

چگونه سرطان را دقیق و بهموقع پیش‌بینی کنیم؟

How to predict the cancer accurately and early?



زهرا توکلی
دانشآموخته‌ی کارشناسی ارشد
بیوتکنولوژی دانشگاه تهران
z.tavakoli72@ut.ac.ir



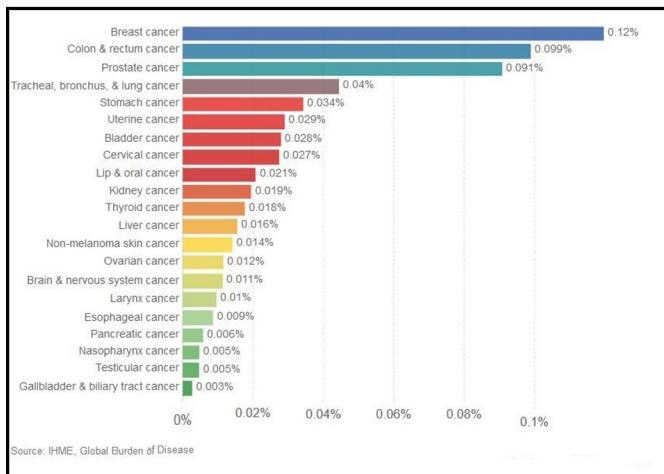
چکیده

سرطان یک بیماری چندعاملی است که سالیانه ۱۳ درصد از موارد مرگ‌ومیر انسانی را شامل می‌شود. این بیماری مراحل پیشرفت متعددی دارد و به چندین نوع متفاوت تقسیم می‌شود. در تحقیقات مربوط به سرطان، تشخیص زودهنگام و پیش‌بینی نوع سرطان مهم‌ترین عاملی است که می‌تواند مدیریت بالینی بیماران را تسهیل کند. با ظهور فناوری‌های جدید در زمینه‌ی پزشکی، مقدار زیادی از اطلاعات و داده‌ها در رابطه با سرطان‌های مختلف جمع‌آوری شد و در اختیار پزشکان قرار گرفت که لزوم بررسی اثر هر یک از عوامل در ارتباط با هم و انتخاب موثرترین عوامل را شدت می‌بخشد. از این رو محققان به سمت دانش هوش صنوعی و استفاده از روش‌های Machine Learning (ML) جهت یافتن، ML، شاخه‌ای از هوش مصنوعی است که با توجه به اطلاعات حاصل از نمونه داده‌ها، یک مفهوم کلی را به استنتاج می‌رساند؛ بنابراین این روش با فراهم آوردن توانایی‌های جدیدی نظری امکان‌شناسایی ویژگی‌های کلیدی از میان مجموعه داده‌های پیچیده، مدل‌سازی پیشرفته و درمان سرطان اهمیت می‌یابد. این تکنیک‌ها می‌توانند الگوها و روابط بین داده‌ها را از میان مجموعه داده‌های پیچیده بررسی کرده، میزان پیشرفته و نوع سرطان را گزارش دهند. طی سالیان اخیر، روش‌های ML میزان دقت در نتایج پیش‌بینی‌های مربوط به سرطان را به میزان قابل توجه ۲۰ تا ۲۵ درصد افزایش داده‌اند و همین موضوع سبب شده است که به طور فزاینده به عنوان ابزاری محبوب برای محققان پزشکی تلقی شوند.

کلیدواژه‌های پیش‌بینی سرطان - مجموعه داده - هوش مصنوعی - Machine Learning

مقدمه

به هوشمندی و قابلیتی که یک سیستم در تحلیل و مواجهه با شرایط مختلف از خودنشان می‌دهد، هوش مصنوعی (Artificial Intelligence) گفته می‌شود. این ابزار معرف سیستم‌هایی هستند که می‌توانند واکنش‌هایی مشابه رفتارهای هوشمند انسانی از جمله درک شرایط پیچیده، شبیه‌سازی فرایندهای تفکری و شیوه‌های استدلال انسانی، توانایی یادگیری و قدرت تحلیل برای حل مسائل را داشته باشند. امروزه با گسترش دانش، توسعه‌ی مجموعه داده‌ها و پیچیده‌تر شدن فرایند تصمیم‌گیری، استفاده از سیستم‌های هوش مصنوعی بسیار اهمیت یافته است. همچنین گسترش دانش پزشکی، لزوم بررسی عوامل متعدد تاثیرگذار در ابتلا و پیشرفته بیماری‌ها و پیچیدگی‌های تصمیمات مرتبط با تشخیص و درمان، توجه متخصصین را به استفاده از سیستم‌های هوشمند با دقت و سرعت بالا جهت داده است. در نگارش پیش رو لزوم استفاده از این سیستم‌ها در مطالعات مربوط به پیش‌بینی سرطان مورد بررسی قرار گرفته است.



