کلان داده و یادگیری ماشینی بر مبنی چارچوب مراقبت های بهداشتی مطمئن

چکیده:

این مقاله یک مقدمه مختصر به کلان داده و نقش ان در برنامه های بهداشتی را ارائه می دهد. مشاهده شده است که استفاده از معماری و تکنیک های کلان داده به طور مداوم به مدیریت سریع رشد داده ها در صنعت مراقبت های بهداشتی کمک می کند. اینجا؛ ابتدا یک مطالعه تجربی برای تجزیه و تحلیل نقش کلان داده در صنعت بهداشتی انجام می شود. مشاهده می شود که کارهای مهمی با استفاده از کلان داده در بخش مراقبت بهداشتی انجام شده است. امروزه؛ پیچیده است تا پیش بینی کنیم که یادگیری ماشینی و کلان داده می تواند بر صنایع بهداشتی تاثیر بگذارد. مشاهده شده است که بیشتر نویسندگانی که استفاده از یادگیری ماشینی و تجزیه و تحلیل کلان داده را در تشخیص بیماری پیاده سازی کرده اند؛ توجه خاصی به حریم خصوصی و امنیت داده ها نداشته اند. در اینجا؛ یک طراحی جدید از سیستم اطلاعات بهداشتی هوشمندانه و مطمئن با استفاده از یادگیری ماشینی و مکانیسم امنیتی پیشرفته برای بکار بردن کلان داده در صنعت پزشکی پیشنهاد شده است. این نوآوری در ترکیب ذخیره سازی بهینه و لایه امنیتی داده مورد استفاده برای حفظ امنیت و حریم خصوصی است. تکنیک های دیگری در ان مانند رمز گذاری پنهانی؛ نظارت بر فعالیت ها؛ کنترل دسترسی گرانشی؛ رمزگذاری داده پویا و تصدیق نقطه پایان گنجانده شده است. مدل ترکیب چهار لایه مراقبت های بهداشتی پیشنهاد شده به نظر می رسد برای سیستم کلان داده در تشخیص بیماری موثرتر است.

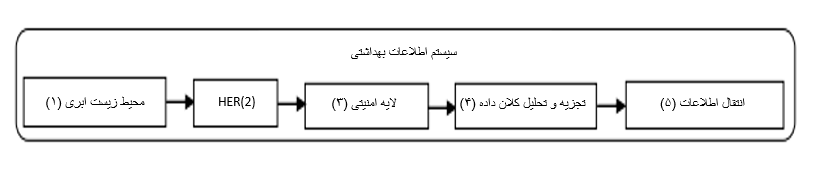
کلمات کلیدی: کلان داده؛ مراقبت های بهداشتی؛ تجزیه و تحلیل کلان داده؛ تشخیص بیماری؛ تجزیه و تحلیل پیش بینی شده؛ امنیت و حریم خصوصی.

1. مقدمه:

امروزه؛ با توجه به توسعه فوری اینترنت و محاسبات ابری؛ داده ها به سرعت با سرعتی غیر قابل کنترل در هر سازمان رشد می کند(۱). Wal-Marts تقریبا هر ساعت 2.5 گیگابایت اطلاعات را به پایگاه داده وارد می کند؛ فیس بوک هر روز بیش از ۲۵۰ میلیون عکس و ۹۰۰ میلیون موضوع و غیره را اداره می کند (۲). با توجه به این رشد انفجاری داده ها؛ توضیحاتی برای به دست آوردن بینش ارزشمند از مجموعه داده ها ضروری است. استفاده موثر از داده ها مهم است زیرا آنها به عنوان یک بلوک ساختمانی برای یک سازمان تلقی می شود. تجزیه و تحلیل موثر داده ها می تواند در پیش بینی فروش؛ تشخیص بیماری؛ تجزیه و تحلیل اقتصادی؛ مدیریت کسب و کار؛ تحلیل شبکه های اجتماعی و غیره بسیار مفید باشد. بعضی از سازمان ها قبلا از طریق تجزیه و تحلیل اطلاعات سازمان یافته در قالب گزارش ها استفاده می کرده اند. ایده اولیه کلان داده در مقاله ( کاوش بصری پایگاه داده گیگا بایت در زمان واقعی) معرفی شد و ۱۹۹۹ منتشر گردید (۳). در ۲۰۰۱؛ Doug Laney ویژگی های کلان داده را در 3Vs یعنی سرعت؛ حجم و تنوع در مقاله ۳ بعدی مدیریت داده خود تعریف کرد: کنترل حجم؛ سرعت و تنوع اطلاعات. هادوپ (Hadoop) یکی از چارچوب های غالب است که برای مدیریت و تجزیه و تحلیل کلان داده بدون ساختار استفاده می شود (۴). به طور کلی، کلان داده به مقدار حجم و پیچیدگی داده های جمع آوری شده از منابع مانند وب، برنامه های سازمانی، دستگاه های تلفن همراه و مخازن دیجیتال اشاره دارند که نمی توانند به راحتی با استفاده از ابزارهای سنتی مدیریت شوند. کلان داده نه تنها در مورد اندازه بزرگ داده ها است بلکه یک عمل ذخیره و مدیریت اطلاعات برای تحلیل نهایی انها است (۵). بنابراین؛ همانطور که انسان ها دیجیتالی می شوند؛ محاسبات داده ها با تنوع؛ حجم و سرعت بیشتر را در بر می گیرد (۶). همانطور که Doug Laney است؛ انواع متنوعی V در کلان داده درگیر است؛ با این وجود؛ مدل 3Vs مدل پایه ای در کلان داده است. اهمیت 3Vs به طور خلاصه در زیر ذکر شده است (۷.۸):

* حجم: کلان داده تجزیه و تحلیل مقایسه ای یک مقدار زیادی داده مانند ترا بایت است (۹.۱۰).
* تنوع: کلان داده داده ها از منابع متنوع را شامل می شود که در صورت های متعددی مانند ساختار یافته؛ بدون ساختار؛ چند منظوره؛ احتمالی ظاهر می شود (۹.۱۰).
* سرعت: کلان داده پردازش سریع داده برای ارتقا فرایند تصمیم گیری را کنترل می کند (۹.۱۰).

