

بسمه تعالی



دانشگاه تهران
دانشکده ریاضی، آمار، علوم کامپیوتر
گروه علوم کامپیوتر

تشخیص بیماری هیپاتیت C از روی ویژگیهای ۱۳۸۵ بیمار مصری

نگارش

محمدحسین شهریاری

استاد راهنما

دکتر هدیه ساجدی

هپاتیت

هپاتیت به معنی التهاب کبد است که عمدتاً بوسیله عوامل ویروسی ایجاد میشود. ۵ ویروس اصلی برای ایجاد بیماری هپاتیت وجود دارند که با نامهای A, B, C, D و E شناخته میشوند. به دلیل شیوع بالا و احتمال مرگ بیماران، این ۵ نوع بیشتر مورد توجه هستند. به طور خاص، انواع هپاتیت B و C باعث بیماری به صورت مزمن در میان صدها تا میلیونها انسان میشوند و از عمده ترین دلایل ابتلا به سیروز و سرطان کبدی میباشند.

راه های انتقال بیماری هپاتیت

به طور کلی عمده راه های انتقال این ویروس ها شامل موارد زیر می باشند:

- دریافت خون یا محصولات خونی آلوده
- آلوده بودن وسایل جراحی و یا هر وسیله ای که با خون و مایعات بدن در تماس باشد
- انتقال از مادر مبتلا به نوزاد در زمان تولد
- انتقال از طریق اعضای خانواده به کودک
- تماس جنسی غیر ایمن

علائم هپاتیت شامل علائم پوستی و ظاهری

علائم هپاتیت نوع حاد عمدتاً قابل تشخیص نیستند (علائمی ندارد) و یا علائم بسیار محدود هستند و در صورت بروز علائم هپاتیت احتمالاً موارد زیر مشاهده میشود:

۱. زردی (زرد شدن رنگ پوست و چشم ها)
۲. تیره شدن رنگ ادرار
۳. خستگی بسیار زیاد (غیر معمول)
۴. حالت تهوع و استفراغ
۵. درد در ناحیه شکم

هپاتیت C

هپاتیت C یک بیماری کبدی است که در اثر ویروس هپاتیت (HCV) ایجاد شده است: این ویروس می تواند هپاتیت حاد و مزمن را ایجاد کند و از شدت بیماری خفیفی که چند هفته به طول می انجامد تا یک بیماری جدی و مادام العمر منجر شود. هپاتیت C یکی از دلایل اصلی سرطان کبد است.

ویروس هپاتیت C یک ویروس منتقل شده از خون است: شایعترین حالت های عفونت از طریق قرار گرفتن در معرض مقادیر کمی خون است. این ممکن است از طریق استفاده از مواد مخدر تزریقی، شیوه های تزریقی ناامن، مراقبت های بهداشتی ناامن، انتقال خون و فرآورده های خونی بدون خون و اقدامات جنسی که منجر به قرار گرفتن در معرض خون می شوند، رخ دهد.

در سطح جهان، تخمین زده می شود ۷۱ میلیون نفر مبتلا به ویروس مزمن و هپاتیت C هستند. تعداد قابل توجهی از مبتلایان مزمن به سیروز یا سرطان کبد مبتلا می شوند. WHO تخمین زده است که در سال ۲۰۱۶، تقریباً ۳۹۹ ۰۰۰ نفر در اثر هپاتیت C درگذشتند، که اکثراً بر اثر سیروز و کارسینوم سلولهای کبدی (سرطان اولیه کبد) است. داروهای ضد ویروسی می توانند بیش از ۹۵٪ از مبتلایان به عفونت هپاتیت C را درمان کنند، از این طریق خطر مرگ ناشی از سیروز و سرطان کبد را کاهش می دهد، اما دسترسی به تشخیص و درمان کم است.

