مقاله پژوهشی

مجله انفورماتیک سلامت و زیست پزشکی
مرکز تحقیقات انفورماتیک پزشکی
دوره دوم، شماره سوم. ص ۱۴۰-۱۳۳، ۱۳۹۴

آینده پژوهی در سلامت: انتخاب بهترین مدل هوشمند مبتنی بر داده کاوی برای پیش‌بینی و تشخیص سرطان کبد در مراحل اولیه

فائزه افضلی[۱]، زهره حیدری[۱]، میترا منتظری[۲،۳]، لیلا احمدیان[۴*]، محمد جواد زاهدی[۵]

• دریافت مقاله: ۹۳/۱۱/۲۸ • پذیرش مقاله: ۹۴/۶/۲۵

مقدمه: سرطان اولیه کبد (HCC) پنجمین سرطان شایع در دنیا و سومین عامل مرگ و میر در جهان می‌باشد. علائم سرطان کبد پس از بروز به سرعت پیشرفت کرده و در صورت عدم تشخیص به موقع متأسفانه بقای عمر بیمار بسیار کم می‌گردد. یکی از مشکلات اصلی پیش روی متخصصین گوارش، پیش بینی و تشخیص زود هنگام سرطان کبد است. داده کاوی از روش‌هایی است که در این زمینه مورد استفاده واقع می‌گردد. هدف از انجام این مطالعه معرفی بهترین مدل هوشمند مبتنی بر داده کاوی برای پیش‌بینی و تشخیص سرطان کبد در مراحل اولیه می‌باشد.

روش: در مقاله حاضر با استفاده از روش مطالعه گذشته نگر، پرونده ۵۱۶ بیمار مبتلا به سرطان کبد اولیه و ثانویه و ۲۲ ریسک فاکتور از هر بیمار، مورد بررسی قرار گرفت. داده‌های جمع‌آوری شده با استفاده از ۵ مدل داده کاوی VFI Classifier، Regression Classifier، HyperPipes Classifier، Functional trees with logistic regression و Meta Muti Class Classifier تحلیل شدند. این مدل‌ها با یکدیگر مقایسه گردیدند.

نتایج: دقت، ویژگی، حساسیت و سطح زیر منحنی Roc مدل VFI Classifier به ترتیب ۷۱/۲۹٪، ۴۹٪، ۵۰٪ و ۶۳/۳۱٪ می‌باشد و این مدل به عنوان بهترین مدل هوشمند مبتنی برداده کاوی برای پیش بینی و تشخیص سرطان کبد در مراحل اولیه شناخته شد.

نتیجه‌گیری: در صورتی که مدل داده‌کاوی VFI Classifier به صورت صحیح طراحی شود، می‌تواند سرطان کبد را پیش بینی نماید یا آن را در مراحل اولیه تشخیص دهد.

کلید واژه‌ها: سرطان کبد، هپاتوسلولار کارسینوما (HCC)، پیش‌بینی و تشخیص، داده‌کاوی، آینده پژوهی در سلامت

•ارجاع: افضلی فائزه، حیدری زهره، منتظری میترا، احمدیان لیلا، زاهدی محمد جواد. آینده پژوهی در سلامت: انتخاب بهترین مدل هوشمند مبتنی بر داده کاوی برای پیش‌بینی و تشخیص سرطان کبد در مراحل اولیه. مجله انفورماتیک سلامت و زیست پزشکی ۱۳۹۴؛ ۲(۳): ۱۴۰-۱۳۳.

۱. کارشناس فناوری اطلاعات سلامت، مرکز تحقیقات مدل سازی در سلامت، پژوهشکده آینده پژوهی در سلامت، دانشگاه علوم پزشکی کرمان، کرمان، ایران.
۲. کارشناس ارشد هوش مصنوعی، مرکز تحقیقات انفورماتیک پزشکی، پژوهشکده آینده پژوهی در سلامت، دانشگاه علوم پزشکی کرمان، کرمان، ایران.
۳. کارشناس ارشد هوش مصنوعی، بخش مهندسی کامپیوتر، دانشکده فنی و مهندسی، دانشگاه شهید باهنر کرمان، کرمان، ایران.
۴. دکترای تخصصی انفورماتیک پزشکی، دانشیار، مرکز تحقیقات انفورماتیک پزشکی، پژوهشکده آینده پژوهی در سلامت، دانشگاه علوم پزشکی کرمان، کرمان، ایران.
۵. فوق تخصص گوارش، استاد، مرکز تحقیقات گوارش و کبد، دانشگاه علوم پزشکی کرمان، کرمان، ایران.
*** نویسنده مسؤول:** کرمان، ابتدای هفت باغ، پردیس دانشگاه علوم پزشکی، پژوهشکده آینده پژوهی در سلامت، مرکز تحقیقات انفورماتیک پزشکی
• **شماره تماس:** ۰۳۴۳۱۳۲۵۴۰۶ • **Email:** ahmadianle@yahoo.com