یکی دیگر از ویژگی های کلان داده پیوستگی همراه با 3Vs مانند صحت است که نشان دهنده اعتبار و سازگاری داده ها برای کاربران است (۹.۱۰). کلان داده فرصت های فراوان در بخش مراقبت های بهداشتی را برای بهبود کارایی و کیفیت مراقبت های بهداشتی، تشخیص تهدیدات بهداشتی، مدیریت سلامت انسان توسط تشخیص بیماری در مراحل اولیه و کمک به تصمیم های بهتر فراهم می کنند (۱۱). با این وجود؛ ذخیره سازی داده ها، استاندارد سازی داده ها، انتقال داده ها و حفظ حریم خصوصی برخی از چالش های عمده در صنعت کلان داده است (۱۲.۱۳). محققان متعددی تلاش کرده اند جنبه های امنیتی را در نظام مراقبت های بهداشتی در نظر بگیرند. دو تقریب متفاوت در زیر ارائه شده است: احمد یوسف (۲۰۱۴) یک چارچوب بر اساس تجزیه و تحلیل کلان داده برای سیستم اطلاعات بهداشتی (HIS) به صورت نشان داده شده در شکل ۱ پیشنهاد کرده است. این چارچوب برای مدیریت و تجزیه و تحلیل مقدار زیادی از داده ها به نفع بیماران و همچنین متخصصان مراقبت های بهداشتی مورد استفاده قرار گرفت. چارچوب HIS شامل ۵ جزء اصلی است که در زیر ذکر شده است.



شکل۱: چارچوب سیستم اطلاعات بهداشتی پیشنهاد شده به وسیله یوسف ۲۰۱۴ (۳).

جزءاول محیط ابری استفاده شده برای فراهم کردن سرویس های متنوع و اجازه اشتراک داده ها با کاربران مجازی می باشد. جزء دوم بایگانی بهداشت الکترونیک مورد استفاده برای کامل کردن داده های بیمار از مکان های مختلف است. جزء سوم لایه امنیتی است؛ جایی که مسائل امنیتی مختلف مانند؛ احراز هویت؛ مجوز؛ محرمانه بودن اطلاعات بیمار به طور موثر مدیریت شد. این لایه شامل محاسبات عددی رمزنگاری مختلف یعنی AES؛ RC4 برای محافظت داده ها است. همچنین تکنیک های مختلف احراز هویت مانند رمز عبور یک بار مصرف OTP؛ دو فاکتور تایید هویت 2FA برای محدود کردن کاربران غیر مجاز استفاده شد. OTP یک روش مورد استفاده برای ایجاد رمز عبور مختلف در هر زمان برای امنیت داده ها در شبکه است. این لایه همچنین برای اعطای امتیاز یا لغو امتیاز به کاربران مجاز استفاده می شود. نه تنها این لایه برای بیماران و متخصصان مراقبت های بهداشتی مفید بود، بلکه سازمان های بیمه درمانی نیز برای تأیید تصمیم گیری در مورد درمان بیماران خود از ان بهره مند شدند. لایه چهارم HIS برای راه اندازی ابزارهای تحلیلی داده های بزرگ مورد استفاده قرار گرفته است و از داده های خام بینش را دریافت می کند. در نهایت سازمان های تحویل مراقبت اطلاعات بهداشت و درمان را در مکان های مختلف توزیع می کنند. این چارچوب مطمئن محرمانه بودن و حریم خصوصی اطلاعات بهداشتی برای بهبود امنیت و کیفیت خدمات بهداشتی را ایجاد می کند (۱۴). برای حفظ حریم خصوصی اطلاعات بهداشتی بیماران Bouhriz Mounia و همکاران (2015) روش دو طرفه را معرفی کردند. روش اول به کنترل دسترسی غیر مجاز در پرونده پزشکی در زمان تولید، پردازش، به اشتراک گذاری و ذخیره سازی داده ها تمرکز می کند. اینها را می توان به وسیله استفاده از محاسبات عددی رمز نگاری مختلف و استراتژی پشتیبان گیری داده ها انجام داد. در نیمه دوم روش؛ پروفایل بیماران برای محافظت از مدارک پزشکی بیماران را تولید می کند. اینها می تواند به وسیله بکار بردن برخی از محدودیت های حریم خصوصی مانند OTPs انجام شود. روش دوم به ایجاد و محافظت پرونده های پزشکی بیماران تمرکز می کند (۱۵). در ابتدا؛ کلان داده و نقش ان در بخش های مختلف به طور خلاصه ارائه شده است. علاوه بر این؛ ۱۹ مقاله مختلف مربوط به کلان داده و مراقبت های بهداشتی مورد بررسی قرار گرفته و نتایج آنها برجسته شده است. این نوآوری در چارچوب مراقبت های بهداشتی پیشنهادی قرار دارد. یک چارچوب چهار لایه پیشنهاد شده است. امنیت داده ها و لایه های حفظ حریم خصوصی برای حفظ میزان محرمانه بودن و حفظ حریم خصوصی داده های بیماران گنجانیده شده است. بخش 2 برنامه های کاربردی کلان داده را در بخش های مختلف ارائه می دهد. در بخش 3، تجزیه و تحلیل برنامه های کلان داده بر اساس پایگاه داده گوگل ارائه شده است. بخش 4 روشی را در مورد ابزار و تکنیک های تجزیه و تحلیل کلان داده ها نشان می دهد. مراقبت های بهداشتی و تجزیه و تحلیل کلان داده در بخش 5 مورد بحث قرار گرفته است. بخش 6 فرایند انتخاب مقاله را ارائه می دهد. طراحی کلان داده مطمئن و چارچوب مراقبت بهداشتی بر اساس یادگیری ماشینی در بخش 7 ارائه شده است. نتیجه گیری در بخش 8 ارائه شده است.

