

ارایه یک روش بهبود یافته در سیستمهای انفورماتیک پزشکی جهت تشخیص برخی از بیماریها به کمک هوش مصنوعی

مهدی بردبار^۱

^۱ کارشناس ارشد کامپیوتر، emehdibordbar@gmail.com

چکیده

دست آوردن روابط مفید بین عوامل خطرزا در بیماریها با توجه به شیوع و سهمی که در مرگومیر انسانها دارند استفاده می شوند (کوفمن^۱، ۲۰۰۶).

داده کاوی دانش استخراج روابط و الگوهای مفید پنهان در حجم زیاد داده است. اینگونه مطالعات و کاوشها را به واقع می توان همان امتداد و استمرار دانش کهن و فراگیر آمار دانست. تفاوت عمده در مقیاس وسعت و گوناگونی زمینه ها و کاربردها و نیز ابعاد و اندازه های داده های امروزی است که شیوه های ماشینی مربوط به یادگیری مدل سازی و تعلم را طلب می نماید (هال^۲، ۲۰۱۱).

تشخیص بیماریهای قلبی یک کار قابل توجه و خسته کننده در علم پزشکی می باشد و وظیفه مهم و کار پیچیده ای است که باید با دقت و کارآمدی انجام گیرد. با این حال ابزارهایی برای تجزیه و تحلیل استخراج داده ها وجود دارد که در دسترس بودن این مجموعه عظیم از داده های پزشکی منجر به تجزیه و تحلیل درستی در این زمینه گردیده است. با استفاده از اطلاعات پزشکی از قبیل سن، جنس، فشارخون و قند خون می توان احتمال پیش بینی بیماریهای قلبی را بیشتر کرد.

در اکثر مقالات پیشین انجام شده در ایران از همه متغیرهای به وجود آورنده بیماری استفاده نشده و براساس حدس و گمان و یا براساس مقالات خارجی انجام شده و کار بر روی تعدادی متغیر صورت گرفته است که باعث شده در هنگام ارزیابی، نتایج آنها از دقت بسیار کمتری برخوردار باشند. اما در این مقاله با جمع بندی تمام متغیرهای شناخته شده دارای اهمیت در مقالات گذشته داخلی و خارجی و نیز تایید آنها براساس نظر پزشک متخصص مهمترین متغیرها جمع آوری و براساس آن مدلسازی انجام گرفته است که نتایج آن نیز از دقت بسیار بالایی برخوردار میباشد.

پیشینه تحقیق

۱- زن^۳ (۲۰۰۶) تحقیق بر روی ۳۱۳ داده در دو کلاس طبیعی و بیمار قلبی انجام داد. در این تحقیق جهت شناسایی و پیشگویی حملات قلبی از روشهای داده کاوی غیر از

تکنیک های داده کاوی در زمینه پزشکی رشد روز افزونی داشته است. استخراج قواعد طبقه بندی نوعی داده کاوی است که در آن دانشی به شکل چندین قانون ساده و فهم پذیر از داده کشف شده و در آینده برای تصمیم گیری و پیشگویی به کار برده می شود. شایع ترین بیماریهایی که امروزه مطرح است بیماری های عروق کرونر، نارسایی قلبی، سکته قلبی، آریتمی قلبی و کاردیومیوپاتی می باشند، رشد چشم گیر این بیماریها و اثرات و عوارض آنها و هزینه هایی که بر جامعه وارد می کند باعث شده است که جامعه پزشکی به دنبال برنامه هایی جهت بررسی بیشتر پیشگیری، شناسایی زود هنگام و درمان موثر آن باشند. هدف این مقاله پیش بینی با خطای کمتر این بیماریها در داده های بیمارستانها با استفاده از ویژگیها و ارتباطات بین ویژگی های مرتبط با بیماری از طریق ترکیب تکنیک های مختلف داده کاوی می باشد. داده ها مورد استفاده در این مقاله با استفاده از متغیرهای شناخته شده در پایگاه تشخیص بیماری قلبی موجود در مرکز UCI از بیمارستانهای سطح کشور جمع آوری شده است که شامل اطلاعات ۹۹۴ بیمار می باشد این اطلاعات در قالب فایل اکسل با ۱۸ ویژگی جمع آوری شده که فیلد آخر نظر پزشک معالج مبنی بر حمله قلبی یا عدم حمله قلبی است. در این مقاله مدل بدست آمده مبتنی بر درخت تصمیم علاوه بر توانایی بالا در تشخیص افراد بیمار و همچنین کارآمدترین الگوریتم در تحلیل داده های عظیم که دارای بیشترین دقت در تشخیص این بیماری می باشد.

واژه های کلیدی

بیماری قلبی، داده های عظیم، درخت تصمیم، داده کاوی، ماتریس اغتشاش

مقدمه

امروزه در دانش پزشکی جمع آوری داده های فراوان در مورد بیماریهای مختلف از اهمیت فراوانی برخوردار است. مراکز پزشکی با مقاصد گوناگون به جمع آوری این داده ها می پردازند. تحقیق روی این داده ها و به دست آوردن نتایج و الگوهای مفید در رابطه با بیماریها یکی از اهداف استفاده از این داده ها است. حجم زیاد این داده ها و سردرگمی حاصل از آن مشکلی است که مانع رسیدن به نتایج قابل توجه می شود. بنابراین از داده کاوی برای غلبه بر این مشکل و به

^۱ Kaufmann

^۲ Hall

^۳ Zhan