[Downloaded from jhbmi.ir on 2026-11-17]

مقدمه

کبد بزرگترین غده بدن [۱] و یکی از مهم‌ترین اعضای بدن است که سم‌زدایی از داروها، دفع محصولات زاید، تولید عوامل انعقادی خون، سوخت و ساز قند و چربی و ... را بر عهده دارد [۲،۳] از آنجایی که محل طبیعی تولید اخلاط در کبد است، هر گونه بیماری کبد و عوارض ناشی از آن تمامی بدن را تحت تأثیر قرار خواهد داد. لذا درمان به موقع و درست بیماری‌های آن ضروری است [۴].

تومورهای کبد ممکن است اولیه یا ثانویه باشند که تومورهای اولیه به طور معمول در بیمارانی که بیماری مزمن کبدی (سیروز) دارند اتفاق می‌افتد [۱].

سرطان کبد پنجمین سرطان شایع و سومین علت مرگ و میر به علت سرطان است [۵،۶] به عنوان مثال هپاتوسلولار کارسینوما (Hepatocellular carcinoma) نوعی سرطان اولیه کبد است [۲] که یکی از مهم‌ترین سرطان‌های رایج در دنیا به خصوص در آفریقا و شرق آسیا می‌باشد که بروز آن در جهان در حال افزایش است. این نوع سرطان کبد سومین و هفتمین عامل مرگ ناشی از سرطان در بین مـردان و زنان می‌باشد و سالانه باعث مرگ یک میلیون نفر در سراسر دنیا می‌شود [۶].

آمارها نشان می‌دهد نرخ سرطان کبد در مردان مبتلا به هپاتیت B و C به طور معمول ۲ تا ۴ برابر بیشتر از زنان مبتلا است [۷]. قوی‌ترین عوامل خطر برای توسعه HCC هپاتیت مزمن B (Hepatitis B virus)، عفونت ویروس هپاتیت C (Hepatitis C virus) و همچنین سیروز کبدی است. عوامل خطر دیگری چون قرار گرفتن در معرض آفلاتوکسین، الکل، توتون و تنباکو، چاقی، دیابت، قرص‌های ضد بارداری، مقاومت به انسولین، اضافه بار آهن، هیپوتیروئیدی، چربی خون و پـراکـسیداسیـون لیپیدی می‌باشد [۱،۸].

امروزه در حوزه علوم پزشکی پیشرفت‌های زیادی صورت گرفته است [۹]، اما همچنان پیش‌بینی و تشخیص زود هنگام سرطان کبد از مشکلات اصلی پیش روی پزشکان می‌باشد. هوش مصنوعی با به کارگیری طیف زیادی از تکنیک‌های هوشمند توانسته است در زمینه‌های مختلفی کاربرد داشته باشد [۱۰]. در حیطه علم پزشکی نیز با استفاده از سیستم‌های تصمیم یار به کمک پزشک در این زمینه پرداخته شده است [۱۱،۱۲].

امروزه کشف دانش از پایگاه داده با استفاده از ابزارهای داده کاوی در تمامی حوزه‌های علمی کاربرد دارد [۱۳،۱۴] که یکی از این حوزه‌ها، علوم پزشکی است. با استفاده از این ابزارها محققان می‌توانند ارتباط بین متغیرها را شناسایی و وقوع سرطان را پیش‌بینی کنند [۱۵].

در مورد کاربرد داده کاوی در تشخیص و پیش‌بینی بیماری‌های مختلف می‌توان به مطالعه Lin اشاره کرد که نتیجه آن تشخیص زود هنگام بیماری کبد، تشخیص انواع بیماری‌های کبدی، بازیابی موارد مشابه، حمایت از درمان بیماری کبد و افزایش دقت تشخیصی به میزان ۹۰٪ با استفاده از روش‌های هوشمند می‌باشد [۱۶].