1. کاربردهای کلان داده:

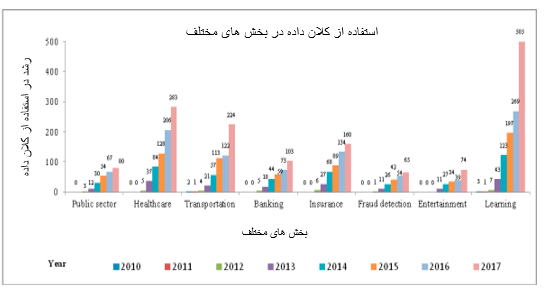
از سال 2010، کلان داده ها همچنان باقی مانده است. کلان داده به طور فزاینده ای در بخش های مختلف صنعتی، اجتماعی و حرفه ای مورد استفاده قرار گرفته است [16]. بعضی از دسته های اصلی و برنامه های کاربردی که کلان داده در ان استفاده می شود، در جدول 1 نشان داده شده است.

جدول۱: کاربردهای کلان داده در دسته بندی های مختلف



1. تجزیه و تحلیل کاربردهای کلان داده بر مبنی پایگاه داده **Google Scholar**

برای یافتن استفاده دقیق از کلان داده در کاربردهای مختلف؛ پرسشنامه های زیر در Google Scholar اجرا شده است و تجزیه و تحلیل هشت سال گذشته (2010 تا 2017) در شکل 2 نشان داده شده است. "برنامه های کلان داده" در "بخش عمومی"، "برنامه های کلان داده" در "بهداشت"، "برنامه های کلان داده" در "حمل و نقل"، "برنامه های کلان داده" در "بانکداری"، "برنامه های کلان داده "در" بیمه "،" برنامه های کلان داده "در" تشخیص کلاه برداری "،" برنامه های کلان داده "در" سرگرمی "،" برنامه های کلان داده "در" یادگیری ". شکل 2 نشان دهنده افزایش استفاده از کلان داده در سال 2010 تا 2017 است. مشاهده شده است که برای به دست آوردن اطلاعات ارزشمند از داده ها، کلان داده ها با موفقیت در بخش های مختلف مانند عمومی، مراقبت های بهداشتی، حمل و نقل، بانکی، بیمه، سرگرمی، یادگیری و تشخیص کلاه برداری مورد استفاده قرار می گیرد.



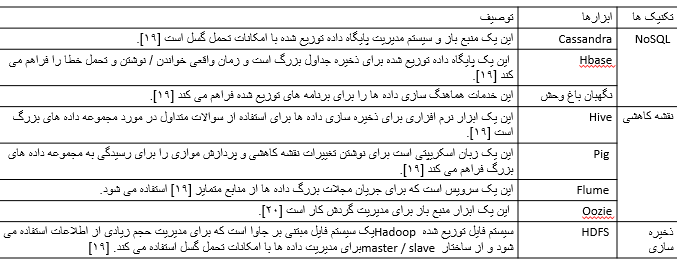
شکل۲: مقایسه کاربردهای کلان داده در بخش های متنوع از ۲۰۱۰ تا ۲۰۱۷

حداکثر کار در بخش بهداشت و درمان و آموزش با کلان داده در سال 2017 انجام شد. بالاترین میزان افزایش استفاده از کلان داده در بخش یادگیری مشاهده شده است. سرعت تشدید در هشت سال گذشته 6/99 درصد است. به همین ترتیب، میزان افزایش 99٪ و 96٪ به ترتیب برای بخش حمل و نقل و بیمه است. با این حال، سرگرمی و بانکداری مناطقی هستند که کمترین میزان استفاده از کلان داده در انها دیده شده است.

1. ابزارها و تکنولوژی های مورد استفاده در تجزیه و تحلیل کلان داده:

کلان داده ها یک سازمان را برای ذخیره سازی و مدیریت مقدار زیادی از داده ها با سرعت بالا برای به دست آوردن بینش مناسب از داده ها قادر می سازد. بیشتر ابزارها و تکنولوژی ها برای کشف داده های ناهمگن و داده های بزرگ برای تبدیل خام داده ها بر روی اطلاعات برای بهبود فرایند تصمیم گیری مورد نیاز است. تجزیه و تحلیل کلان داده BDA به ابزارها و روش هایی اشاره دارد که کلان داده را به اطلاعات مفید برای استفاده تحلیلی تبدیل می کند. به طور کلی؛ BDA ترکیبی از فن آوری هایی شامل IT، کسب و کار حرفه ای و اطلاعات دانشمندان است. این معادلات با درک عمیق تر از سازمان های کسب و کار به این نتیجه میرسد که سازمان می تواند در مسیر درست حرکت کند. با استفاده از BDA، ممکن است یکی از طرفین رقیب خود را به رقابت بفرستد. علت ظهور BDA رشد بی نظیری از داده های سالانه است. امروزه بیش از میلیون؛ تریلیون ترافیک داده بیش از یک روز یا گاهی در یک ساعت وجود دارد. در توییتر و فیس بوک؛ صدها میلیون؛ "توییت" و میلیاردها دلار در یک روز پست گذاشته می شود. به طور متوسط؛ ۲.۵ quintillions بایت در یک روز تولید می شود مانند یک حجم داده از ماموت است. برای مدیریت کلان داده تکنیک های متنوعی وجود دارد؛ اما اینها می تواند بسته به نوع داده ها مورد تجزیه و تحلیل قرار گیرد و سوال یا مشکل تحقیق مورد نظر حل شود (۱۷.۱۸). بر اساس مشکلات و نوع داده ها که مورد تجزیه و تحلیل قرار می گیرد، می توان از ابزارهای مختلف کلان داده استفاده کرد. جدول ۲ برخی از ابزارها و تکنیک ها را در زمینه کلان داده نشان می دهد.

جدول۲: ابزارها و تکنیک هایی برای تجزیه و تحلیل کلان داده



1. کیفیت مراقبت های بهداشتی و تجزیه و تحلیل کلان داده:

داده مراقبت های بهداشتی و پزشکی نوعی از کلان داده است که نه تنها به خاطر حجم و اندازه بزرگ انها بلکه به خاطر پیچیدگی و به روز بودن انها است. اطلاعات بهداشتی شامل اطلاعات مرتبط با بیمار یعنی گزارش های بهداشتی الکترونیکی بیمار (EHR)، گزارش های تشخیصی، نسخه پزشک، تصاویر پزشکی، سوابق داروخانه و داده های تحقیق از مجلات دارویی است (۲۱). امروزه، ضروری است که داده های تولید شده توسط صنایع بهداشتی به منظور بهبود کیفیت درمان، تشخیص زودهنگام بیماری ها برای جلوگیری از عوامل خطر و مدیریت بهتر سیستم اطلاعات در بیمارستان ها، دیجیتالی شود. فرصت های شگرف برای تاثیر گذاری بر کیفیت و بهره وری مراقبت های بهداشتی با استفاده از تجزیه و تحلیل کلان داده وجود دارد. بینش های تحلیلی از تجزیه و تحلیل قابل توجهی از داده های پزشکی حاصل شده است که می تواند احتمالا به مدل های بالینی و بهره وری را از طریق تحویل عقلانی مراقبت بهبود می بخشد (۲۲). برخی از وظایف چالش برانگیز برای صنعت بهداشت عبارتند از:

- چگونگی تصمیم گیری برای درمان موثر تر بیماری خاص است؟

- چطور سیاست های خاصی بر هزینه و رفتار تاثیر می گذارند؟

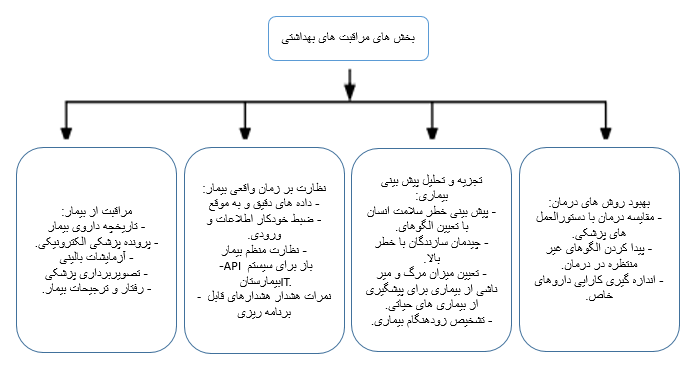
- چقدر هزینه های مراقبت های بهداشتی در آینده برای جنبه های مختلف افزایش می یابد؟

- چطور ادعا می شود کلاه بردار شناسایی شده است؟

- آیا هزینه های مراقبت های بهداشتی جغرافیایی متفاوت است؟

چهار رکن اصلی برای مراقبت های بهداشتی با کیفیت وجود دارد. شکل 3 این ستون ها را نشان می دهد.

شکل۳: کلان داده در بخش مراقبت های بهداشتی



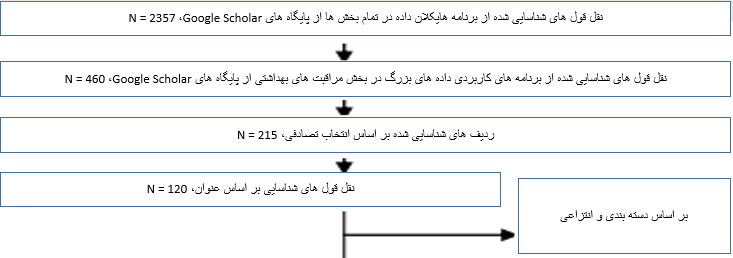
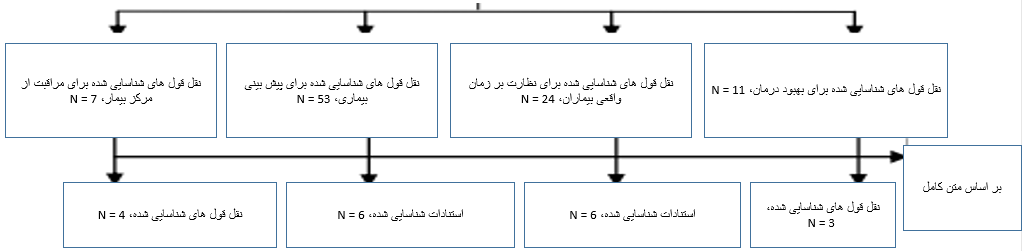
تمام این چهار ستون مراقبت های بهداشتی با کیفیت می تواند با استفاده از تکنیک های تجزیه و تحلیل کلان داده ها توصیف، پیش بینی و تجربی قابل استفاده باشد.

