



مدلی برای پیش‌بینی نیاز به جراحی ارتوپدی با استفاده از تکنیک‌های داده‌کاوی

سیدسینا فاطمی رضوی

دانش آموخته کارشناسی ارشد دانشگاه آزاد اسلامی واحد علوم تحقیقات
30na.rzv2@gmail.com

سید عبدالله امین موسوی

استادیار دانشگاه آزاد اسلامی واحد علوم و تحقیقات (مسئول مکاتبات)
Dr.AminMousavi@gmail.com

تاریخ دریافت: ۹۷/۰۷/۳۰ تاریخ پذیرش: ۹۷/۱۱/۱۳

چکیده

با گسترش استفاده از رایانه در جنبه‌های مختلف زندگی افراد، حجم بسیار زیادی از داده‌ها تولید می‌شود که در بسیاری از اوقات این داده‌ها شامل اطلاعات ارزشمندی هستند. برای استخراج این اطلاعات و بهره‌بردن از آنها می‌توان از علم داده‌کاوی^۱ بهره‌برد. با استفاده از داده‌کاوی می‌توان، الگوهای پنهان موجود در داده‌ها را کشف نمود و برای پیش‌بینی موارد جدید مورد استفاده قرار داد. از جمله حوزه‌هایی که با تولید حجم انبوه داده‌ها روبرو می‌باشد، حوزه‌ی درمان است. در این پژوهش به‌طور خاص در زمینه‌ی ارتوپدی تمرکز خواهد شد. این تحقیق به دنبال این است که با استفاده از تکنولوژی و تکنیک‌های داده‌کاوی بتواند از داده‌ها موجود در دیتابیس بیمارستان به اطلاعات ارزشمندی دست یابد و از طریق آن اطلاعات بتواند احتمال شکستگی و همچنین نیازمند بودن بیمار به جراحی را پیش‌بینی کند و تصمیم‌گیری را برای پزشکان ساده‌تر و سریع‌تر کند. بدین صورت می‌توان با سرعت و دقت بالاتری نسبت به روش‌های موجود به تفکیک بیماران و ارائه‌ی خدمات به آن‌ها پرداخت. این پژوهش بر مبنای متدولوژی CRISP بنا نهاده شده است و نتایج حاصل از تحقیق بیانگر این امر است که استفاده‌ی تلفیقی از الگوریتم‌های CHAID و شبکه‌ی عصبی^۲ تقویت شده با روش تجمعی^۳ Boosting، می‌تواند دقت مطلوبی در پیش‌بینی نیاز به جراحی در بیماران ارتوپدی را ارائه دهد.

واژه‌های کلیدی: داده‌کاوی، ارتوپدی، درمان جراحی.

۱- مقدمه

عمق و گستردگی آن را نشان می‌دهد. بطور کلی ارتوپدی به دو دسته تقسیم می‌شود: ارتوپدی در زمینه آسیب‌های حادثه‌ای خارجی که با تصادفات و ضربه‌ها ایجاد می‌شود و ارتوپدی در زمینه ناهنجاری‌ها و آسیب‌های غیرحادثه‌ای که می‌تواند مادرزادی یا اکتسابی باشد. در ارتوپدی دو دسته آسیب ممکن است اتفاق بیفتد: یکی آسیب‌هایی که به استخوان‌ها می‌رسد و به آن شکستگی گفته می‌شود؛ دسته دیگر هم آسیب‌هایی هستند که به بافت‌های نرم وارد شده و به پارگی عضله و شریان‌ها منجر می‌شود که با توجه به اهمیت و حاد بودن این آسیب‌ها بیمار جهت دریافت خدمات ارتوپدی به اورژانس یا مرکز فوریت‌های پزشکی انتقال داده می‌شود.

۲- مروری بر پیشینه تحقیق

در رابطه با داده‌کاوی در عرصه پزشکی تحقیقات گسترده‌ای صورت گرفته است. جدول ۱ تحقیقات مرتبط با کاهش ابعاد و جدول ۲ تحقیقات مرتبط با روش‌های دسته‌بندی^۴ و جدول ۳ تحقیقات مرتبط با روش‌های خوشه‌بندی^۵ و جدول ۴ روش‌های مرتبط با روش‌های قواعد انجمنی^۶ را نشان می‌دهند.

امروزه با گسترش استفاده از رایانه در حوزه‌های مختلف به خصوص حوزه سلامت، شاهد تولید حجم انبوهی از داده‌ها می‌باشیم که در بسیاری از اوقات شامل اطلاعات ارزشمندی هستند که می‌توانند راهگشای بسیاری از مشکلات فعلی در حوزه درمان باشند. برای استخراج اطلاعات مفید و موثر از داده‌ها و کشف الگوهای موجود در آن‌ها نیاز به فرآیندهایی تحت عنوان داده‌کاوی می‌باشیم. داده‌کاوی یکی از تکنیک‌های پرکاربرد می‌باشد که پردازش خود را روی داده‌های انبوه انجام می‌دهد. همان‌طور که گفته شد با توجه به فراوانی داده‌های پزشکی و بیماری‌ها، این موضوع می‌تواند بستر مناسبی برای به‌کارگیری تکنیک‌های مختلف داده‌کاوی باشد. در این پژوهش سعی می‌شود با بهره‌گیری از روش‌های داده‌کاوی، داده‌های بخش ارتوپدی را مورد بررسی قرار داد.