در مطالعه منتظری و همکاران از ۸ روش هوشمند داده کاوی به منظور پیدا کردن خطر فاکتورهای مؤثر در سرطان کبد استفاده گردید [۱۷]. در این مطالعه پیدا کردن ریسک فاکتور مهم از بین تمامی ریسک فاکتورها مسئله مهمی در داده کاوی به شمار می‌آید [۱۸،۱۹].

در مطالعه Avci و همکاران یک سیستم هوشمند برای تشخیص بیماری دریچه قلب معرفی شد که دقت طبقه بندی این سیستم هوشمند برای افراد سالم ۹۶٪ و برای افراد غیر طبیعی ۹۳/۱٪ می‌باشد [۲۰].

هدف از مطالعه Castanho و همکاران بررسی عملکرد برخی از سیستم‌های خبره فازی در طبقه بندی بیماران با سرطانی محدود یا غیر محدود می‌باشد. این سیستم برای پیش بینی مرحله سـرطان پـروستات مورد استـفاده قرار می‌گیرد [۲۱].

Pinheiro و همکاران کاربرد تکنیک‌های داده کاوی را به منظور تعیین عوامل در ارتباط با بروز سرطان کبد بررسی کردند [۲۲].

در ایران جمارانی و همکاران مدل هوشمندی برای تشخیص سرطان پستان با استفاده از ترکیبی از شبکه عصبی مصنوعی و تجزیه تصویر ارائه نمودند [۲۳].

در مطالعه علی‌پور و همکاران یک سیستم هوشمند برای تشخیص دقیق سرطان پستان با استفاده از ۲۲ ویژگی به دقت شناسایی ۱۰۰٪ دست یافت. این سیستم از لحاظ دقت و تعداد ویژگی‌های مورد نیاز بر سیستم‌های موجود برتری دارد [۲۴].

از آنجایی که شیوع سرطان کبد در ایران در حال افزایش است و با توجه به این که سرطان کبد در مراحل انتهایی این بیماری تشخیص داده می‌شود بر آن شدیم بهترین مدل هوشمندی که می‌تواند سرطان کبد را پیش‌بینی کرده و آن را در مراحل اولیه تشخیص دهد معرفی کنیم تا بدین گونه سال‌های عمر بیمار را

[Downloaded from jhbmi.ir on 2026-11-17]

افزایش داده و کیفیت زندگی وی را بهبود بخشیم.

روش

در مقاله حاضر با استفاده از روش مطالعه گذشته نگر، با حفظ محرمانگی اطلاعات بیماران، داده‌های مورد نیاز از پرونده ۳۶۶ بیمار مبتلا به سرطان کبد اولیه و ۱۵۰ بیمار مبتلا به سرطان کبد ثانویه استخراج گردید.

جامعه مورد مطالعه بیماران مبتلا به سرطان کبد می‌باشند که در فاصله زمانی سال ۹۳-۱۳۸۱ به بیمارستان افضلی پور شهر کرمان مراجعه کردند.

از هر پرونده ۲۲ ریسک فاکتور مربوط به سرطان کبد از قسمت اطلاعات هویتی، بالینی و برگه آزمایش بیمار استخراج گردید. این ریسک فاکتورها در جدول ۱ نمایش داده شدند. میزان ۱۲/۵۴٪ از مقادیر مربوط به متغیرهای جمع‌آوری شده در پرونده بعضی از بیماران موجود نبود. به همین منظور از داده‌های جمع‌آوری شده از پرونده بیماران میانگین‌گیری شد و نتیجه حاصل از هر متغیر جایگزین مقادیر خالی آن متغیر گردید.

در این مقاله از ۵ روش داده کاوی تحت عنوان VFI [۲۵]، Regression [۲۶]، HyperPipes [۲۷]، Functional trees with logistic regression [۲۸،۲۹] و Meta MutiClass [۳۰] استفاده شده است.

این روش‌ها جزء روش‌های جدید، متداول و معروف در حیطه داده کاوی بوده و طیف وسیعی از مدل‌های داده کاوی را در بر می‌گیرند و منجر به تنوع در روش‌های داده کاوی می‌شوند.