در حال حاضر هیچ واکسن مؤثر علیه هپاتیت C وجود ندارد. با این حال، تحقیقات در این زمینه ادامه دارد. ویروس هپاتیت C باعث عفونت حاد و مزمن می شود. عفونت HCV جدید معمولاً بدون علامت است. برخی از افراد به هپاتیت حاد مبتلا می شوند که منجر به یک بیماری تهدید کننده زندگی نمی شود. حدود ۳۰٪ (۱۵-۴۵٪) از افراد آلوده به طور خودجوش ویروس را ظرف ۶ ماه از عفونت بدون هیچگونه درمانی پاک می کنند. ۷۰٪ باقیمانده (۵۵-۸۵٪) از افراد به عفونت مزمن HCV مبتلا می شوند. در بین مبتلایان به عفونت مزمن HCV، خطر سیروز بین ۲۰ تا ۳۰ درصد در طی ۲۰ سال متغیر است.

انتقال

ویروس هپاتیت C یک ویروس منتقله از خون است. بیشترین موارد از طریق:

- تزریق استفاده از مواد مخدر از طریق به اشتراک گذاشتن تجهیزات تزریق
- استفاده مجدد یا عقیم سازی نامناسب تجهیزات پزشکی ، به ویژه سرنگ و سوزن در تنظیمات مراقبت های بهداشتی.
- انتقال خون و فرآورده های خونی ضد خون ،
- اقدامات جنسی که منجر به قرار گرفتن در معرض خون می شود (به عنوان مثال ، در بین مردانی که با مردان رابطه جنسی دارند ، به ویژه با عفونت HIV یا افرادی که پیشگیری از مواجهه با عفونت HIV را ندارند).

HCV همچنین می تواند از نظر جنسی منتقل شود و از مادر آلوده به نوزاد خود منتقل شود. با این حال ، این حالت های انتقال کمتر رایج است. هپاتیت C از طریق شیر مادر ، غذا ، آب یا تماس گاه به گاه مانند آغوش گرفتن ، بوسیدن و به اشتراک گذاشتن غذا یا نوشیدنی با فرد آلوده پخش نمی شود. WHO تخمین می زند که در سال ۲۰۱۵ ، ۱.۷۵ میلیون عفونت HCV جدید در جهان وجود داشته است (۲۳.۷ عفونت HCV جدید در هر ۱۰۰۰۰۰ نفر).

علائم

دوره نهفتگی هپاتیت C از ۲ هفته تا ۶ ماه متغیر است. به دنبال عفونت اولیه ، تقریباً ۸۰٪ از افراد علائمی از خود نشان نمی دهند. افرادی که علامت حاد دارند ممکن است تب ، خستگی ، کاهش اشتها ، تهوع ، استفراغ ، درد شکم ، ادرار تیره ، مدفوع به رنگ خاکستری ، درد مفاصل و زردی (زردی پوست و سفیدی چشم) نشان دهند.

کاربرد هوش مصنوعی در سلامت

صنایع بسیاری توسط خیل تکنولوژی جدید در عصر اطلاعات مختل شده اند. بهداشت و سلامت هم از این دسته هستند. خصوصا در زمینه ی اتوماسیون، فراگیری ماشین، هوش مصنوعی، پزشکان، بیمارستانها، شرکت های بیمه و صنایع مربوط به سلامت و بهداشت همگی متاثر از این قضیه هستند- در بیشتر موارد تاثیرات مثبت نسبت به صنایع دیگر. طبق گزارشی از CB Insight تقریبا ۸۶ درصد از سازمانهای تامین بهداشت و سلامت، شرکت های علم زندگی و فروشندگان فناوری به خدمات بهداشتی از تکنولوژی هوش مصنوعی استفاده می کنند. این سازمانها میانگینی برابر ۵۴ میلیون دلار را تا سال ۲۰۲۰ بر روی پروژه های هوش مصنوعی سرمایه گذاری خواهند کرد. آنها دنبال چه راه حل هایی هستند؟ این بخش ده مورد از تاثیرات هوش مصنوعی بر سلامت و بهداشت را در زمان حال و آینده بیان می کند.