* مراقبت مرکزی بیمار: تسکین دادن بیمار با ارائه تشخیص بیماری بر اساس داده های بالینی در مراحل اولیه و کم کردن دوز دارو برای اجتناب از عوارض جانبی است. این کار به کاهش سرعت پذیرش در بیمارستان ها و نیز کاهش هزینه بیماران کمک می کند.
* پیش بینی تحلیل بیماری: پیش بینی اختلالات ویروسی در مرحله اولیه قبل از گسترش یر اساس تحلیل زندگی است. اینها می تواند به وسیله تحلیل کارنامه اجتماعی بیمارانی که از یک بیماری در یک مکان خاص رنج می برند تعیین شود. اینها بیشتر به متخصصان بهداشت و درمان کمک می کند تا به قربانیان با اقدامات پیشگیرانه لازم توصیه لازم را انجام دهند (۲۳).
* زمان واقعی نظارت بر بیمار: نظارت بر این که ایا بیمارستان با توجه به تنظیم هنجارهای شورای پزشکی هند، تنظیمات را انجام می دهد. این بررسی دوره ای به دولت کمک می کند تا اقدامات لازم را علیه بیمارستان های رد صلاحیت انجام دهد (۲۳).
* بهبود روش های درمان: نظارت بر درمان بیماران سفارش شده وابسته به تجزیه و تحلیل دوز دارو می تواند برای تسکین سریع درد تغییر داده شود. تحلیل اطلاعات بیمارانی که دارای علائم مشابهی هستند به بیماران کمک می کند تا داروهای موثر برای بیماران جدید ارائه دهند (۲۳).

بعضی از بیماری ها روانشناسی، قلب و عروق، اختلال کبدی و دیابت با مسائل پزشکی درازمدت مرتبط است. تجزیه و تحلیل کلان داده به بهبود مراقبت های بهداشتی به وسیله پیش بینی اولیه بیماری هایی که بهبود بیشتر انها امید به زندگی بیماران را بالا می برد؛ کمک می کند. با توجه به پیش بینی های اولیه، کارشناسان مراقبت های بهداشتی همچنین می توانند درمانی را در مرحله اولیه پیشنهاد دهند (۲۱).

1. کلان داده در مراقبت های بهداشتی:

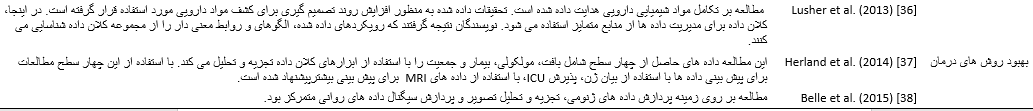
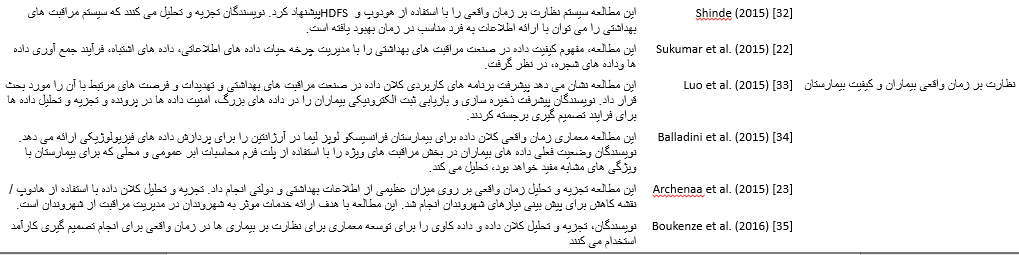
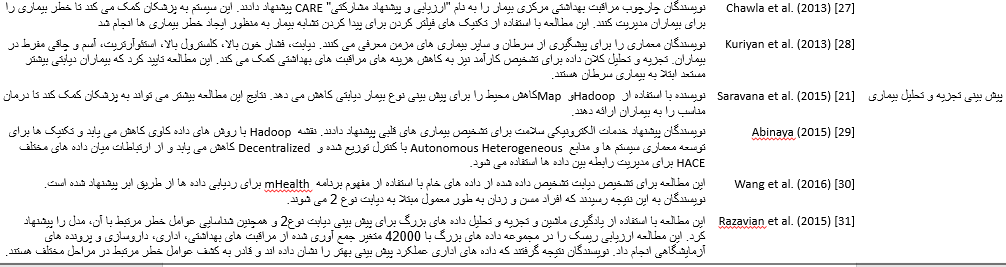
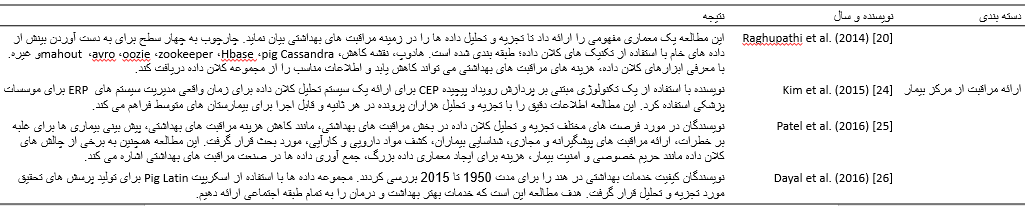
این مطالعه شامل بررسی سیستماتیک در برنامه های کاربردی کلان داده در بخش مراقبت های بهداشتی می باشد. تعداد 2357 مقاله از سال 2010 تا 2016 برای برنامه های کاربردی کلان داده در همه بخش ها منتشر شده است.