شکل ۱: میزان شیوع انواع مختلف سرطان در جمعیت انسانی

گسترش دانش پزشکی، لزوم بررسی عوامل متعدد تأثیرگذار در ابتلا و پیشرفت بیماری‌ها و پیچیدگی‌های تصمیمات مرتبط با تشخیص و درمان، توجه متخصصین را به استفاده از سیستم‌های هوشمند با دقت و سرعت بالا جهت داده است.

سرطان؛ شیوع و مراحل پیشرفت

سرطان یک بیماری چند عاملی است که طیف وسیعی از عوامل ژنتیکی و محیطی در بروز آن دخالت دارند. این بیماری سالیانه ۱۳٪ از موارد مرگ انسانی را شامل می‌شود و طی آن، سلول‌ها از مسیر تقسیم طبیعی خود خارج شده و تکثیر مهار نشده‌ی سلول‌ها منجر به شکل‌گیری انواع تومور می‌شود. شکل ۱ میزان شیوع سرطان‌های مختلف در جمعیت انسانی را نشان می‌دهد. پیشرفت سرطان عموماً در ۴ مرحله رخ می‌دهد. این مراحل به ترتیب عبارت‌اند از: کارسینوم (Carcinoma) درجا یا مرحله‌ی ۱ که تنها لایه‌ای از سلول‌های سرطانی شده در میان سلول‌های سالم شروع به رشد کرده‌اند. در مرحله‌ی ۲ تعداد سلول‌های سرطانی شده در حال افزایش است و گستردگی تومور به بافت و اندامی که در آن تشکیل شده است محدود می‌شود. در مرحله‌ی ۳ سرطان از محل استقرار اولیه‌ی خود به گره‌های لنفی، بافت‌ها یا اندام‌های نزدیک انتشار می‌یابد (متاستاز) و مرحله‌ی ۴ که تومور بسیاری از بافت‌ها و اندام‌های بدن را درگیر می‌کند؛ بنابراین با تشخیص به موقع سرطان در مراحل اولیه، شناخت مسیر تکاملی سرطان و جلوگیری از پیشرفت آن در مراحل اولیه می‌توان به مقابله با آن پرداخت. از سیستم TMN برای توصیف مراحل پیشرفت سرطان استفاده می‌شود. T اندامی تومور را مشخص می‌کند، N گره‌های لنفی را نشان می‌دهد و برای توصیف میزان گسترش تومور بدخیم به گره‌های لنفی استفاده می‌شود و M بیانگر وقوع متاستاز و درگیری اندام‌های مختلف است.

دانش یادگیری ماشین (ML) یا Machine Learning فرایند یادگیری در انسان با استفاده از تجربیات شخصی و در سیستم‌های ماشینی با دنبال کردن دستورالعمل‌های داده شده به سیستم، محقق خواهد شد. یادگیری ماشین یا به اختصار (ML) شاخه‌ای از هوش مصنوعی است که تنوعی از تکنیک‌های آماری، احتمالی و بهینه‌سازی را به کار می‌برد و اجزا می‌دهد که کامپیوترها براساس مثال‌های ورودی، دستورالعمل‌هایی را یاد بگیرند و از آن برای شناسایی الگوهای هدف از میان مجموعه داده‌های بزرگ، شلوغ و پیچیده بهره ببرند. این دانش با توجه به الگوهای داده‌ها و روابط بین آن‌ها یک مفهوم کلی را به استنتاج می‌رساند. در هوش مصنوعی، هر فرایند یادگیری از ۲ مرحله تشکیل می‌شود: برآورد وابستگی‌های ناشناخته موجود در میان مجموعه داده‌های یک سیستم واستفاده از وابستگی‌های برآورد شده برای پیش‌بینی خروجی‌های جدید در سیستم. دونوع معمول از روش‌های ML شامل یادگیری تحت نظارت (Supervised) و یادگیری بدون نظارت