۱- کنترل پیشینه ی پزشکی و سایر اطلاعات

از آنجا که اولین قدم در سلامت و بهداشت جمع آوری و بررسی اطلاعات (مانند پیشینه ی پزشکی و تاریخ) است، مدیریت اطلاعات رایج ترین کاربرد هوش مصنوعی و اتوماسیون دیجیتال است. ربات ها اطلاعات را جمع آوری، ذخیره و تغییر فرمت می کنند و باعث دسترسی بهتر و سرعت بالا می شوند.

۲- انجام کارهای تکراری

بررسی آزمایشات، رادیولوژی، سی تی اسکن، ورود اطلاعات و دیگر امور توسط ربات ها سریعتر و دقیق تر انجام می شود. قلب شناسی و پرتوشناسی دو مورد هستند که حجم اطلاعات آنها بسیار زیاد و زمان بر است. قلب شناسان و پرتوشناسان در آینده باید تنها به مواردی نگاه کنند که نظارت انسان در آن ضروری است.

۳- طرح درمان

سیستم های هوش مصنوعی برای این منظور طراحی شده اند که اطلاعات را بررسی کنند- نکات و گزارشات پرونده ی بیمار، تحقیقات خارجی، اختصارات پزشکی- و راه منحصر به فرد و بهتری را برای درمان طراحی کنند.

۴- مشاوره ی دیجیتال

برنامه هایی چون بابلون در بریتانیا از هوش مصنوعی برای مشاوره ی پزشکی با توجه به پیشینه ی پزشکی و اطلاعات عمومی استفاده می کنند. کاربران علائم بیماری خود را در برنامه تایپ می کنند، که از شناسایی صدا برای مقایسه ی این علائم با بانک اطلاعاتی انواع بیماری استفاده می کند. بدین صورت بابلون راهکارهایی با توجه به تاریخچه ی پزشکی فرد پیشنهاد می کند.

۵- پرستاران مجازی

The startup sense.ly "مولی" را خلق کرده است، پرستاری مجازی که شرایط بیماران را کنترل می کند و پیگیر درمان در بین دو ملاقات پزشک است. این برنامه از فراگیری ماشینی برای پشتیبانی بیماران استفاده می کند که در بیماری های مزمن تخصصی سازی شده. در سال ۲۰۱۶، Boston Children's Hospital برنامه ای برای "آمازون الکسا" طراحی کرد که اطلاعات سلامت پایه ای و توصیه های لازم برای والدین کودکان بیمار را داشت. این برنامه به سوالات درمانی پاسخ داده و تعیین می کند که آیا فرد نیازمند مراجعه به پزشک است.

۶- مدیریت دارو

موسسات ملی سلامت برنامه ی AiCure را به منظور کنترل مصرف دارو توسط بیمار طراحی کردند. یک وب کم تلفن هوشمند در کنار هوش مصنوعی فعالیت می کند که می تواند به طور مستقل پیگیری درمان توسط بیمار را تایید کند و شرایط آنها را مدیریت نماید. کاربران رایج این برنامه می توانند افرادی با شرایط پزشکی وخیم، بیمارانی که خلاف توصیه های پزشکی عمل می کنند و شرکت کنندگان در آزمایشات کلینیکی باشند.

۷- ساخت دارو

توسعه ی دارو سازی از طریق آزمایشات کلینیکی می تواند یک دهه زمان ببرد و میلیاردها دلار خرج بردارد. ارزان کردن و سریعتر کردن این روند می تواند جهان را تغییر دهد. حین فراگیری اخیر ویروس ایبولا، از برنامه ای که با هوش مصنوعی کار می کرد برای بررسی داروهای موجود که می توانستند با طراحی دوباره برای مقابله با بیماری مصرف شوند استفاده شد. این برنامه دو دارو یافت که ممکن بود عفونت ایبولا را در یک روز کاهش دهد، در حالی که این نوع بررسی به طور کلی ماه ها و سال ها طول می کشد- تفاوتی که باعث نجات هزاران زندگی می شود.

