزشکی، ۱۳۹۶؛ ۴(۴): ۲۶۸-۲۷۸
- ۱۳-صبوری، اسماعیلی، حبیب الله، شهید ثالث، عمادی. بررسی عوامل مرتبط با بقای بیماران مبتلا به سرطان کولورکتال با استفاده از رگرسیون کاکس. مجله دانشکده پزشکی دانشگاه علوم پزشکی مشهد، ۱۳۹۷؛ ۶۱(۴): ۱۰۸۲-۱۰۸۹.
- ۱۴- حقیقت شهر، الفت بخش آسیه، سجادیان اکرم السادات، مهرداد ندا، انصاری مریم، ابراهیمی ماندانا، هاشمی عصمت السادات، مختاری حصاری پریرا. بررسی میزان بقا و عوامل مرتبط با آن در بیماران سرطان پستان مراجعه کننده به مرکز بیماری های پستان. فصلنامه بیماری های پستان ایران، ۱۳۹۲؛ ۶(۳).
- 15-Forouzanfar MH, Foreman KJ, Delossantos AM, Lozano R, Lopez AD, Murray CJ, et al. Breast and cervical cancer in 187 countries between 1980 and 2010: a systematic analysis. Lancet 2011; 378(9801): 1461-84.
- ۱۶- ستاره، سوگند، ظهیری اصفهانی، میثاق، زارع بندامیری، رئیسی، عباسی. پیش بینی بقا بیماران مبتلا به سرطان روده بزرگ در بخش پرتو درمانی بیمارستان نمازی شیراز با استفاده از روش های داده کاوی ماشین بردار پشتیبان و بگینگ. مجله تخصصی اپیدمیولوژی ایران، ۱۳۹۷؛ ۱۴(۱): ۱۹-۲۹.
- 17-Goli S, Mahjub H, Faradmal J, Mashayekhi H, Soltanian AR. Survival prediction and feature selection in patients with breast cancer using support vector regression. Computational and mathematical methods in medicine. 2016;2016.
- 18- Kim H, and Bredel M. "Feature selection and survival modeling in The Cancer Genome Atlas. International Journal of Nanomedicine, vol. 8, supplement 1, pp. 57-62, 2013.
- 19- Van Belle V, Pelckmans K, Van Huffel S, Suykens JAK. Support vector methods for survival analysis: a comparison between ranking and regression approaches. Artificial Intelligence in Medicine, 2011; 53(2): 107-118, 2011.
- 20- Ding Z. X. "The application of support vector machine in survival analysis," in Proceedings of the 2nd International Conference on Artificial Intelligence, Management Science and Electronic Commerce (AIMSEC '11), pp. 6816-6819, IEEE, Dengfeng, China, August 2011.
- 21-Delen D, Walker G, Kadam A. Predicting breast cancer survivability: a comparison of three data mining methods. Artificial intelligence in medicine. 2005 Jun 1; 34(2):113-27.
- 22-Liu YQ, Wang C, Zhang L. Decision tree based predictive models for breast cancer survivability on imbalanced data. In 2009 3rd international conference on bioinformatics and biomedical engineering 2009 Jun 11 (pp. 1-4). IEEE.
- 23-Noori Dalooi MR, Tabarestani S. Molecular genetics, diagnosis and treatment of breast cancer: Review article. J Sabzevar Univ Med Sci 2010; 17(2): 74-87