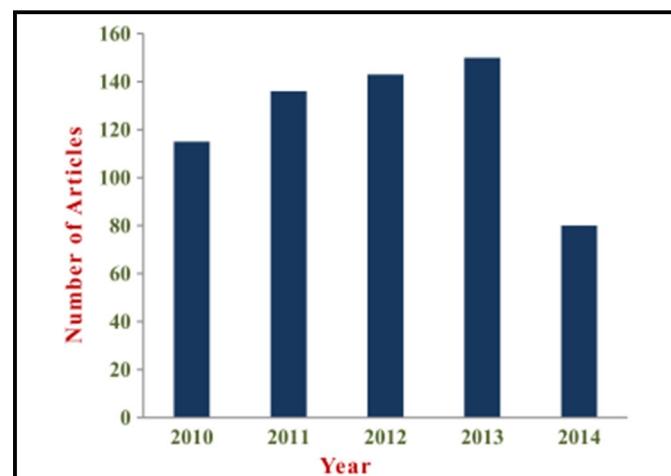
با تشخیص به موقع سرطان در مراحل اولیه، شناخت مسیر تکاملی سرطان و جلوگیری از پیشرفت آن در مراحل اولیه می‌توان به مقابله با آن پرداخت. از سیستم TMN برای توصیف مراحل پیشرفت سرطان استفاده می‌شود. T اندامی تومور را مشخص می‌کند، N گره‌های لنفی را نشان می‌دهد و برای توصیف میزان گسترش تومور بدخیم به گره‌های لنفی استفاده می‌شود و M بیانگر وقوع متاستاز و درگیری اندام‌های مختلف است.

دو نوع معمول از روش‌های ML شامل یادگیری تحت ناظارت (Supervised) و یادگیری بدون ناظارت (Unsupervised) است.

اهمیت بهره‌گیری از ابزارهای ML در پیش‌بینی دقیق و به موقع سرطان

طی دهه‌های گذشته، تکامل پیوسته‌ای در تحقیقات مرتبط با سرطان شکل گرفته است. به طور کلی، پیش‌بینی دقیق یک سرطان در مراحل مختلف وقوع، اعم از مرحله‌ی تشخیص و شناسایی نوع آن، شناخت مرحله‌ای که در آن قرار دارد، مدل‌سازی پیشرفت و پیش‌بینی پاسخی که در مراحل اولیه‌ی درمان نسبت به دارو ایجاد می‌کند، مهم‌ترین عواملی هستند که می‌توانند توانایی ما را در کنترل و مقابله با این بیماری افزایش دهند. دانشمندان در مراحل اولیه‌ی وقوع سرطان از روش‌های غربالگری بر پایه‌ی اطلاعات حاصل از بیان ژن و تولید پروتئین‌های ویژه در جهت تشخیص و شناسایی سرطان گام برداشتند. با این وجود، به دلیل غلظت کم این پروتئین‌ها در مراحل اولیه‌ی بیماری و حساسیت کم دستگاه های آنالیز (تکنیک‌های ایمونوفلوروئورسنت) نسبت به این مقدار کم و هزینه‌ی بالای آنالیزها این روش‌ها مناسب غربالگری نیستند. همچنین این تکنیک‌ها نمی‌توانند در تشخیص تومور خوش‌خیم از بدخیم مورد استفاده قرار گیرند. از سویی دیگر با ظهور و توسعه‌ی تکنولوژی‌های جدید، حجم زیادی از داده‌های ژنتیکی، بافت شناسی، جمعیت‌شناسی و کلینیکی در ارتباط با سرطان جمع آوری شده است که لزوم بررسی اثر هر یک از عوامل در ارتباط با هم و انتخاب موثرترین عوامل را شدت می‌بخشد. از این رو تکنیک های ML با فراهم آوردن توانایی‌های جدیدی نظیر امکان بررسی الگوها و روابط بین داده‌ها از میان مجموعه داده‌های پیچیده اهمیت می‌یابند. سرانجام، محققان با الهام گرفتن از دانش زیست