یکی از خدمات تخصصی که در بخش اورژانس به بیماران ارائه می‌شود خدمات مرتبط با ارتوپدی است. ارتوپدی به‌طور عام شامل تمام معضلات، مشکلات، حوادث و اتفاقاتی است که در اثر حادثه یا به‌صورت مادرزادی یا در اثر تغییر شکل‌های تدریجی اندام‌ها اعم از فوقانی، تحتانی، ستون فقرات و لگن به وجود می‌آید و همین

جدول ۱- تحقیقات مرتبط با کاهش ابعاد

ردیف	عنوان پژوهش	نام پژوهشگر
۱	استخراج دانش از داده‌های بیماران دیابتی با استفاده از درخت تصمیم C5.0	عامری و همکاران (۱۳۹۲)
۲	تشخیص بیماری هپاتیت با ترکیب روش‌های داده‌کاوی	کاظمی و همکاران (۱۳۹۲)
۳	شناسایی مدل‌ها، پیش‌بینی بیماری‌های عروق کرونر با استفاده از شبکه‌های عصبی و انتخاب متغیر بر اساس طبقه‌بندی و درخت رگرسیون	محمودی و همکاران (۱۳۹۲)

جدول ۲- تحقیقات مرتبط با روش‌های دسته‌بندی

ردیف	عنوان پژوهش	نام پژوهشگر
۱	پیش‌بینی داده‌کاوی برای تشخیص پزشکی: بررسی کلی از پیش‌بینی بیماری‌های قلبی	(Soni et al, 2011)
۲	استفاده از یک شبکه عصبی مصنوعی در پیش‌بینی بیماری قلبی	(Khalid et al, 2013)
۳	تشخیص بیماری سرطان ریه با استفاده از منطق فازی نوری	(Malathi et al, 2013)
۴	بررسی تجربی کاربرد تکنیک‌های داده‌کاوی در مراقبت‌های بهداشتی	(Kaur et al, 2006)
۵	داده‌کاوی داده‌های پزشکی و بیماری‌ها	(موسوی، ۱۳۹۳)
۶	تجزیه و تحلیل تطبیقی الگوریتم‌های طبقه‌بندی تصمیم‌گیری درخت	(Priyama et al, 2013)
۷	بررسی مقایسه‌ای روش‌های طبقه‌بندی داده‌کاوی در پیش‌بینی بیماری قلبی عروقی	(Kumari et al, 2011)

ردیف	عنوان پژوهش	نام پژوهشگر
۸	مقایسه هفت الگوریتم برای پیش بینی بقای سرطان پستان	(Endo et al, 2008)
۹	مقایسه و تجزیه و تحلیل الگوریتم های طبقه بندی درخت تصمیم گیری	(Anyanwu, et al, 2009)

جدول ۳- تحقیقات مرتبط با روش های خوشه بندی

ردیف	عنوان پژوهش	نام پژوهشگر
۱	توسعه انبار داده های پزشکی بالینی مرسوم چینی برای کشف دانش پزشکی و حمایت از تصمیم گیری.	(Zhou et al, 2010)
۲	ارائه راه حلی برای تشخیص بیماری به کمک تکنیک های داده کاوی	کلاهی و رافع (۱۳۹۲)
۳	تشخیص تومورهای مغزی با استفاده از الگوریتم خوشه بندی فازی و هوش ازدحامی	مهرآفرید و اکبرپورسکه (۱۳۹۷)
۴	بهبود قطعه بندی تصاویر تومور مغزی با استفاده از ترکیب خوشه بندی فازی و الگوریتم PSO	قوبدل و همکاران (۱۳۹۷)
۵	انتخاب خوشه بندی ترکیبی بر پایه یک اندازه گیری جدید بر روی داده های پزشکی	باقریان وهمکاران (۱۳۹۶)

جدول ۴- تحقیقات مرتبط با روش های کشف قواعد انجمنی

ردیف	عنوان پژوهش	نام پژوهشگر
۱	کشف قوانین انجمنی در داده های پزشکی	(DODDI et al, 2001)
۲	استفاده از الگوریتم بیز ساده و قوانین انجمنی جهت کشف روابط پنهان بین پارامترهای خطرزای بیماری سل	فراهانی و ربیعی (۱۳۹۶)
۳	قوانین انجمنی داده کاوی در سیستم پیش بینی بیماری قلبی با استفاده از الگوریتم	جواهریان (۱۳۹۵)

۳- روش پژوهش

در این تحقیق مدل پیشنهادی بر اساس CRISP ارائه می شود که شامل ۵ فاز است با توجه به متدولوژی CRISP ابتدا به درک هدف انجام کار و شناخت از بیمارستان و شناخت مفاهیم ارتوپدی می پردازیم. فاز دوم فاز شناخت داده ها و آماده سازی آنهاست که در این مرحله به بررسی رابطه بین فیلدها، شناسایی داده های خاص و حذف رکوردهای ناقص و دسته بندی اطلاعات در گروه های مفید می پردازیم. مرحله بعد مرحله مدل سازی است، روش های داده کاوی بسیاری برای مدل سازی وجود دارد. در این فاز با استفاده از تکنیک های مختلف داده کاوی به پیدا کردن مدل و الگوی بهینه می پردازیم. مدل سازی را با استفاده از نرم افزار IBM SPSS Modeler انجام می دهیم. پس از مدلسازی به ارزیابی نتایج حاصل از مدلسازی می پردازیم. نتایج ارزیابی باعث بهبود مدل می شود و مدل را قابل استفاده می نماید. مجموعه داده ی مورد استفاده در این پژوهش اطلاعات حدود ۳۱۰ بیمار می باشد و از بیمارستان اختر