داده‌های بیماران با استفاده از نرم‌افزارهای Matlab و Weka 3.6.2 مورد آنالیز قرار گرفتند. دسته بندی (Classification) داده‌ها با در نظر گرفتن فیلد نوع سرطان از نظر اولیه و ثانویه به عنوان برچسب کلاس، صورت پذیرفت. برای پیش بینی و تشخیص سرطان کبد از الگوریتم و برای ارزیابی مدل ایجاد شده از روش K-Fold که در آن K=۱۰ می‌باشد، استفاده شد. در این پژوهش از اعتبارسنجی ضربدری از نوع K– لایه استفاده شد. در این روش داده‌ها به K زیر مجموعه تقسیم می‌شوند و از این زیر مجموعه‌ها هر بار یکی برای اعتبار سنجی و K-۱ تای دیگر برای آموزش سیستم مورد استفاده قرار می‌گیرند. در پایان میانگین نتایج K بار اعتبار سنجی به عنوان تخمین نهایی حاصل می‌گردد [۱۲].

جدول ۱: ریسک فاکتور های مؤثر در سرطان کبد

ردیف	عنوان متغیر	واحد اندازه گیری	میانگین	missing
۱	سن	سال	۵۸	٪۰
۲	جنس	مرد	٪۶۸	٪۰
		زن	٪۳۲	٪۰
۳	فشارخون	mmHg	۱۳۰/۷۰	٪۶
۴	WBC	u/l	۱۱/۵	٪۱
۵	RBC	Mg/dl	۶/۸	٪۱
۶	HGB	Mg/dl	۱۱/۵	٪۱
۷	HCT	second	۳۶/۹	٪۱
۸	ALT	grams/cell	۶۶/۳	٪۳
۹	AST	grams/cell	۷۵/۷	٪۳
۱۰	PLT	liters/cell	۲۲۳/۹	٪۱
۱۱	AFL	grams/cell	این متغیر به علت هزینه بالا در ایران اندازه گیری نمی‌شود	---
۱۲	α/AT	grams/cell	این متغیر به علت هزینه بالا در ایران اندازه گیری نمی‌شود	---
۱۳	هپاتیت B	دارد	ندارد	٪۱
		ندارد		
۱۴	هپاتیتC	دارد	ندارد	٪۲
		ندارد		
۱۵	دیابت	دارد	ندارد	٪۱
		ندارد		
۱۶	سیگار	مصرف می‌کند	مصرف نمی‌کند	٪۱۰
		مصرف نمی‌کند		

[Downloaded from jhbmi.ir on 2026-11-17]

جدول ۱: ریسک فاکتورهای مؤثر در سرطان کبد (ادامه)

۱۷	الکل	مصرف می‌کند مصرف نمی‌کند	مصرف نمی‌کند	٪۱۱
۱۸	سیروز کبدی	دارد ندارد	ندارد	٪۲
۱۹	آفلاتوکسین	دارد ندارد	دارد	٪۱
۲۰	آهن	نرمال غیرنرمال	نرمال	٪۳
۲۱	تری گلیسرید	نرمال غیرنرمال	نرمال	٪۶
۲۲	کلسترول	نرمال غیرنرمال	نرمال	٪۹

نتایج

در پیش پردازش انجام شده داده‌های گم شده (Missing) کنترل و جایگزین شد، به این صورت که گم شدگی در داده های گسسته با بیشترین تکرار جای‌گذاری گردید و برای داده های پیوسته از میانگین داده‌ها در ستون مربوطه استفاده گردید. این رهیافت، فرآیند متداولی در داده کاوی به منظور بازشناسی الگو (Pattern Recognition) در مقالات علمی می‌باشد [۳۱].

با استفاده از مدل دسته بندی توسط نرم‌افزار Weka 3.6.2 ماتریس برخورد تشکیل شد که با استفاده از آن ویژگی و حساسیت آن مدل محاسبه گردید.

در این مقاله ۵ مدل ماشین یادگیری از نظر دقت، حساسیت، ویژگی و سطح زیر نمودار Roc با یکدیگر مقایسه شدند. نتایج حاصل از آن در جدول ۲ نشان داده شده است.

مدل VFI Classifier بیشترین میزان دقت دسته‌بندی، ویژگی و سطح زیر منحنی Roc را به خود اختصاص داده است.