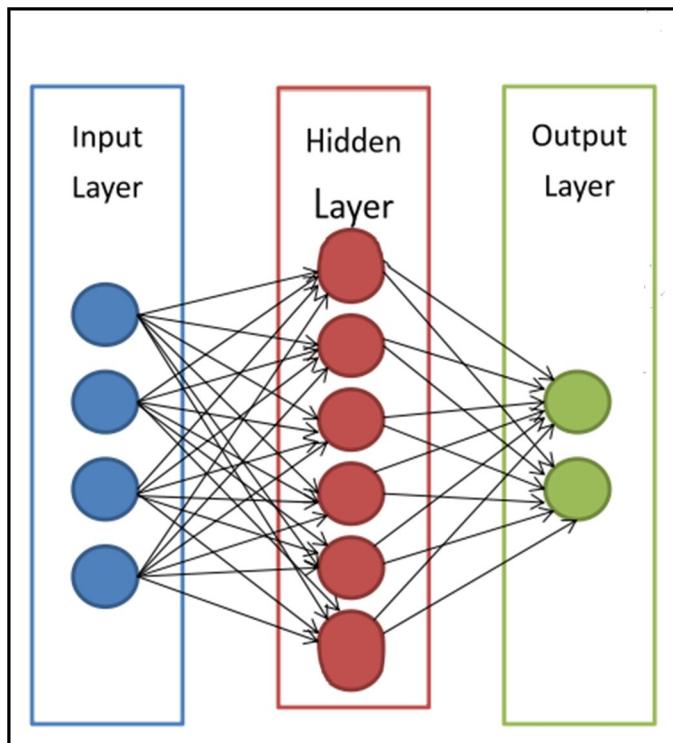
(Unsupervised) است. در روش اول، به سیستم یک مجموعه داده‌ی نشانه‌گذاری شده برای هدایت داده‌های ورودی به سمت خروجی مدنظر داده می‌شود؛ اما در روش دوم هیچ نمونه‌ی نشانه‌گذاری شده ارائه نمی‌شود، در طول فرایند یادگیری هیچ تصویری از داده‌های خروجی وجود ندارد و در پایان داده‌ها برآسانس ویژگی‌های مختلف از هم متمایز می‌شوند. هنگام استفاده از روش ML، هر نمونه با چندین ویژگی و هر ویژگی با انواع مختلفی از مقدار توصیف می‌شود. طی دوره‌ی گذشته، انواعی از تکنیک‌های ML و الگوریتم‌های انتخاب ویژگی برای پیش‌بینی بیماری و برجسته نمودن مهم‌ترین ویژگی‌های تاثیرگذار در وقوع و پیشرفت آن به کاررفته‌اند. بیشتر این مطالعات بر روی طبقه‌بندی انواع تومور (خوش‌خیم یا بدخیم)، مدل‌سازی پیشرفت و شناسایی عوامل موثر در وقوع هر نوع سرطان (از میان مجموعه داده‌ها) متمرکز شده است. شکل ۲ میزان توزیع مقالات چاپ شده طی ۵ سال اخیر در این حوزه را نشان می‌دهد.



شکل ۲: نمودار مربوط به میزان توزیع مقالات چاپ شده در زمینه کاربرد روش‌های ML در پیش‌بینی‌های مربوط به سرطان (۲۰۱۰-۲۰۱۴)

فرایند یادگیری در انسان با استفاده از تجربیات شخصی و در سیستم‌های ماشینی با دنبال کردن دستورالعمل‌های داده شده به سیستم، محقق خواهد شد.

را در بر می‌گیرد. این موضوع خود پایه‌ای در رویکرد درمان فردی (Personalized Medicine) است که تکنیک‌های محاسباتی با امکان شناسایی کم‌هزینه، موثر و دقیقی که فراهم می‌کنند، می‌توانند در چنین گروه‌های کوچکی از بیماران مورد استفاده قرار گیرند.



شکل ۳: تصویر شماتیکی از یک شبکه عصبی مصنوعی در روش ML

هرچه تعداد لایه‌های پنهان بیشتر باشد بدین معنی است که وزن‌های بیشتری را بین لایه‌ی ورودی تا لایه‌ی خروجی خواهیم داشت و ما می‌توانیم رفتار پیچیده‌تری را مدل‌سازی کنیم.

با ظهور و توسعه‌ی تکنولوژی‌های جدید، حجم زیادی از داده‌های ژنتیکی، بافت‌شناسی، جمعیت شناسی و کلینیکی در ارتباط با سرطان جمع‌آوری شده است که لزوم بررسی اثر هر یک از عوامل در ارتباط با هم و انتخاب موثرترین عوامل را شدت می‌بخشد.