جمع آوری شده است که متشکل از هفت ویژگی می باشد. شش مورد از آن ها، ویژگی های بیومکانیکی مربوط با بیماران ارتوپدی می باشد. این ویژگی های بیومکانیکی با توجه به شکل و جهت گیری لگن و ستون فقرات کمری مشخص می گردند. این ویژگی به صورت زیر می باشند:

- ویژگی برخورد لگنی^۷ : به عنوان زاویه ی بین خط عمود بر صفحه ی ساکرال^۸ (نام یکی از استخوان های تشکیل دهنده لگن خاصره است) در نقطه ی میانی خود و خط ارتباط دهنده ی این نقطه به محور اسکلت ران می باشد.
- شیب لگن^۹ : شیب لگن، جهت لگن در ارتباط با استخوان ران و بقیه بدن است.
- زاویه لوردوز کمری^{۱۰} : لوردوز یا کاوشتی، اصطلاحی پزشکی است که انحنای داخلی در بخشی از ستون مهره های بدن را بیان می کند. به بیان دیگر، کاوشتی، افزایش نابهنجار گودی کمر در ستون مهره ها می باشد.

۴-۱- نمای کلی داده‌ها

در این مرحله، مواردی همچون شناخت نقش، نوع و جزییات کاربردی داده مورد بررسی قرار می‌گیرد. شکل ۱ نمایی از داده‌های مورد استفاده را نشان می‌دهد. داده‌های مورد استفاده در این پژوهش در قالب دو مجموعه داده ارائه شده است. در یکی از مجموعه داده‌ها که به صورت دو کلاسه می‌باشد، صرفاً افراد عادی (با برچسب NO) و غیرعادی (با برچسب AB) از هم تفکیک شده‌اند. در مجموعه داده دوم که به صورت سه کلاسه می‌باشد، علاوه بر تفکیک افراد عادی از غیرعادی، موارد غیرعادی نیز بر اساس نوع بیماری به دو گروه فتق دیسک (با برچسب DH) یا لغزش مهره (با برچسب SL) تقسیم شده‌اند.

نحوه‌ی توزیع کلاس‌های عادی و غیرعادی در شکل ۲ نشان داده شده است. همانطور که مشاهده می‌شود، بیشتر افراد ثبت شده در مجموعه داده‌ی مورد بررسی دارای مشکلات فتق دیسک یا لغزش مهره می‌باشند.

شکل ۳، نحوه‌ی توزیع کلاس‌های عادی و غیرعادی با در نظر گرفتن هر دو حالت کلاس غیرعادی یعنی فتق دیسک یا لغزش مهره را نشان می‌دهد.

- شیب ساکرالی^{۱۱} : به عنوان زاویه‌ی بین صفحه‌ی ساکرال و یک خط افقی تعریف شده است.
- شعاع لگن^{۱۲}
- درجه اسپوندیلولیزتازی^{۱۳} : اسپوندیلولیزتازی یا لغزش مهره، در پزشکی، به رانده شدن یک مهره به طرف جلو بر روی مهره زیرین، گفته می‌شود. با توجه به ویژگی‌های بیان شده، وضعیت بیمار را می‌توان به دسته‌های عادی و غیرعادی طبقه‌بندی نمود. حالت غیرعادی می‌تواند شامل موارد زیر باشد:
 - فتق دیسک^{۱۴} : عارضه‌ای است که به علت ضربه یا بلندکردن جسم سنگین یا گاهی خود به خود در ستون فقرات اتفاق می‌افتد و طی آن، آزاد شدن بخش مرکزی دیسک یا هسته دیسک از قسمت محیطی آن، باعث فشار بر روی ریشه‌های عصبی و ایجاد علائم بالینی می‌گردد.
 - اسپوندیلولیزتازیس: بیمار دارای مشکل لغزش مهره.
- با توجه به ویژگی‌های بیان شده، می‌توان با اعمال روش‌های داده‌کاوی بر روی شش ویژگی ورودی، به وضعیت عادی یا غیرعادی بیمار پی برد. در حالت غیرعادی نیز می‌توان بررسی نمود که مشکل بیمار، کدامیک از دو مورد فتق دیسک یا لغزش مهره می‌باشد.

۴- یافته‌های پژوهش

	pelvic_incidence	pelvic_tilt	lumbar_lordosis_angle	sacral_slope	pelvic_radius	degree_spondylolisthesis	class
1	63.030	22.550	39.610	40.480	98.670	-0.250	AB
2	39.060	10.060	25.020	29.000	114.410	4.560	AB
3	68.830	22.220	50.090	46.610	105.990	-3.530	AB
4	69.300	24.650	44.310	44.640	101.870	11.210	AB
5	49.710	9.650	28.320	40.060	108.170	7.920	AB
6	40.250	13.920	25.120	26.330	130.330	2.230	AB
7	53.430	15.860	37.170	37.570	120.570	5.990	AB
8	45.370	10.760	29.040	34.610	117.270	-10.680	AB
9	43.790	13.530	42.690	30.260	125.000	13.290	AB
10	36.690	5.010	41.950	31.680	84.240	0.660	AB
11	49.710	13.040	31.330	36.670	108.650	-7.830	AB
12	31.230	17.720	15.500	13.520	120.060	0.500	AB
13	48.920	19.960	40.260	28.950	119.320	8.030	AB
14	53.570	20.460	33.100	33.110	110.970	7.040	AB
15	57.300	24.190	47.000	33.110	116.810	5.770	AB
16	44.320	12.540	36.100	31.780	124.120	5.420	AB
17	63.830	20.360	54.550	43.470	112.310	-0.620	AB
18	31.280	3.140	32.560	28.130	129.010	3.620	AB
19	38.700	13.440	31.000	25.250	123.160	1.430	AB
20	41.730	12.250	30.120	29.480	116.590	-1.240	AB