جدول۲: مقایسه روش‌های داده کاوی از دیدگاه دقت، ویژگی، حساسیت و سطح زیر منحنی Roc

ردیف	مدل	دقت	ویژگی	حساسیت	سطح زیر منحنی Roc
۱	VFI Classifier	۷۱/۲۹	۰/۴۹	۰/۵۰	۶۳/۳۱
۲	Regression Classifier	۶۷/۶۵	۰/۴۸	۰/۵۱	۶۲/۰۸
۳	HyperPipes Classifier	۶۳/۸۱	۰/۴۶	۰/۵۲	۵۸/۹
۴	Functional trees with logistic regression	۶۳/۳۹	۰/۴۱	۰/۵۵	۵۸/۷
۵	Meta MutiClass Classifier	۶۲/۱۶	۰/۴۲	۰/۵۴	۵۵/۴۲

در این پژوهش مدل VFI Classifier دارای بیشترین دقت می‌باشد. مدل VFI Classifier یک الگوریتم یادگیری طبقه‌بندی است. این مدل ویژگی‌های هر یک از کلاس‌های طبقه‌بندی را در نظر می‌گیرد و با ویژگی‌های نمونه جدید مقایسه می‌کند. سپس بر اساس آن تشخیص می‌دهد که نمونه مورد نظر در کدام دسته قرار می‌گیرد [۳۲].

سرطان یکی از شایع‌ترین بیماری‌های جهان امروز محسوب می‌شود. سرطان اولیه کبد (HCC) پنجمین سرطان شایع در دنیا و سومین عامل مرگ و میر در جهان می‌باشد [۳۳]. درصد قابل توجهی از سرطان‌های کبد در دنیا مربوط به سرطان کبد اولیه می‌باشد [۳۴،۳۵]. بروز HCC در آسیا معادل ۵۰۰ در صد هزار مورد در سال است در حالی که بروز آن در جوامع غربی کمتر از این میزان می‌باشد [۳۵،۳۶]. سرطان کبد پس از بروز به سرعت پیشرفت کرده و در صورت عدم تشخیص به موقع و

[Downloaded from jhbmi.ir on 2026-11-17]

اقدامات درمانی در زمان مناسب متأسفانه بیمار پس از ۳ تا ۶ ماه فوت می‌کند [۳۵،۳۷].

با وجود این که امروزه در حوزه علوم پزشکی پیشرفت‌های زیادی صورت گرفته است، اما پیش‌بینی و تشخیص زود هنگام سرطان کبد هنوز هم به عنوان یکی از مشکلات اصلی پیش روی متخصصین گوارش می‌باشد. سیستم‌های تصمیم یار بسیاری در این زمینه وجود دارد که پزشک را در تشخیص زود هنگام بیماری یاری می‌رسانند [۱۲]. داده‌کاوی از روش‌هایی است که در پیش‌بینی و تشخیص سرطان‌ها مورد استفاده قرار می‌گیرد [۱۳].

در این پژوهش همان طور که در روش اجرای پژوهش توضیح داده شد از ۵ روش داده‌کاوی VFI، Regression، HyperPipes، Functional trees with logistic regression و Meta MutiClass استفاده شده است. اطلاعات جمع آوری شده از پرونده بیماران مبتلا به سرطان کبد با استفاده از نرم‌افزارهای Matlab و Weka 3.6.2 مورد آنالیز قرار گرفتند. در پژوهش صورت گرفته مدل VFI Classifier که دارای بیشترین دقت، ویژگی و سطح زیر منحنی Roc می باشد به عنوان بهترین مدل پیش‌بینی و تشخیص مبتنی بر داده کاوی برای پیش‌بینی و تشخیص سرطان کبد در مراحل اولیه شناسایی شد.

در نهایت می‌توان گفت که به منظور پیش‌بینی و تشخیص زود هنگام سرطان کبد استفاده از الگوریتم‌های داده کاوی امری ضروری است [۳۸]. در صورت ایجاد مدلی به منظور پیش‌بینی و تشخیص زود هنگام سرطان با دقت، ویژگی و حساسیت بالا می‌توان به پزشکان در تشخیص به موقع و انجام اقدامات درمانی مناسب در مراحل اولیه کمک کرد و بدین صورت مانع از پیشرفت سرطان شد. همچنین می‌توان عمر بیمار و کیفیت زندگی او را افزایش داد. به همین منظور تصمیم به ایجاد مدلی گرفتیم که چنین قابلیت‌هایی را داشته باشد.

از جمله نقاط قوت این مطالعه می‌توان به موارد ۱ تا ۳ اشاره کرد:

۱- در کشور ایران تاکنون در زمینه سرطان کبد مطالعات زیادی صورت نگرفته است.

۲- این مطالعه توانسته است سرطان کبد را با دقت قابل قبولی در مراحل اولیه تشخیص دهد.