شناسی و ساختار نورون‌های مغز انسان به سمت طراحی و استفاده از شبکه‌های عصبی مصنوعی (ANN یا Artificial Neural Network) چهت یافتد که پرکاربردترین تکنیک از انواع روش‌های ML است و نزدیک به سهدهه است که در مطالعات مربوط به پیش‌بینی سرطان مورد استفاده قرار می‌گیرد. شکل ۳ تصویر شماتیکی از یک شبکه‌ی عصبی مصنوعی و لایه‌های موجود در آن را نشان می‌دهد. لایه‌ی ورودی (Input Layer) دربرگیرنده‌ی هر اطلاعات خامی از مجموعه داده‌های است که به سیستم وارد می‌شود. در لایه‌ی پنهان (Hidden Layer)، هر یک از داده‌های خام در وزن مشخص ضرب شده و الگوریتم‌های ابزار ML بر روی آن اعمال می‌شوند. هرچه تعداد لایه‌های پنهان بیشتر باشد بدین معنی است که وزن‌های بیشتری را بین لایه‌ی ورودی تا لایه‌ی خروجی خواهیم داشت و ما می‌توانیم رفتار پیچیده‌تری را مدل‌سازی کنیم. در لایه‌ی خروجی نتیجه را خواهیم داشت که با توجه به مدل‌سازی انجام شده براساس داده‌های ورودی پیش‌بینی می‌کند که نمونه‌ی تهیه شده از فرد دارای سرطان است یا خیر، نوع سرطان چیست، در چه مرحله‌ای قرار دارد و یا چه پاسخی نسبت به داروی به کاربرده شده نشان می‌دهد. همچنین امروزه بر پایه‌ی نقص‌های ژنتیکی منحصر به فرد، زیرگروه‌های جداگانه‌ای در میان یک نوع سرطان وجود دارد که نتایج بالینی متفاوت و رویکردهای درمانی مختلفی



نتیجه‌گیری

همچنین با استفاده از ابزار ML، در ابتدا سیستم یکبار براساس اطلاعات ورودی و الگوی موجود در مجموعه داده‌ها راهاندازی می‌شود و به دنبال آن می‌توان نمونه‌ی جدیدی با صرف هزینه و زمان کمتر، در بررسی‌های احتمال ابتلا به سرطان، نوع سرطان، میزان پیشرفت و پیش‌بینی‌های درمانی مورد ارزیابی قرارداد.

اکثر مطالعات مربوط به سرطان از انواع مختلف داده‌های ورودی که شامل اطلاعات ژنتیکی، کلینیکی، بافت‌شناسی، تصویربرداری و اپیدمیولوژی می‌شود، در بررسی احتمالات مربوط به سرطان بهره می‌برند. از آنجایی که بررسی اثر هریک از این عوامل به تنها یافته دقت کافی است و دامنه اطلاعات وسیعی را فراهم نمی‌کند لزوم بهره‌گیری از ابزار ML تقویت می‌یابد. این تکنیک‌ها امکانی را فراهم می‌کنند تا محققان بتوانند تمامی عوامل موثر در بروز بیماری را به طور همزمان مورد بررسی قرار دهند و از این میان موثرترین عوامل برگزیده شوند. همچنین با استفاده از ابزار ML، در ابتدا سیستم یکبار براساس اطلاعات ورودی و الگوی موجود در مجموعه داده‌ها راهاندازی می‌شود و به دنبال آن می‌توان نمونه‌ی جدیدی با صرف هزینه و زمان کمتر، در بررسی‌های احتمال ابتلا به سرطان، نوع سرطان، میزان پیشرفت و پیش‌بینی‌های درمانی مورد ارزیابی قرارداد. بر این اساس دقت و سهولت در نتایج مربوط به پیش‌بینی و مدل‌سازی سرطان با استفاده از تکنیک ML به میزان ۱۵-۲۰ درصد افزایش یافته است و همین عامل سبب شده است که به عنوان ابزاری محبوب در تحقیقات پزشکی مورد استفاده قرار بگیرد. در مطالعات آینده سرطان می‌توان بر روی توسعه‌ی ابزارهای ML با بهره‌گیری از الگوریتم‌های جدید متمرکز شد که می‌توانند از اطلاعات ورودی ساده، میلیون‌ها پردازش اطلاعاتی انجام دهند و میزان نیاز به استفاده از تصویربرداری‌های گران‌قیمت و اقدامات آزمایشگاهی را به حداقل برسانند.