شکل ۱- نمایی از داده‌های مورد استفاده

Value ▲	Proportion	%	Count
AB		67.74	210
NO		32.26	100

شکل ۲- نحوه ی توزیع کلاس های عادی و غیرعادی

Value ▲	Proportion	%	Count
DH		19.35	60
NO		32.26	100
SL		48.39	150

شکل ۳- نحوه ی توزیع کلاس های عادی و غیرعادی

۲-۴- پیش پردازش^{۱۵}

می باشد. برای استاندارد سازی مقادیر راهکارهای متعددی وجود دارد که در این پژوهش از روش نمره استاندارد^{۱۶} استفاده می شود. برای اعمال این روش باید قدم های زیر برای هر مقدار از هر ویژگی انجام شود:

- (۱) محاسبه ی میانگین و انحراف معیار ویژگی
 - (۲) کم کردن مقدار مورد نظر از میانگین
 - (۳) تقسیم کردن حاصل بدست آمده از مرحله ی ۲ بر انحراف معیار
- بخشی از نتیجه ی حاصل از محاسبه ی نمره استاندارد در شکل ۴ نشان داده شده است.

توضیحات گسترده ای پیرامون پیش پردازش و آماده سازی داده ها در منابع مرتبط با داده کاوی ارائه شده است. با بررسی مجموعه داده ی این تحقیق، می توان دریافت که بسیاری از این فعالیت ها برای این داده ها ضرورت ندارند. به عنوان مثال از آنجایی که در مجموعه داده مورد نظر، شاهد مقادیر گم شده نیستیم، بنابراین نیازی به فعالیت های مرتبط با حذف مقادیر گم شده، نمی باشد.

با توجه به اینکه همه ی ویژگی های این مجموعه داده به غیر از ویژگی کلاس، از نوع عددی می باشد، مهم ترین فعالیتی که باید صورت داد، استاندارد سازی مقادیر

class	pelvic_incidence...	pelvic_tilt...	lumbar_lordosis_angle...	sacral_slope...	pelvic_radius...	degree_spondylolisthesis...
AB	0.147	0.500	-0.664	-0.184	-1.445	-0.707
AB	-1.244	-0.748	-1.450	-1.040	-0.264	-0.579
AB	0.483	0.467	-0.099	0.272	-0.896	-0.794
AB	0.511	0.710	-0.411	0.126	-1.205	-0.402
AB	-0.626	-0.789	-1.273	-0.216	-0.732	-0.489
AB	-1.175	-0.362	-1.445	-1.238	0.932	-0.641
AB	-0.410	-0.168	-0.796	-0.401	0.199	-0.541
AB	-0.878	-0.678	-1.234	-0.622	-0.049	-0.985
AB	-0.969	-0.401	-0.498	-0.946	0.532	-0.346
AB	-1.381	-1.252	-0.538	-0.840	-2.529	-0.683
AB	-0.626	-0.450	-1.110	-0.468	-0.696	-0.909
AB	-1.698	0.018	-1.964	-2.193	0.161	-0.687
AB	-0.672	0.242	-0.629	-1.043	0.105	-0.486
AB	-0.402	0.291	-1.015	-0.733	-0.522	-0.513
AB	-0.185	0.664	-0.266	-0.733	-0.083	-0.547
AB	-0.939	-0.500	-0.853	-0.832	0.466	-0.556
AB	0.193	0.281	0.141	0.038	-0.421	-0.717
AB	-1.695	-1.439	-1.044	-1.104	0.833	-0.604
AB	-1.265	-0.410	-1.128	-1.319	0.393	-0.662
AB	-1.089	-0.529	-1.176	-1.004	-0.100	-0.733

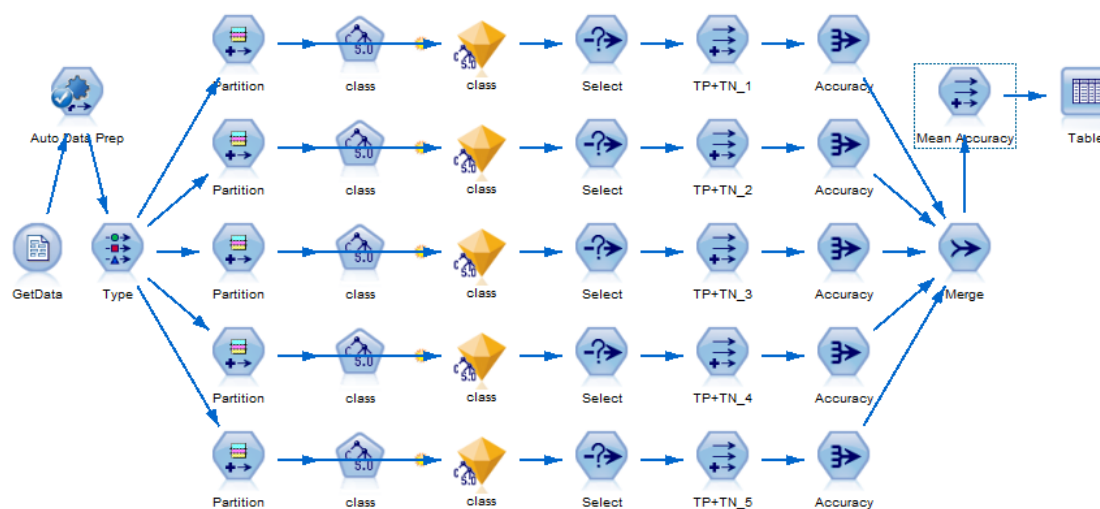
شکل ۴- محاسبه ی نمره استاندارد

۳-۴- مدل‌سازی

در این بخش پژوهش، به ساخت مدل‌های داده‌کاوی بر روی مجموعه داده مورد نظر پرداخته می‌شود. هدف از این بخش تعیین دقت روش‌های مختلف برای پیش‌بینی نیاز به جراحی بیماران می‌باشد. برای مقایسه کردن روش‌های مختلف داده‌کاوی از راهکار ارزیابی متقابل K-fold^{۱۷} استفاده خواهد شد. در راهکار ارزیابی متقابل، در طول فرآیند آموزش، داده‌های

آموزش به K بخش تقسیم می‌شوند و هر بار از یکی از بخش‌ها برای ارزیابی و از سایر بخش‌ها برای آموزش استفاده می‌شود.