۸- دقت دارو

ژنتیک و ژنوم شناسی بدنبال جهش ها و پیوند های بیماری از طریق اطلاعات DNA هستند. با کمک هوش مصنوعی اسکن بدنی می تواند سرطان و بیماری های قلبی عروقی را زودتر پیش بینی کند و شرایط سلامت افراد را با توجه به ژنتیک آنها بسنجد.

۹- کنترل سلامت

گیرنده های سلامتی که قابلیت پوشش دارند- مانند گیرنده های فیت بیت، اپل، گارمین و...- ضربان قلب و دیگر فعالیت ها را کنترل می کند. این وسایل می توانند به کاربر هشدار بدهند که به ورزش بیشتری نیاز دارند و همچنین می توانند این اطلاعات را با پزشکان و دیگر سیستم های هوش مصنوعی برای تکمیل نیازهای بیمار در میان گذارند.

۱۰- بررسی سیستم سلامت

در هلند، ۹۷ درصد فاکتورهای سلامتی بصورت دیجیتال هستند. یک شرکت هلندی از هوش مصنوعی برای بررسی اطلاعات و مشخص کردن اشتباهات درمان و بی کفایتی های کار استفاده می کند که باعث پرهیز سیستم سلامت منطقه از بستری های غیر ضروری می شود.

اینها تنها نمونه ای از راهکارهایی هستند که هوش مصنوعی به صنایع بهداشت و سلامت پیشنهاد می کند. باتوجه به توانایی اتوماسیون سازی که با تکنولوژی پیشرفت کرده و نیروی کار دیجیتال از طریق تهیه کنندگانی چون Novatio، راه حل های بیشتری برای صرفه جویی در وقت، کاهش هزینه و افزایش دقت پیشنهاد خواهد شد.

کاربرد هوش مصنوعی در درمان هپاتیت

تشخیص بیماری هپاتیت با توجه به علائم بالینی اندکی که در مراحل اولیه دارد، بسیار دشوار است. بنابراین روشی که بتواند این بیماری را در مراحل اولیه آن پیش‌بینی کند می‌تواند بسیار ارزشمند باشد و باعث کاهش عوارض احتمالی بعدی آن شود.

نمونه های مرتبط

Barakat و همکاران در سال ۲۰۱۷ انواع مختلفی از الگوریتم‌های شبکه عصبی شامل شبکه‌های عصبی سریع، چندگانه، پویا و شعاعی را برای تشخیص بیماری هپاتیت با استفاده از داده‌های هپاتیت موجود در مخزن داده UCI مورد بررسی قرار دادند.

Husain و همکاران در سال ۲۰۱۷ جهت تشخیص بیماری هپاتیت، مدل ماشین بردار پشتیبان و الگوریتم بهبود یافته مورچه را ارائه دادند. در این تحقیق پس از کاهش داده‌ها جهت تعیین پارامترهای مطلوب ماشین بردار پشتیبان از الگوریتم بهبود یافته مورچه استفاده کردند.

Elaboudi و همکاران در سال ۲۰۱۶ تکنیک ترکیبی PCA-CE SVM را برای افزایش دقت تشخیص بیماری هپاتیت پیشنهاد دادند. به این صورت که ابتدا با استفاده از الگوریتم PCA ابعاد بردار ویژگی‌ها کم می‌شود و سپس با استفاده از روش بهینه‌سازی و الگوریتم SVM آموزش شبکه انجام می‌شود.

Karthikeyan و همکاران در سال ۲۰۱۵ براساس بهترین مناسب و جستجوی حریمانه مبتنی بر الگوریتم‌های طبقه‌بندی Naïve Bayes و CFS به تشخیص بیماری هپاتیت پرداختند. در این تحقیق از روش‌های جستجوی بهترین مناسب و جستجوی حریمانه و ارزیابی کننده ویژگی CFS استفاده شده است. الگوریتم طبقه‌بندی Naïve Bayes بر روی پایگاه داده بیماری هپاتیت اعمال می‌شود و طبقه‌بندی را انجام می‌دهد.

Neshat و همکاران در سال ۲۰۱۴ از شبکه عصبی هاپفیلد فازی برای تعیین شدت بیماری هپاتیت استفاده کردند. در این مقاله از مجموعه داده‌های هپاتیت از مخزن یادگیری ماشین UCI استفاده شد.