شکل۴: فرایند انتخاب مقاله

در اینجا، 460 مقاله نشان دهنده عملکرد کلان داده در مراقبت های بهداشتی است. بعد از ازمایشات موفقیت امیز (شش سطح)؛ در نهایت ۱۹ مقاله در این مطالعه یکپارچه شده است. این مقاله منتخب در چهار بخش تحت بخش مراقبت های بهداشتی مراقبت از مرکز بیمار ( ۴ مقاله)؛ پیش بینی بیماری ( ۶ مقاله)؛ نظارت بر زمان واقعی بیماران ( ۶ مقاله) و بهبود روش های درمان ( ۳ مقاله) تقسیم می شود. جدول 3 نتایج تحقیقات سال 2013 تا 2016 را در صنعت بهداشت و درمان با استفاده از کلان داده نشان می دهد.

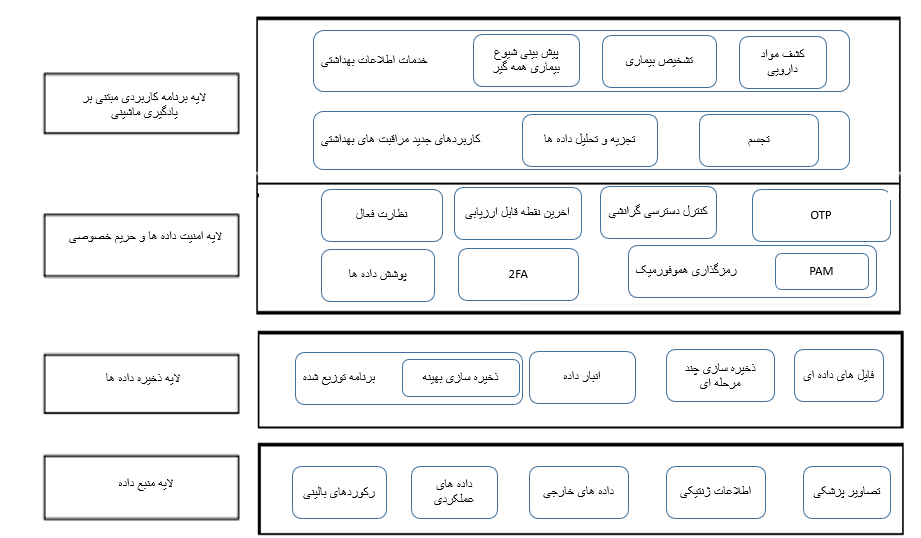
جدول۳: مروری بر کاربردهای کلان داده در صنعت مراقبت های بهداشتی



امروزه در بخش مراقبت های بهداشتی اندازه داده ها نه تنها یک مسئله مهم است، بلکه امنیت و حفظ حریم خصوصی پرونده های پزشکی نیز به همان اندازه مهم است. به طور کلی؛ نویسندگان در این مطالعات تنها بر روی قابلیت همکاری و برهمکنش کلان داده برای تحلیل الگوهای اموزنده و اکثر نگرانی ها همراه امنیت متمرکز شده اند. در حالیکه امنیت پرونده پزشکی چالشی عمده است.

1. طراحی کلان داده مطمئن و یادگیری ماشینی بر مبنی چارچوب مراقبت های بهداشتی:

انگیزه مشورت در بالا برای استفاده از کلان داده های زیست بوم در مراقبت های بهداشتی برای ادرس های مرتبط با راه حل ها جهت حریم خصوصی و بهینه سازی داده ها مورد نیاز است. شکل 5 نشان دهنده طراحی سیستم بهداشتی مطمئن و مطلوب است. سیستم دارای چهار لایه مانند منبع داده، ذخیره سازی داده ها، امنیت و لایه کاربردی مبتنی بر یادگیری ماشینی است. این یک نوآوری در بهینه سازی داده ها و ویژگی های امنیتی است که در این چارچوب متحد گنجانده شده است. محیط توزیع شده و ذخیره سازی بهینه به بهینه سازی منابع ذخیره سازی کمک می کند که در نهایت کارایی عملکردی سیستم را بالا می برد. علاوه بر این؛ تلاش برای حفظ امنیت و حریم خصوصی داده های بیماران به وسیله ترکیب تکنیک های حفاظت از داده های مختلف ساخته شده است. لایه اول با منبع داده ناهمگن سروکار دارد. پرونده های بالینی، داده های عملیاتی (فایل های صاف، داده های ارتباطی و فرمت ASCII)، اطلاعات ژنوم، تصاویر پزشکی (X-ray، MRI، CT Scan و غیره) منبع اصلی لایه داده هستند. لایه دوم مسئول ذخیره داده های ناهمگن است. بسته به نوع داده، می توان آن را در قالب فایل های داده یا در انبار داده ذخیره کرد. بنیاد توزیع شده امکان تسهیل تجزیه داده ها را فراهم می کند و آن را در سایت های مختلف ذخیره می کند. در نهایت این لایه از بهینه سازی ذخیره سازی مراقبت می کند تا منابع حافظه به طور موثر مورد استفاده قرار گیرد. در بخش مراقبت های بهداشتی، مهمترین نکته امنیت است. در این مدل، برای اطمینان از اطلاعات بیماران، تکنیک های مختلف مانند رمزگذاری پنهانی، رمزنگاری داده ها، کنترل دسترسی دانه ها و واحد نظارت بر فعالیت در لایه امنیت و حریم خصوصی گنجانده شده است. یادگیری ماشینی بر مبنی لایه کاربردی لایه با اهمیت دیگری از چارچوب پیشنهاد شده است. این پنج واحد زیر تشخیص زودهنگام، کشف مواد دارویی، پیشگیری از شیوع بیماری همه گیر، تجزیه و تحلیل و تجسم داده ها نامیده می شود. در اینجا روش های مختلف یادگیری ماشینی مانند بله ساده؛ پشتیبانی از بردار ماشینی؛ درخت تصمیم گیری؛ محاسبات عددی ژنتیک و غیره ممکن است برای واحد های زیر استفاده شود. بسته به عملکرد و دقت، تکنیک های یادگیری ماشین ممکن است با سایر تکنیک های محاسباتی نرم افزاری ترکیب شود تا نتایج بهتری به دست آید. در ۴ لایه، مدل های پیش بینی شده با استفاده از روش های کلان داده برای بهبود فرآیند تصمیم گیری و کسب اطلاعات از داده ها توسعه می یابد. این لایه اطلاعات خدمات بهداشتی و برنامه های کاربردی را پوشش می دهد. در نهایت، اطلاعات تولید شده را می توان با استفاده از یادگیری ماشین، پرس و جو SQL و تجزیه و تحلیل تصویربرداری پزشکی ارائه نمود. این چارچوب پیشنهاد شده است تا یکپارچه سازی و تجزیه و تحلیل داده های توزیع شده را ارائه دهد که به بیماران بیشتر کمک می کند تا درمان های بهتری انجام شود و کارشناسان مراقبت های بهداشتی تصمیم گیری موثرتری داشته باشند. حفظ حریم خصوصی و امنیت به عنوان یک مسئله مهم در کلان داده محسوب می شود. حریم خصوصی داده ها شامل استفاده مناسب از داده ها برای محافظت از دسترسی غیر مجاز می باشد. در حقیقت؛ امنیت عمل حفاظت از داده ها و منابع داده ای را با استفاده از حفاظت از فن آوری و روش های آماده سازی داده ها با بازرسی، دسترسی غیر مجاز، ضبط و غیره است. حفظ حریم خصوصی داده ها بدون هیچگونه شیوه امنیتی دشوار است. روش های متنوع یعنی رمزگذاری پنهان، نظارت بر فعالیت، کنترل دسترسی گرانشی و رمزنگاری داده ها برای مدیریت حریم خصوصی و امنیت کلان داده در بخش مراقبت های بهداشتی استفاده می شود.



شکل۵: طراحی کلان داده مطمئن و یادگیری ماشینی بر اساس چارچوب مراقبت های بهداشتی

پوشش داده ها محرمانه بودن داده به وسیله رمزگذاری بخش های اصلی اطلاعات شناسایی شخصیت (PII) مانند اطلاعات شخصی بیمار را تضمین می کند. در انها داده های حساس به وسیله جایگزین کردن با داده های دیگر برای تولید داده های اصلی مبهم یا سردر گم می شود. در مراقبت های بهداشتی؛ داده های پنهان می توان در زمان واقعی با استفاده از تکنیک های مختلف EHR انعکاسی مانند مبهم کردن داده ها، اختلال در داده ها، حذف داده ها، خرد کردن داده ها و غیره بدست اید. برخی از محاسبات عددی مانند: محاسبات عددی رمزنگاری، رمزگذاری ماتریسی، محاسبات عددی MASK و برنامه های امنیتی استفاده می شود تا مقدار زیادی از اطلاعات اصلی بیماران را به طور موثر پنهان کنند. این محاسبات عددی همچنین می توانند برای رمزگذاری اطلاعات تصویر مانند اشعه ایکس و گزارش های MIR بیماران استفاده شوند (۳۹). راه دیگری برای حفاظت از داده ها رمزنگاری داده است. این کار شامل تبدیل داده های اصلی به الگوهای کد شده (به عنوان مثال متن رمز) برای محدود کردن دسترسی غیر مجاز است. هنگامی که داده ها رمزنگاری شده است، یک روش رمزنگاری برای دریافت اطلاعات اصلی از متن رمز وجود دارد. رمزگذاری داده ها را می توان در سطوح مختلف مانند سطح ستون یا میدان، سطح دیسک، فایل های HDFS، سطح بسته در خوشه ها برای محافظت از داده ها انجام داد. الگوریتم های مختلف متقارن و نامتقارن برای رمزگذاری پرونده پزشکی بیمار مانند استانداردهای رمزگذاری داده (DES)، DES سه گانه، استاندارد رمزگذاری پیشرفته (AES) و blowfish مورد استفاده قرار می گیرند.