با توجه به اینکه بسیاری از الگوریتم‌های محیط Spss Modeler فاقد امکان داخلی ارزیابی متقابل می‌باشند، نیاز به پیاده‌سازی دستی این روش می‌باشد. شکل ۵ پیاده‌سازی روش ارزیابی متقابل (5-Fold) بر روی روش C5 را نشان می‌دهد.



شکل ۵- پیاده‌سازی روش ارزیابی متقابل (5-Fold)

روش‌های C5 و Neural Net بر روی داده‌های دو کلاس، نتایج نسبتاً بهتری ارائه می‌دهند. میانگین ۱۰ بار اجرای روش ارزیابی متقابل بر روی داده‌های سه کلاس به عنوان نتیجه در جدول ۶ قید شده است.

برای اطمینان از دقت هر یک از روش‌ها، فرآیند تقسیم داده‌ها به داده‌های آموزش و ارزیابی به صورت تصادفی صورت می‌گیرد. میانگین ۱۰ بار اجرای روش ارزیابی متقابل بر روی داده‌های دو کلاس به عنوان نتیجه در جدول ۵ قید شده است. برای اطمینان از کیفیت روش‌ها، تمام محاسبات بر روی داده‌های ارزیابی (تست) صورت گرفته است.

جدول ۵- نتایج حاصل از اجرای ارزیابی متقابل برای حالت دو کلاس

ردیف	روش مورد استفاده	میانگین نتایج	بهترین نتیجه	بدترین نتیجه
۱	C5	۰,۸۹	۰,۹۲۳	۰,۸۶۶
۲	CHAID	۰,۸۴۹	۰,۸۷۸	۰,۸۳۸
۳	C&R Tree	۰,۸۵۶	۰,۸۷۱	۰,۸۳۳
۴	Neural Net	۰,۸۷۴	۰,۸۹۳	۰,۸۵۶
۵	SVM	۰,۸۱۴	۰,۸۴۴	۰,۷۸۸
۶	KNN	۰,۸۴۳	۰,۸۷۹	۰,۸۰۳

تاثیر ساختار شبکه‌ی عصبی بر کیفیت نتیجه‌ی بدست آمده را بررسی نمود.

جدول ۷- تاثیر روش های تجمعی (داده های دو کلاسه)

ردیف	روش مورد استفاده	میانگین نتایج	بهترین نتیجه	بدترین نتیجه
۱	C5 (Boosting)	۰,۹۲۳	۰,۹۵۲	۰,۹۰۱
۲	Neural Net(Bagging)	۰,۹۰۵	۰,۹۲۹	۰,۸۸
۳	Neural Net(Boosting)	۰,۹۶۳	۰,۹۷۹	۰,۹۴۲

جدول ۸- تاثیر روش های تجمعی (داده های سه کلاسه)

ردیف	روش مورد استفاده	میانگین نتایج	بهترین نتیجه	بدترین نتیجه
۱	C5 (Boosting)	۰,۹۵۶	۰,۹۷۷	۰,۹۲۹
۲	Neural Net(Bagging)	۰,۹۰۴	۰,۹۲۷	۰,۸۸۵
۳	Neural Net(Boosting)	۰,۹۶۷	۰,۹۸۱	۰,۹۵۱

۴-۵- بهبود ساختار روش شبکه‌ی عصبی

شبکه‌های عصبی در محیط Spss Modeler به شبکه‌های Multi Layer Perceptron (MLP) و Radial Basis Function (RBF) محدود می‌شوند. ساختار پیش فرض شبکه‌های عصبی در محیط Spss Modeler، مبتنی بر یک لایه‌ی پنهان می‌باشد که بهترین تعداد نورون برای آن توسط نرم افزار محاسبه می‌شود. همچنین امکان افزایش تعداد لایه‌های پنهان حداکثر به دو مورد می‌باشد (برای شبکه‌ی عصبی RBF، تنها می‌توان یک لایه‌ی پنهان در نظر گرفت). بنابراین، برای بررسی سایر ساختارهای شبکه‌ی عصبی، تمرکز را بر روی روش MLP، با تعداد دو لایه خواهیم گذاشت و سعی می‌شود که تعداد مطلوب نورون در هر لایه شناسایی شود. این فرآیند با تست ساختارهای مختلف شبکه‌ی عصبی صورت خواهد گرفت. جدول ۹ نتایج حاصل از بررسی ساختارهای مختلف شبکه‌ی عصبی بر روی داده‌های دو کلاسه را نشان می‌دهد. دقت شود که در تمام حالات، روش تجمعی Boosting لحاظ شده است.

جدول ۱۰ نتایج حاصل از بررسی ساختارهای مختلف شبکه‌ی عصبی بر روی داده‌های سه کلاسه را نشان می‌دهد.