مجموعه داده های هپاتیت

مجموعه داده‌های بکار رفته در این تحقیق، از پایگاه داده UCI دانشگاه کالیفرنیا گرفته شده است. این مجموعه داده، دارای ۱۳۸۶ نمونه میباشد که هر نمونه دارای ۲۸ ویژگی ورودی و دو ویژگی هدف است که نشان دهنده

میزان سلامت هر نمونه میباشد. این ویژگیها شامل سن ، جنسیت ، BMI ، تب ، حالت تهوع، سر درد ، تعداد گلبولهای سفید و گلبولهای قرمز و میزان هموگلوبین خون و غیره می‌باشند که جمع آوری گردیده و بصورت یک فایل ذخیره شده است تا بتوان روشهای یادگیری ماشین را بر روی آن اعمال کرد. به طور دقیق تر ، برای بررسی از آخرین ویژگی هدف برای دسته بندی به پنج دسته " [No Fibrosis], [Portal Fibrosis], [Few Septa], [Many Septa], [Cirrhosis] استفاده میکنیم تا میزان پیشرفت بیماری را تشخیص دهیم.

روش پیشنهادی

در این تحقیق، قصد داریم جهت پیشبینی و تشخیص بیماری هپاتیت C از الگوریتم "k-نزدیک‌ترین همسایگی" استفاده کنیم. (k-Nearest Neighbors) یک روش ناپارامتری است که در داده‌کاوی، یادگیری ماشین و تشخیص الگو مورد استفاده قرار می‌گیرد. برای این منظور پس از انجام پیش پردازشهای لازم بر روی مجموعه داده، مختصات دکارتی KNN ایجاد و آموزش داده میشود و سپس وزنهاى آن توسط الگوریتم مورد ارزیابی قرار میگیرد. با توجه به اینکه ما برای حل مسئله از روش یادگیری ماشین و برای دسته بندی از روش KNN استفاده می‌کنیم بنابراین از حالت طبقه بندی این الگوریتم استفاده خواهیم کرد. در حالت طبقه بندی با توجه به مقدار مشخص شده برای k ، به محاسبه فاصله نقطه ای که میخواهیم برچسب آن را مشخص کنیم با نزدیک ترین نقاط میپردازد و با توجه به تعداد رای حداکثری این نقاط همسایه، در رابطه با برچسب نقطه مورد نظر تصمیم گیری می‌کنیم. برای محاسبه این فاصله میتوان از روش های مختلفی استفاده کرد که یکی از مطرح ترین این روش ها، فاصله اقلیدسی است.

داده‌های اولیه، بردارهایی در یک فضای چند بعدی هستند که هر کدام شامل برچسبی به نام دسته می‌باشند. فاز یادگیری (training phase) الگوریتم، شامل ذخیره‌سازی بردارهای ویژگی و برچسب دسته نمونه‌های اولیه است. در فاز طبقه‌بندی، k یک ثابت توسط کاربر تعریف می‌شود و بردار بدون برچسب (نقطه تست) از دسته ای است که بیشترین تعداد را در k نزدیک‌ترین همسایه آن نقطه داشته باشد. به این ترتیب برچسب نقطه تست نیز مشخص می‌شود. معیار فاصله برای متغیرهای پیوسته معمولاً فاصله اقلیدسی است.

مراحل الگوریتم KNN شامل موارد زیر خواهد بود:

۱. داده‌های را بارگیری کنید.
۲. K به عنوان تعداد نزدیک‌ترین همسایگان انتخاب کنید.
۳. برای هر یک از داده‌های اولیه:

۱. فاصله بین داده مورد سؤال و هر یک داده‌های اولیه را محاسبه کنید.
 ۲. فاصله و اندیس نمونه را به یک مجموعه اضافه کنید.
 ۳. مجموعه را بر اساس فاصله از کوچک به بزرگ مرتب کنید.
 ۴. نقاط K عضو اول مجموعه مرتب شده را انتخاب کنید.
 ۵. بسته به حالت یا حالت طبقه‌بندی، خروجی را اعلام کنید.
- کد پایتون محاسبه فاصله اقلیدسی و KNN را در ادامه می‌بینید :

```
def EuclideanDistance(x1, x2, length):  
    distance = 0  
    for x in range(length):  
        distance += np.square(x1[x] - x2[x])  
    return np.sqrt(distance)
```

پیاده سازی در زبان پایتون

کتابخانه Scikit-learn که مبتنی بر SciPy می‌باشد، برای کاربردهای یادگیری ماشین از جمله خوشه بندی، رگرسیون و کلاس بندی ابزار قدرتمند و در عین حال ساده ای می‌باشد. یادگیری استفاده از آن بسیار سریع و آسان بوده و انعطاف آن باعث می‌شود تا بتوان آن را در مسائل متنوعی مورد استفاده قرار داد. روش‌های متنوعی در آن فراهم آمده است که برای مثال در زمینه کلاس بندی می‌توان موارد زیر را نام برد:

- K-Nearest Neighbors
- Linear SVM
- RBF SVM
- Gaussian Process
- Decision Tree
- Random Forest
- Neural Networks

برای پیاده سازی الگوریتم پیشنهادی در زبان پایتون از کتابخانه Scikit-learn استفاده می‌کنیم.

- [١] <https://archive.ics.uci.edu/ml/datasets/hepatitis>
- [٢] <http://archive.ics.uci.edu/ml/machine-learning-databases/hepatitis/hepatitis.data>
- [٣] G.K. Singh, R.K. Jain, P. Dubey, Study of Classification Techniques on Medical Datasets, Computing, Communication and Signal Processing, Vol. 810, pp. 557-565, 2019
- [٤] T.P. Vital, M.M. Krishna, G.V.L. Narayana, P. Suneel, P. Ramarao, Empirical Analysis on Cancer Dataset with Machine Learning Algorithms, Soft Computing in *Data Analytics*, Vol. 758, pp. 789-801, 2019.
- [٥] K. Polat, S. Gunes, Prediction of hepatitis disease based on principal component analysis and artificial immune recognition system, Applied Mathematics and Computation, Vol. 189, Issue 2, pp. 1282-1291, 2007
- [٦] P. Shunmugapriya, S. Kanmani, A hybrid algorithm using ant and bee colony optimization for feature selection and classification (AC-ABC Hybrid), Swarm and Evolutionary Computation, Vol. 36, pp. 27-36, 2017.
- [٧] <https://www.who.int/news-room/q-a-detail/what-is-hepatitis>
- [8] <https://www.who.int/news-room/q-a-detail/what-is-hepatitis>
- [9] <https://www.who.int/news-room/factsheets/detail/hepatitis-c>
- [10] M. Ashraf, G. Chetty, D. Tran, and D. Sharma, Hybrid Approach for Diagnosing Thyroid Hepatitis and Breast Cancer Based on Correlation Based Feature Selection and Naïve Bayes, Springer-Verlag Berlin Heidelberg, ICONIP 2012, Part IV, LNCS 7666, pp. 272–280, 2012.
- [11] M. Ghaemi, M.R. Feizi-Derakhshi, Feature selection using Forest Optimization Algorithm, Pattern Recognition, Vol. 60, pp. 121-129, 2016.
- [12] <https://www.cdc.gov/>
- [13] Barakat Saeed Alshamrani , Ahmed Hamza Osman, “ Investigation of Hepatitis Disease Diagnosis using Different Types of Neural Network

Algorithms” International Journal of Computer Science and Network Security, 17(2). (2017)

[14] Zhang, Y., Gue, L., Li, Q., Li, J.,. “Elrctricity consumption forecasting method based on MPSO-BP neural

network model”. Advances in Computer Science Research, 50. 674-678. (2016)

[15] Çalışır, D., & Dogantekin, E. “A new intelligent hepatitis diagnosis system: PCA–LSSVM.” Expert Systems with Applications, 38(8), 10705-10708, (2011)