۳- مدل داده کاوی VFI Classifier توانست علاوه بر تشخیص سرطان کبد در مراحل اولیه، به پیش‌بینی آن نیز بپردازد.

در این پژوهش با محدودیت‌هایی اعم از عدم اندازه گیری میزان AFL و α/AT به علت هزینه بالا در ایران، وجود رکوردهای خالی از مقادیر متغیرهای اندازه گیری شده بیماران در پرونده و گم شدگی برگه‌های آزمایش تعدادی از بیماران مواجه بودیم. در صورتی که این مقادیر از دست رفته در پرونده بیمار موجود بودند می‌توان پیش‌بینی کرد دقت مدل‌های مورد بررسی به میزان بسیار زیادی بهبود می‌یافت.

با توجه به قابلیت‌های داده‌کاوی در پیش‌بینی و تشخیص بیماری‌ها می‌توان به این موارد اشاره کرد:

- داده کاوی از روش‌های مؤثر در حوزه‌های تشخیص و درمان پزشکی می‌باشد.
- خاصیت ذاتی داده کاوی الگوریتم‌های آن را به ابزاری مناسب برای پیش‌بینی و تشخیص سرطان کبد تبدیل کرده است.
- الگوریتم‌های داده کاوی همچنین می‌توانند در تمامی حوزه‌های پزشکی به خصوص در پیش‌بینی و تشخیص سرطان ها مورد استفاده قرار گیرند.
- با استفاده از داده کاوی، حجم زیادی از داده‌ها را می‌توان به سرعت مورد آنالیز قرار داد در حالی که امکان آنالیز آن توسط پزشک به تنهایی وجود ندارد.

در صورت طراحی صحیح این سیستم پزشکان می‌توانند سرطان کبد را پیش‌بینی و یا در مراحل اولیه آن را تشخیص داده و اقدامات درمانی مؤثر را برای بیمار شروع کنند. همچنین با تشخیص زود هنگام سرطان کبد، می‌توان بقای عمر بیمار را افزایش داد و کیفیت زندگی وی را بهبود بخشید.

در پایان به کادر پزشکی پیشنهاد می‌شود تمامی داده‌های مربوط به بیماران را در پرونده بیمار ثبت نموده و یک نسخه از برگه آزمایش بیمار را در پرونده بیمار قرار دهند.

به پژوهشگران علاقه‌مند در این زمینه پیشنهاد می‌شود به پژوهش در مورد کاربرد داده کاوی در زمینه سایر بیماری‌ها بپردازند. همچنین می‌توان داده‌های سرطان کبد را در سایر شهر ها جمع آوری کرد.

تشکر و قدردانی

از تمامی کادر پزشکی و پرسنل بخش مدارک پزشکی بیمارستان افضلی پور تشکر و قدردانی به عمل می‌آید.

[Downloaded from jhbmi.ir on 2026-11-17]

References

1. Nobahar M, Vafai A. General Medicine. 2th ed. Tehran: Boshra; 2011

2. Jamali R, Jamali A. Non-alcoholic fatty liver disease Feyz. 2010; 14(2):169-81. Persian.

3. Fazelipour S, Toutian Z, Davoudi P. The effect of the heroin on histological structure of liver in mouse (Balb/C). Medical Sciences Journal of Islamic Azad University Tehran Medical Branch. 2006; 16(2): 91-4. Persian.

4. Shamsi-Baghbanan H, Sharifiyan A, Kamali H, Saghebi R, Ghobanifar Z, Minaei M. Assessment of phlebotomy in liver diseases in Iranian traditional medicine. Journal of Islamic and Iranian Traditional Medicine. 2013; 3 (4) :461-76. Persian.

5. Makhdoumi P. Evaluation of anti carcinogenicity effects of new analogues of selective COX II inhibitor in human liver carcinoma cell line; 2012 Sep 4-6; Mashhad: Mashhad University of Medical Sciences; 2011.

6. Goodarzi Z, Jazayeri SM. Hepatitis B virus pres variant and hepatocellular carcinoma. Hepatitis Monthly. 2008;8(2):129-33.

7. Bosch FX, Ribes J, Cleries R, Diaz M. Epidemiology of hepatocellular carcinoma. Clin Liver Dis. 2005;9(2):191-211 .

8. Bosch FX, Ribes J, Diaz M, Cleries R. Primary liver Cancer: worldwide incidence and trends. Gastroenterology. 2004;127(5): 5-16.