جدول ۶- نتایج حاصل از اجرای ارزیابی متقابل برای

حالت سه کلاسه

ردیف	روش مورد استفاده	میانگین نتایج	بهترین نتیجه	بدترین نتیجه
۱	C5	۰,۸۷۴	۰,۹۱۳	۰,۸۵۱
۲	CHAID	۰,۸۲۳	۰,۸۵۶	۰,۸۰۱
۳	C&R Tree	۰,۸۳۵	۰,۸۷۵	۰,۷۹۴
۴	Neural Net	۰,۸۸۲	۰,۹۰۳	۰,۸۴۱
۵	SVM	۰,۸۰۱	۰,۸۴۲	۰,۷۸۱
۶	KNN	۰,۸۲۳	۰,۸۴۹	۰,۷۹۱

برای داده‌های سه کلاسه نیز روش‌های C5 و Neural Net، نتایج نسبتاً بهتری ارائه می‌دهند.

۴-۴- تاثیر روش های تجمعی بر بهبود دقت

با بهره‌گیری از روش‌های تجمعی می‌توان دست‌بندها را تقویت نمود. در محیط Spss Modeler برخی از الگوریتم‌ها، روش‌های تجمعی Boosting و Bagging لحاظ شده است. در این بخش از این روش‌ها بر روی مدل‌های C5 و شبکه‌ی عصبی که در آزمایش قبلی نتایج مناسب تری نسبت به سایر راهکارها داشتند، استفاده می‌شود.

جدول ۷ نتایج حاصل از اعمال روش‌های تجمعی را بر روی مدل‌های C5 و شبکه‌ی عصبی با در نظر گرفتن داده‌های دو کلاسه، نشان می‌دهد. همان‌طور که ملاحظه می‌شود، روش شبکه‌ی عصبی به همراه Boosting نتیجه‌ی مطلوب تری را ارائه می‌دهد. دقت شود که برای روش C5 در محیط نرم افزار Spss Modeler، پیاده‌سازی روش Bagging لحاظ نشده است.

جدول ۸ نتایج حاصل از اعمال روش‌های تجمعی را بر روی مدل‌های C5 و شبکه‌ی عصبی با در نظر گرفتن داده‌های سه کلاسه، نشان می‌دهد. همان‌طور که ملاحظه می‌شود مشابه جدول ۷، روش شبکه‌ی عصبی به همراه Boosting نتیجه‌ی مطلوب تری را ارائه می‌دهد.

با توجه به نتایج جداول ۷ و ۸، می‌توان مشاهده کرد که روش شبکه‌ی عصبی بهبود داده شده با Boosting، نتایج مطلوب تری ارائه می‌دهد. در ادامه سعی می‌شود که

نتایج حاصل از تلفیق برخی از روش‌ها بر روی شبکه عصبی MLP با دو لایه ی پنهان که لایه ی اول ۵ نورون و لایه ی دوم ۱۰ نورون را دارا می‌باشد، در جدول ۱۱ قید شده است. نتایج این جدول، مربوط به داده‌های دو کلاسه است. همان‌طور که مشاهده می‌شود، تلفیق روش‌های شبکه عصبی و روش CHAID، بهترین نتیجه را ارائه می‌دهد. دقت شود که برای تمام روش‌های قید شده در جدول ۱۱ و جدول ۱۲، از روش تجمعی Boosting بهره برده شده است.

جدول ۱۱- نتایج روش‌های ترکیبی بر روی داده‌های

دو کلاسه

ردیف	مدل‌های استفاده شده	میانگین نتایج	بهترین نتیجه	بدترین نتیجه
۱	NeuralNet+ C5	۰,۹۸۵	۰,۹۹۶	۰,۹۷۸
۲	NeuralNet+ CHAID	۰,۹۹۷	۱,۰۰	۰,۹۹۳
۳	NeuralNet+ C&R Tree	۰,۹۸۶	۱,۰۰	۰,۹۴۲

جدول ۱۲، مشابه جدول ۱۱ طراحی شده است، با این تفاوت که از داده‌های سه کلاسه برای آموزش و ارزیابی روش استفاده شده است. در این جدول نیز همان‌طور که مشاهده می‌شود، تلفیق روش‌های شبکه عصبی و روش CHAID، بهترین نتیجه را ارائه می‌دهد.

جدول ۱۲- نتایج روش‌های ترکیبی بر روی داده‌های سه

کلاسه

ردیف	مدل‌های استفاده شده	میانگین نتایج	بهترین نتیجه	بدترین نتیجه
۱	NeuralNet+ C5	۰,۹۹	۰,۹۹۷	۰,۹۸۳
۲	NeuralNet+ CHAID	۰,۹۹۸	۱,۰۰	۰,۹۹۶
۳	NeuralNet+ C&R Tree	۰,۹۹۵	۰,۹۹۷	۰,۹۹۳

جمع بندی

همان‌طور که بیان شد، روش مبتنی بر شبکه عصبی MLP با دو لایه ی پنهان که لایه ی اول ۵ نورون و لایه ی دوم ۱۰ نورون را دارا می‌باشد، که به کمک روش تجمعی Boosting تقویت شده، از سایر روش‌ها، عملکرد بهتری داشته است. تلفیق روش مذکور با روش درخت تصمیم CHAID، بهترین راه حل تلفیقی ارائه

همان‌طور که در جداول ۹ و ۱۰ مشاهده می‌شود، روش شبکه عصبی متشکل از دو لایه پنهان که در لایه ی اول ۵ نورون و در لایه دوم ۱۰ نورون قرار داشته باشد، میانگین نتایج بهتری ارائه می‌دهد.