- 24-Mirmalek SA, Elham Kani F. Clinical application of breast cancer biology review of literature. Iran J Surg 2009; (17): 1-6. .
- 25- Kleinbaum DG, Klein M. Survival analysis: a self-learning text: Springer; 2005.
- 26-Salehi M, Gohari MR, Vahabi N, Zayeri F, Yahyazadeh SH, Kafashian MR. Comparison of artificial neural network and cox regression models in survival prediction of breast cancer patients,2013.
- 27-Afshar HL, Ahmadi M, Roudbari M, Sadoughi F. Prediction of breast cancer survival through knowledge discovery in databases. Global journal of health science. 2015 Jul;7(4):392.
- 28-Khalkhali HR, Lotfnezhad Afshar H, Esnaashari O, Jabbari N. Applying data mining techniques to extract hidden patterns about breast cancer survival in an Iranian cohort study. Journal of research in health sciences. 2016 Mar 21;16(1):31-5.
- 29-Montazeri M, Montazeri M, Montazeri M, Beigzadeh A. Machine learning models in breast cancer survival prediction. Technology and Health Care. 2016 Jan 1;24(1):31-42.
- 30-Tapak L, Shirmohammadi-Khorram N, Amini P, Alafchi B, Hamidi O, Poorolajal J. Prediction of survival and metastasis in breast cancer patients using machine learning classifiers. Clinical Epidemiology and Global Health. 2018 Oct 6.
- 31-Miri Rostami S, Ahmadzadeh M. Extracting predictor variables to construct breast cancer survivability model with class imbalance problem. Journal of AI and Data Mining. 2018 Jul 1;6(2):263-76.
- 32-Bellaachia A, Guven E. Predicting breast cancer survivability using data mining techniques. Age. 2006;58(13):10-10.
- 33-Endo A, Shibata T, Tanaka H. Comparison of seven algorithms to predict breast cancer survival (< special issue> contribution to 21 century intelligent technologies and bioinformatics). International Journal of Biomedical Soft Computing and Human Sciences: the official journal of the Biomedical Fuzzy Systems Association. 2008;13(2):11-6
- 34-Khan U, Shin H, Choi JP, Kim M. wFDT: weighted fuzzy decision trees for prognosis of breast cancer survivability. InProceedings of the 7th Australasian Data Mining Conference-Volume 87 2008 Nov 27 (pp. 141-152). Australian Computer Society, Inc..
- 35-Thongkam J, Xu G, Zhang Y, Huang F. Toward breast cancer survivability prediction models through improving training space. Expert systems with Applications. 2009 Dec 1;36(10):12200-9.
- 36-Kim J, Shin H. Breast cancer survivability prediction using labeled, unlabeled, and pseudo-labeled patient data. Journal of the American Medical Informatics Association. 2013 Mar 6;20(4):613-8.
- 37-Park K, Ali A, Kim D, An Y, Kim M, Shin H. Robust predictive model for evaluating breast cancer survivability. Engineering Applications of Artificial Intelligence. 2013 Oct 1;26(9):2194-205.
- 38-Rathore N, Agarwal S. Predicting the survivability of breast cancer patients using ensemble approach. In2014 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT) 2014 Feb 7 (pp. 459-464). IEEE.
- 39-Wang KJ, Makond B, Chen KH, Wang KM. A hybrid classifier combining SMOTE with PSO to estimate 5-year survivability of breast cancer patients. Applied Soft Computing. 2014 Jul 1;20:15-24.

۴۱- محمد صنیعی آباده، سینا محمودی، محدثه طاهرپور، داده کاوی کاربردی، ، چاپ سوم، انتشارات نیاز دانش، ۱۳۹۱

- 42-National Cancer Institute, Available at <https://www.cancer.gov/types/breast> .

جدول زمانی مراحل اجرا و پیشرفت کار

ردیف	نوع فعالیت	فرد مسئول	طول مدت به ماه	زمان اجرا											
				ماه 1	ماه 2	ماه 3	ماه 4	ماه 5	ماه 6	ماه 7	ماه 8	ماه 9	ماه 10	ماه 11	ماه 12
1	مرور متون		۱	*											
2	انتخاب عنوان		۱	*											
3	تدوین پروپوزال		۲		*										
4	گردآوری		۱			*	*								
5	تحلیل		۱			*	*	*							
6	گزارش نهایی		۲			*									
7	مقاله		۱			*									
	جمع =		۹												

																ماه
--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	-----

اطلاعات مربوط به هزینه‌ها

آیا برای این طرح از سازمانهای دیگر نیز درخواست اعتبار شده است؟ ☐ بلی ☒ خیر

جدول ۳: هزینه پرسنلی با ذکر مشخصات کامل و میزان اشتغال هر یک و حق الزحمه آنها:

نوع فعالیت	تعداد افراد	درجه تحصیلی	مجموع ساعات	حق الزحمه در ساعت	جمع
جمع آوری داده	۲	استادیار	۱۲۰	۱۴۰۰۰۰	۱۶,۸۰۰,۰۰۰
ورود اطلاعات*	۲	استادیار	۲۲	۱۴۰۰۰۰	۳۰,۸۰۰,۰۰۰
تجزیه و تحلیل اطلاعات	۱	استادیار	۳۰	۱۴۰۰۰۰	۴۲,۰۰۰,۰۰۰
جمع کل					24,080,000

* تعداد نمونه ۱۵۴۰ است تعداد متغیر ۲۰ است. حق الزحمه ورود هر داده ۱۰۰ ریال است.

$$۱۰۰ * ۲۰ * ۱۵۴۰ = 3080000$$

جدول ۴: هزینه آزمایشات و خدمات تخصصی :

موضوع آزمایش یا خدمات تخصصی	مرکز سرویس دهنده	تعداد کل دفعات	هزینه برای هر دفعه	جمع (ریال)
تکثیر فرم جمع آوری داده	زیراکس دانشکده مدیریت	۱۵۴۰	۳۰۰۰	۴۶۲۰,۰۰۰
جمع هزینه های آزمایشات و خدمات تخصصی				۴۶۲۰,۰۰۰

جدول ۵: فهرست وسایل و موادی که باید از اعتبار این طرح از داخل یا خارج کشور خریداری شود.

نام دستگاه و یا مواد	شرکت سازنده یا فروشنده	کشور	مصرفی یا غیرمصرفی	تعداد یا مقدار لازم	قیمت کل
نرم افزار رپید ماینر*					۶۰۰۰۰۰
جمع کل					۶۰۰۰۰۰

* به منظور استفاده از لایسنس اورجینال نرم افزار و همه ی قابلیت های موجود در نرم افزار

جدول ۶: هزینه مسافرت (در صورت لزوم)

مقصد	دفعات مسافرت در مدت اجرای طرح و منظور آن	نوع وسیله نقلیه	تعداد افراد	هزینه به ریال
بیمارستان سید الشهدا	۲۰ مرتبه رفت و برگشت	آژانس	۲	3000000

۴-۶ هزینه های دیگر :

الف) هزینه تکثیر نشریات، کتاب و پرسشنامه ها: ۰ ریال
ب) سایر هزینه ها 2000000 ریال

2000000 ریال

جمع هزینه های دیگر :

جدول ۶: جمع هزینه های طرح

الف) جمع هزینه های پرسنل	24080000 ریال
ب) جمع هزینه های آزمایشات و خدمات تخصصی	4620000 ریال
پ) جمع هزینه های مسافرت	3000000 ریال
ت) جمع هزینه های وسایل	6000000 ریال
ث) جمع هزینه های دیگر	2000000 ریال
ج) جمع کل	39700000 ریال

مبلغی که از منابع دیگر کمک خواهد شد و نحوه مصرف آن :

باقیمانده هزینه‌های طرح که تامین آن درخواست می‌شود:

صحت مطالب مندرج در پرسشنامه را تایید می‌نمایم.

محل امضای معاون تحقیقات و فناوری دانشگاه

محل امضای مجری اصلی طرح

محدودیت ها و مشکلات اجرای طرح و روش حل آنها

پیوست ها:

۱- در صورت نیاز به اخذ رضایت نامه کتبی از واحدهای مورد پژوهش، نمونه‌ای از فرم مذکور پیوست گردد.

۲- در صورتی که روش و یا ابزار گردآوری داده‌ها پرسشنامه، سیاهه واریسی و یا کارت مشاهده می‌باشد. لطفاً نمونه‌ای از آن پیوست شود. در صورتی که ابزار را شما ترجمه یا بومی نموده اید اصل ابزار نیز پیوست شود.

۳- در صورت نیاز موافقت کتبی محیط پژوهش اخذ و ضمیمه گردد