9. Montazeri M, Bahaadinbeigy K, Rahnama Z, Montazeri M. Comparison of the accuracy of digital image-based and patient visit-based diagnoses in an Iranian dermatology clinic. J Basic Appl Sci Res. 2013; 3(11):28-33.

10. Montazeri M, Bahrololoum A, Nezamabadi-pour H, Soleymani Baghshah M, Montazeri M. Cooperating of local searches based hyperheuristic approach for solving traveling salesman problem. 3th International Conference on Evolutionary Computation Theory and Applications; 2011 Oct 24-26; Paris: Sci TePress; 2011.

11. Bahrampour A, Montazeri M. Prediction of breast cancer mortality by hidden markov model [dissertation]. Kerman: Kerman University of Medical Sciences, Kerman; 2015.

12. Montazeri M, Montazeri M. Machine learning models for predicting the diagnosis of liver disease. Koomesh. 2014; 16(1): 53-9. Persian.

13. Montazeri M, Nezamabadipour H. Automatic extraction of eye field from a gray intensity image using intensity filtering and hybrid projection function. Communications, Computing and Control Applications (CCCA), 2011 International Conference on; 2011 Mar 3-5; Hammamet: IEEE; 2011. p. 1-5.

14. Montazeri M, NezamabadiPour H, Montazeri M. automatically eye detection with different gray intensity image conditions. Computer Technology and Application. 2012; 3: 525-32.

15. Montazeri M, Montazeri M, Montazeri M, Beigzadeh A. Machine learning models in breast cancer survival prediction. Technol Health Care. 2015.

16. Lin RH. An intelligent model for liver disease diagnosis. Artif Intell Med. 2009;47(1):53-62.

17. Montazeri, M, Montazeri M, Beygzadeh A, Zahedi MJ. Identifying efficient clinical parameters in diagnose of liver disease. HealthMED. 2014; 1115.

18. Montazeri M, Naji HR, Faraahi A. A novel memetic feature selection algorithm. Information and Knowledge Technology (IKT). 5th Conference on; 2013 May 28-30; Shiraz: IEEE; 2013.

19. Montazeri M, Naji HR, Montazeri M. Memetic feature selection algorithm based on efficient filter local search. J Basic Appl Sci Res. 2013; 3(10): 126-33.

20. Avci E, Turkoglu I. An intelligent diagnosis system based on principle component analysis and ANFIS for the heartvalve diseases. Expert Systems with Applications. 2009;36(2):2873-8.

21. Castanho MJ, Hernandes F, De Ré AM, Rautenberg S, Billis A. Fuzzy expert system for predicting pathological stage of prostate cancer. Expert Systems with Applications. 2013;40(2):466-7.

22. Pinheiro F, Kuo MH, Thomo A, Barnett J. Mining Association Rules in the BCCA Liver Cancer Data Set. Stud Health Technol Inform. 2015;208:308-13.

23. Jamarani SM, Moradi MH, Behnam H, Rad GR. Intelligent system for breast cancer prognosis using multiwavelet packets and neural network. International Enformatika Conference, IEC'05; 2005 Aug 26-28; Prague, Czech Republic; 2005.

24. Alipoor M, Haddadnia J. An accurate intelligent breast cancer diagnosis system. Iranian Journal of Breast Disease. 2009;2(2):33-40. Persian.

25. Demiröz G, Güvenir HA. Classification by voting feature intervals. Machine Learning: Springer Berlin Heidelberg; 1997.

26. Frank E, Wang Y, Inglis S, Holmes G, Witten IH. Using model trees for classification. Machine Learning. 1998;32(1):63-76.

27. Webb S, Caverlee J, Pu C. Predicting web spam with HTTP session information. Proceedings of the 17th ACM conference on Information and knowledge management; 2008 Oct 26-30; Napa Valley, California, USA: ACM; 2008.

28. Gama J. Functional trees. Machine Learning. 2004;55(3):219-50.

29. Landwehr N, Hall M, Frank E. Logistic model trees. Machine Learning. 2005;59(1-2):161-205.

30. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. The WEKA data mining software: an update. ACM SIGKDD Explorations Newsletter. 2009;11(1):10-8.

31. Montazeri M, Bahrololoum A, Enhesari A, Hyper-heuristic algorithm for finding efficient features in diagnose of lung cancer disease. J Basic Appl Sci Res. 2013; 3(10): 134-40.