جدول ۹- نتایج حاصل از بررسی ساختارهای مختلف

شبکه عصبی بر روی داده‌های دو کلاسه

ردیف	ساختار شبکه	میانگین نتایج	بهترین نتیجه	بدترین نتیجه
۱	MLP (Default Structure)	۰,۹۶۳	۰,۹۷۹	۰,۹۴۲
۲	RBF(Default Structure)	۰,۸۵۵	۰,۸۸۳	۰,۸۱۹
۳	MLP(1N, 1N)	۰,۸۵۶	۰,۸۸۴	۰,۸۳۶
۴	MLP(5N, 10N)	۰,۹۶۹	۰,۹۸۴	۰,۹۵۲
۵	MLP(10N, 10N)	۰,۹۶۲	۰,۹۸۳	۰,۹۵۶
۶	MLP(10N, 5N)	۰,۹۶۳	۰,۹۷۸	۰,۹۵۷

جدول ۱۰- نتایج حاصل از بررسی ساختارهای مختلف

شبکه عصبی بر روی داده‌های سه کلاسه

ردیف	ساختار شبکه	میانگین نتایج	بهترین نتیجه	بدترین نتیجه
۱	MLP (Default Structure)	۰,۹۶۷	۰,۹۸۱	۰,۹۵۱
۲	RBF(Default Structure)	۰,۸۴۸	۰,۸۶۹	۰,۸۳
۳	MLP(1N, 1N)	۰,۸۸۵	۰,۹۱۳	۰,۸۵۱
۴	MLP(5N, 10N)	۰,۹۷۱	۰,۹۸۷	۰,۹۵۴
۵	MLP(10N, 10N)	۰,۹۶۹	۰,۹۸	۰,۹۶
۶	MLP(10N, 5N)	۰,۹۶۶	۰,۹۸۴	۰,۹۳۶

۴-۶- تلفیق مدل‌های مختلف

برای بهبود عملکرد روش‌های داده‌کاوی، می‌توان اقدام به تلفیق این روش‌ها نمود. در این پژوهش برای تلفیق روش‌های مختلف از ایده ی Stacking بهره برده شده است. در این ایده، ابتدا یکی از روش‌ها بر روی داده‌ها اعمال می‌شود و نتایج حاصل از اجرای روش، به مجموعه داده ی اصلی اضافه می‌شود. به‌طور معمول این نتایج شامل برچسب پیش‌بینی شده توسط روش می‌باشد. در ادامه، مجموعه داده ی جدید، تحویل روش اصلی می‌شود تا هم‌زمان از داده‌های اصلی و همین‌طور برچسب پیش‌بینی شده توسط روش اول، بهره‌برد.

روابط پنهان بین پارامترهای خطرهای بیماری سل، دومین کنفرانس بین‌المللی پژوهش‌های دانش بنیان در مهندسی کامپیوتر و فناوری اطلاعات، تهران، دانشگاه مجلسی

(۲) باقریان، زهره، اکبری، ابراهیم، مومنی، همایون (۱۳۹۶). انتخاب خوشه بندی ترکیبی بر پایه یک اندازه گیری جدید بر روی داده های پزشکی، دومین کنفرانس بین‌المللی پژوهش‌های دانش بنیان در مهندسی کامپیوتر و فناوری اطلاعات، تهران، دانشگاه مجلسی

(۳) جواهریان، محمد میثاق، (۱۳۹۵). قوانین انجمنی داده کاوی در سیستم پیش بینی بیماری قلبی با استفاده از الگوریتم، دومین کنفرانس بین‌المللی یافته های نوین پژوهشی در مهندسی برق و علوم کامپیوتر، رامسر، موسسه آموزش عالی غیرانتفاعی کسری رامسر

(۴) عامری حکیمه، علیزاده، سمیه، بزرگری، اکبر (۱۳۹۲). استخراج دانش از داده های بیماران دیابتی با استفاده از روش درخت تصمیم C5.0. مجله مدیریت سلامت (۵۳) ۱۶

(۵) قویدل، سهیلا، جعفری، احسان، پدیداران مقدم، فرهنگ، (۱۳۹۷). بهبود قطعه بندی تصاویر تومور مغزی با استفاده از ترکیب خوشه بندی فازی و الگوریتم PSO، چهارمین کنفرانس ملی علوم و مهندسی کامپیوتر و فناوری اطلاعات، بابل، موسسه علمی تحقیقاتی کومه علم آوران دانش

(۶) موسوی، فاطمه (۱۳۹۳). "داده کاوی داده های پزشکی و بیماری ها"، پایان نامه کارشناسی ارشد. رشته مهندسی کامپیوتر - گرایش نرم افزار. دانشکده مهندسی برق و کامپیوتر. دانشگاه تبریز.

(۷) مهرآفرید، سمیرا، اکبرپورسکه، محمد، (۱۳۹۷)، تشخیص تومورهای مغزی با استفاده از الگوریتم خوشه بندی فازی و هوش ازدحامی، فصلنامه پژوهش های کاربردی در فنی و مهندسی ۲ (۸)

8) Anyanwu, M. N., & Shiva, S. G. (2009). Comparative analysis of serial decision tree classification algorithms. *International Journal of Computer Science and Security*, 3(3), 230-240.

شده در این پژوهش می باشد که دقت نزدیک به ۱۰۰ را ارائه می دهد.

بنابراین، بهترین روش با توجه به نتایج حاصل در این پژوهش به صورت زیر قابل بیان است:

(۱) دریافت داده های ورودی.