در میان تمام محاسبات عددی رمزنگاری، blowfish رمزنگاری کلان داده را در زمان کمتر با بهترین شرایط امنیتی انجام می دهد (۴۰). یکی از راه های ردیابی امنیت داده ها، نظارت بر فعالیت است که در آن فعالیت های کاربران ممتاز تحت مشاهده قرار دارد. این روش به کنترل دسترسی به داده های مدیریت شده توسط کاربران مهمان کمک می کند، فعالیت های مظنون را مسدود می کند، تغییرات در امتیاز ها را شناسایی می کند و حقوق استفاده نشده را لغو می نماید. برای مدیریت سطوح قابل قبول برای کاربر مجاز، روش های کنترل دسترسی گرانشی برای محدود کردن اقدامات کاربران استفاده می شود (۴۱). رمزنگاری هومورفیک را می توان برای رمزگشایی نتیجه یک پرس و جو کلان داده استفاده کرد. اینها نتایج پرس و جو را تجزیه می کند به طوری که در اجرای پرس و جو کاربران غیر مجاز قادر به دریافت آن نیست. سپس کاربران مجاز می توانند از محاسبات عددی رمزگشایی همگن بر روی نتایج استفاده کنند تا آن را به متن ساده تبدیل نمایند. علاوه بر این، برای ایمن سازی داده های غیر مرتبط، روش های رمزنگاری مانند AES، RSA، SHA-256 می تواند مورد استفاده قرار گیرد. علاوه بر این، امکانات مانند TLS و SSL نیز می تواند گنجانیده شود. استفاده ازنمونه تصدیق شده قابل تنظیم (PAM) می تواند در حفاظت از داده ها بیشتر کمک کند. امنیت پایانی می تواند به طور موثر برای شناسایی و اتصال دستگاه های سیمی یا تلفن همراه مورد اعتماد مورد استفاده قرار گیرد (۴۲).

1. نتیجه گیری:

این مقاله معرفی مختصر کلان داده و برنامه های کاربردی آن را فراهم می کند. هدف اصلی این مطالعه ارائه یک چارچوب مراقبت بهداشتی با داده های بهینه و مطمئن است. با توجه به مقالات موجود مشاهده شده است که استفاده از کلان داده در صنعت مراقبت های بهداشتی، با ارائه سلامتی و اطلاعات بهتر به بیماران، صنعت پزشکی یک انقلاب صنعتی ایجاد می نماید. علاوه بر این، استفاده از فناوری اطلاعات به کاهش هزینه های مربوط به تشخیص مراقبت های بهداشتی کمک می کند. در اینجا، در ابتدا مفهوم کلان داده همراه با کاربردهای اصلی ان به اختصار معرفی شده است. از تجزیه و تحلیل آماری منتشر شده، مشخص است که در هشت سال گذشته، تحقیقات زیادی در زمینه استفاده از کلان داده در بخش های مجزا مانند عمومی، مراقبت های بهداشتی، حمل و نقل، بانکداری، بیمه، تشخیص کلاه برداری، سرگرمی و یادگیری انجام شده است. علاوه بر این، تحقیق عالی با استفاده از کلان داده در بخش مراقبت های بهداشتی و بخش یادگیری انجام شده است. چهار بخش اصلی بخش مراقبت های بهداشتی یعنی: مراقبت از بیمار، نظارت بر بیمار در زمان واقعی، پیش بینی بیماری واگیردار و روش های بهبود درمان برجسته شده است. علاوه بر این، یک مدل مراقبت بهداشتی چهار لایه مطمئن و اختراع شده طراحی و پیشنهاد شده است. لایه پایین، یعنی لایه منبع داده، با منابع داده ناهمگن سروکار دارد. این لایه برای مدیریت داده های ناهمگن و تبدیل ناهمگنی به داده همگن در صورت نیاز قرار دارد. لایه دوم لایه ذخیره داده ها باید فرایند ذخیره سازی بهینه را مدیریت کند. لایه امنیتی داده ها و لایه های حفظ حریم خصوصی چندین ویژگی امنیت و حفظ حریم خصوصی پیشرفته مانند پنهان کردن داده، نظارت فعالیت، رمزگذاری همگرا، PAM و کنترل دسترسی گرانشی را فراهم می کند. در نهایت، لایه کاربردی مبتنی بر یادگیری ماشین، از وظایف مختلف مانند تشخیص بیماری، کشف مواد دارویی، تجزیه و تحلیل داده ها و پشتیبانی تجسم مراقبت می کند. در اینجا، برای تشخیص بیماری های اولیه، تکنیک های یادگیری ماشین مانند داده کاوی سنتی و محاسبات الهام گرفته از طبیعت در حال ظهور است. دقت در تشخیص بیماری ها می تواند با ترکیب مفهوم منطق فازی و نظریه اطلاعات بیشتر شود. چهار لایه از این چارچوب وجود دارد. ویژگی منحصر به فرد این مدل؛ نهفته در بهینه سازی و امنیت اطلاعات بیماران است. برای استفاده بهینه از منابع سیستم، تکنیک های مختلف بهینه سازی مانند شاخص بندی، نرمال سازی استفاده شده است. علاوه بر این، ویژگی های امنیتی مختلف مانند رمزگذاری پنهانی، نظارت بر فعالیت، کنترل دسترسی گرانشی و رمزنگاری داده ها، اطلاعات بیمار را خصوصی و امن می کند. مدل پیشنهادی برای بررسی نتایج تجربی قابل اجرا می باشد. طراحی می تواند به طریقی که چارچوب بر روی دستگاه های مختلف تلفن همراه مانند تلفن همراه، تبلت، phablet و غیره کار کند؛ افزایش یابد. طبیعت در حال ظهور الهام بخش تکنیک ها است و پیوند زدن انها می تواند به منظور بهبود دقت این چارچوب تشخیصی اجرا شود. در نهایت تلاش ها می تواند برای بررسی اثر پیش پردازش و انواع مختلف داده ها در کل کارآیی این سیستم گسترش یابد.