32. Guvenir HA, Demiroz G, Ilter N. Learning differential diagnosis of erythemato-squamous diseases using voting feature intervals. Artif Intell Med. 1998;13(3):147-65.

[Downloaded from jhbmi.ir on 2026-11-17]

33. Reddy A, Dash C, Leerapun A, Mettler TA, Stadheim LM, Lazaridis KN, et al. Hypothyroidism: a possible risk factor for liver cancer in patients with no known underlying cause of liver disease. Clin Gastroenterol Hepatol. 2007;5(1):118-23.
34. McGlynn KA, London WT. Epidemiology and natural history of hepatocellular carcinoma. Best Pract Res Clin Gastroenterol. 2005;19(1):3-23.
35. Somi M, Fakhrjo A, Shakeri M, Farhang S. Skull metastasis as the first presentation of hepatocellular carcinoma. Govaresh. 2005; 10(4): 208-49.
36. Chan CH, Trost N, McKelvie P, Rophael JA, Murphy MA. Unusual case of skull metastasis from hepatocellular carcinoma. ANZ J Surg. 2004;74(8):710-3.
37. Qureshi SS, Shrikhande SV, Borges AM, Shukla PJ. Chest wall metastases from unknown primary hepatocellular carcinoma. J Postgrad Med. 2005;51(1):41-2.
38. Sabnani SV. Computer security: A machine learning approach [dissertation]. England: Department of Mathematics Royal Holloway, University of London Egham; 2008.

[Downloaded from jhbmi.ir on 2026-11-17]

Original Article

Journal of Health and Biomedical Informatics
Medical Informatics Research Center
2015; 2(3): 133-140

Futures Studies in Health: Choosing the Best Intelligent Data Mining Model to Predict and Diagnose Liver Cancer in Early Stage

Faezeh Afzali[1], Zohreh Heidari[1], Mitra Montazeri[2,3], Leila Ahmadian[4]*, Mohammad Javad Zahedi[5]

• **Received:** 17 Feb, 2015 • **Accepted:** 16 Sep, 2015

Introduction: Primary liver cancer (HCC), is the fifth most common type of cancer and the third leading cause of death in the world. Symptoms of liver cancer will progress rapidly after the onset of the disease, and unfortunately, the patients' survival rate is very low. One of the main problems for gastroenterologists is the prediction and early detection of liver cancer. Data mining techniques can be used to understand and predict cancer. The aim of this study was to identify the best model based on intelligent data mining to predict and diagnose liver cancer in an early stage.

Method: In the present article, a retrospective study was conducted on 516 cases of primary and secondary liver cancer, and 22 risk factors were examined. Data were collected from the patients' files and analyzed using 5 data mining models including VFI Classifier, Regression Classifier, Hyper Pipes Classifier, Functional trees with logistic regression, and Meta Multi Class Classifier with the highest precision (Precision). These models were compared.

Results: The precision, sensitivity, specificity, and the area under the curve of VFI Classifier model were respectively 71.29%, 49%, 50%, and 63.31%, and VFI Classifier model is the best model based on intelligent data mining to predict and diagnose liver cancer in an early stage.

Conclusion: If properly designed, data mining model VFI Classifier can predict liver cancer or detect it in an early stage.

Key words: Liver Cancer, Hepatocellular Carcinoma (HCC), Prediction and Detection, Data Mining, Futures studies on Health

•**Citation:** Afzali F, Heidari Z, Montazeri M, Ahmadian L, Zahedi MJ. Futures Studies in Health: Choosing the Best Intelligent Data Mining Model to Predict and Diagnose liver Cancer in Early Stage. Journal of Health and Biomedical Informatics 2015; 2(3): 133-140.

1. B.Sc. of Health Information Technology, Research Center for Modeling in Health, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran.
2. M.Sc. in Artificial Intelligence, Medical Informatics Research Center, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran.
3. M.Sc. in Artificial Intelligence, Computer Engineering Dept., Shahid Bahonar University, Kerman, Iran.
4. Ph.D. in Medical Informatics, Associate Professor, Medical Informatics Research Center, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran.
5. Gastroenterologist, Professor, Research Center of Gastroenterology and Hepatology, Kerman University of Medical Sciences, Kerman, Iran.

***Correspondence:** Medical Informatics Research Center, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Haft - bagh Highway

•**Tel:** 03431325406 •**Email:** ahmadianle@yahoo.com

[Downloaded from jhbmi.ir on 2026-11-17]