(۲) تبدیل ویژگی ها (اعمال تغییرات برای تغییر نوع ویژگی از اسمی به عددی و برعکس، بسته به نیاز) و محاسبه ی نمره ی استاندارد برای ویژگی های عددی

(۳) اعمال الگوریتم CHIAD بر روی داده ها

(۴) اعمال فرآیند تجمعی Boosting بر روی شبکه ی عصبی MLP با دو لایه ی پنهان که لایه ی اول ۵ نورون و لایه ی دوم ۱۰ نورون دارد، با در نظر گرفتن داده های آموزش

(۵) اعمال مدل به دست آمده در مرحله ی ۶، بر روی داده های تست.

در پایان می توان گفت کارایی روش ها و الگوریتم های داده کاوی، تا حد زیادی به مجموعه داده های مورد بررسی وابسته می باشند. به نحوی که برخی از الگوریتم ها، بر روی برخی از داده ها نتایج خوبی ارائه می دهند و بر روی داده های متفاوت، نتایج حاصل، شاید چندان مناسب نباشند. از این رو، پژوهشگران می توانند، از بهترین راهکارهای ارائه شده در این تحقیق، بر روی مجموعه داده های متفاوت در حوزه ی درمان بیماران ارتوپدی استفاده کنند و دقت حاصل از آن را مورد ارزیابی قرار دهند. مورد دیگری که از آن می توان به عنوان افق های آتی این پژوهش به آن اشاره کرد، بهره گیری از امکانات برنامه نویسی خارج از محیط Spss Modeler، به منظور بررسی دقیق تر ساختارهای متفاوت شبکه ی عصبی بر روی مجموعه داده ی مورد نظر می باشد. برای این منظور می توان از روش های تکاملی نظیر الگوریتم ژنتیک، برای تعیین ساختار مناسب شبکه ی عصبی بهره برد که متاسفانه این امکانات در محیط نرم افزار فراهم نشده است.

فهرست منابع

(۱) اشقلی فراهانی، مونا، ربیعی، محمد (۱۳۹۶). استفاده از الگوریتم بیز ساده و قوانین انجمنی جهت کشف

- algorithms. *International Journal of Current Engineering and Technology*, 3(2), 334-337.
- 22) Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS one*, 10(3), e0118432.
- 23) Soni, J, Ansari, U, Sharma, D, Soni, S (2011), "Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction", *International Journal of Computer Applications* (0975 – 8887), Volume 17–No.8
- 24) Zhou, X., Chen, S., Liu, B., Zhang, R., Wang, Y., Li, P. & Yan, X. (2010). Development of traditional Chinese medicine clinical data warehouse for medical knowledge discovery and decision support. *Artificial Intelligence in medicine*, 48(2-3), 139-152
- 9) Babu, M. S. P., Swetha, R., Ramana, B. V., & Murty, N. V. R. (2013). A web-based soya bean expert system using bagging algorithm with C4. 5 decision trees. *International Journal of Agriculture Innovations and Research*, 1(4), 91-96.
- 10) DODDI, S, MARATHE, A, RAVIOE, S.S, TORNEY, D.S (2001), "Discovery of association rules in medical data", *med. Inform* , vol. 26, no. 1. 25± 33
- 11) Endo A, Shibata T, & Tanaka H. (2008). Comparison of seven algorithms to predict breast cancer survival. *Biomed Soft Comput Hum Sci*; 13: 11-6 .
- 12) Han, J., Pei, J., & Kamber, M. (2011). *Data mining: concepts and techniques*. Elsevier.
- 13) Kazemi A, Yousof zadeh A, & Azimi P. (2014). Detect of Hepatitis by combining data mining methods. *The first national conference application of intelligent systems in science and technology*.
- 14) Kaur, H, Wasan, S.K (2006), "Empirical Study on Applications of Data Mining Techniques in Healthcare", *Journal of Computer Science* 2 (2): 194-200
- 15) Khalid, A. M, Siraj, F (2013), "Utilization of an Artificial Neural Network in the Prediction of Heart Disease," *International Journal of Bio-Science and Bio-Technology*, vol. 5, pp. 159-166.
- 16) Kolahi S, & Rafe W. (2014). Presentation a solution to diagnosis using data mining techniques. *The first national conference electrical and computer southern Iran*.
- 17) Kumari, M, Godara, S (2011), "Comparative Study of Data Mining Classification Methods in Cardiovascular Disease Prediction", *International Journal of Computer Science and Technology*, Vol. 2, Issue 2
- 18) Mahmoodi A, Asgarimoghadam R, Moazam M, & Saeghian S. (2014). Identify models predict coronary artery disease using Neural networks and variable selection based on classification and regression tree. *J Shahrekord Med Sci Uni*; 15:22-7. (Persian)
- 19) Malathi, A., & Santra, A. K. (2013). Diagnosis of lung cancer disease using neuro-fuzzy logic. *CARE Journal of applied research*, 1(1), 6-9.
- 20) Maldonado, M., Dean, J., Czika, W., & Haller, S. (2014). Leveraging ensemble models in SAS® Enterprise Miner™. In *Proceedings of the SAS Global Forum 2014 Conference*.
- 21) Priyama, A., Abhijeeta, R. G., Ratheeb, A., & Srivastava, S. (2013). Comparative analysis of decision tree classification

یادداشت‌ها

1. Data Mining
2. Neural Network
3. Ensemble method
4. Classification
5. Clustering
6. Association Rule
7. Pelvic incidence
8. Sacral plate
9. Pelvic tilt
10. Lumbar lordosis angle
11. Sacral slope
12. Pelvic radius
13. Grade of spondylolisthesis
14. Disk Hernia
15. Preprocessing
16. Z Score
17. K-Fold Cross Validation