

## کاربرد هوش مصنوعی در تشخیص سلول سرطانی در جریان خون

**چکیده:** تشخیص سلول های سرطانی در جریان خون، به عنوان یک نشان گر زیستی ارزشمند کم تهاجمی انقلابی در مدیریت سرطان ایجاد کرده است. این مقاله به مرور جامع کاربردهای هوش مصنوعی شامل یادگیری ماشین و یادگیری عمیق در شناسایی و شمارش خودکار سلول های سرطانی در جریان خون از نمونه های خونی واقعی می پردازد. با بررسی سیستماتیک مطالعات پیشین، مزایا مانند سرعت و دقت بالا و معایب مانند وابستگی شدید به حجم و کیفیت داده های آموزشی در هر روش تحلیل شده است. یافته ها نشان می دهند که عملکرد مدل های هوش مصنوعی به طور قابل توجهی به کیفیت تصاویر وابسته است و همچنین فرآیند آموزش آنها اغلب زمان بر می باشد. برای غلبه بر این چالش، راهکارهایی مانند پیش پردازش داده، تکنیک های تقویت مصنوعی داده و استفاده از حلقه های بازخوردی پیشنهاد می شوند. با وجود نتایج امیدوار کننده ای که از پژوهش های انجام شده پیشین بدست آمده است برای کاربردی کردن این روش در محیط های بالینی واقعی، نیاز مبرمی به توسعه الگوریتم های عمومی تر، قوی تر و کمتر وابسته به داده های حجیم به عنوان نیاز اصلی این حوزه تحقیقاتی باقی است.

**واژه های راهنما:** سلول سرطانی در جریان خون، نشان گر زیستی، خون، یادگیری ماشین، یادگیری عمیق

هاجر مقدس\*  
استادیار

حمید پاکباز

دانشجوی کارشناسی ارشد،  
گروه مهندسی مکانیک،  
دانشگاه یاسوج، یاسوج

مقاله مروری

دریافت: ۱۴۰۴/۰۴/۲۱

بازنگری: ۱۴۰۴/۰۶/۱۶

پذیرش: ۱۴۰۴/۰۸/۱۹

Hajar Moghadas\*  
Assistant Professor

Hamid Pakbaz  
MSc Student,  
Faculty of Mechanical  
Engineering,  
Yasouj University, Yasouj

## Application of artificial intelligence in the detection of circulating tumor cell

**Abstract:** The detection of circulating cancer cells in the blood as a valuable minimally invasive biomarker has revolutionized cancer management. This paper provides a comprehensive review of the applications of artificial intelligence in the automatic identification and enumeration of circulating cancer cells in the real blood samples. By reviewing previous studies, the advantages, such as high speed and accuracy, and the disadvantages, such as strong dependence on the volume and quality of training data in each method, are analyzed. The findings show that the performance of artificial intelligence models is significantly dependent on the quality of images, and also their training process is often time-consuming. To overcome this challenge, solutions such as data preprocessing, artificial data augmentation techniques, and the use of feedback loops are proposed. Despite the promising results obtained from previous research to apply this method in real clinical settings, there remains an urgent need to develop more general, robust, and less dependent on large data sets as the main need of this research area.

**Keywords:** Circulating tumor cell, Biomarker, Blood, Machine learning, Deep learning

## ۱- مقدمه

نمونه‌های واقعی خون استفاده کرده‌اند. در ادامه، ساختارهای عمومی و مبانی هوش مصنوعی به کار رفته در این حوزه توضیح داده خواهد شد. سپس، مزایا، محدودیت‌ها و چالش‌های هر یک از روش‌های مبتنی بر هوش مصنوعی بررسی می‌شود.

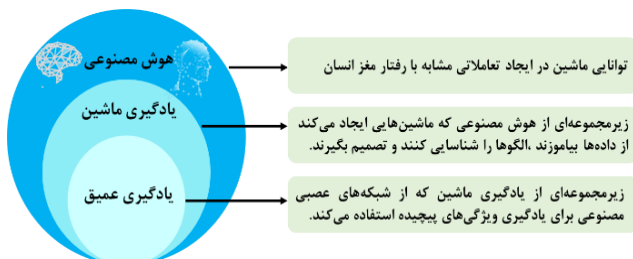
## ۲- هوش مصنوعی

هوش مصنوعی به عنوان یک فناوری نوین و رو به گسترش، کاربردهای وسیعی در تمامی حوزه‌های علمی پیدا کرده است [۱۶-۱۷]. این فناوری با خودکارسازی فرآیندها منجر به کاهش چشمگیر هزینه‌های انسانی و رفع چالش‌های وابسته به آن شده است [۱۸-۱۹].

در حوزه پزشکی، هوش مصنوعی در تشخیص انواع بیماری‌ها از جمله سرطان پستان [۲۰]، دیابت [۲۱]، بیماری‌های قلبی-عروقی [۲۲]، بیماری‌های ریوی [۲۳، ۲۴]، بیماری روده بزرگ [۲۵] و بیماری مزمن کلیه [۲۶] به نتایج قابل قبولی دست یافته است.

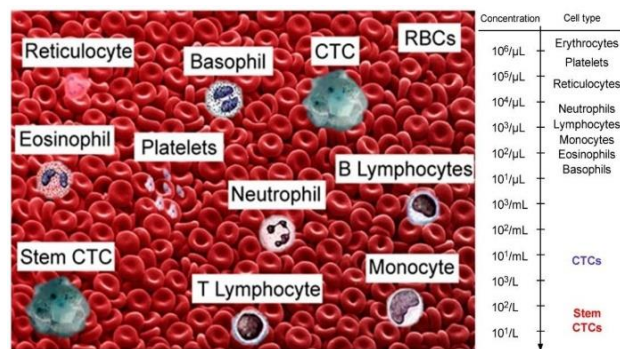
هوش مصنوعی شاخه‌ای از علوم کامپیوتر است که به ماشین آموزش می‌دهد مشابه هوش انسانی عمل کند. در این حوزه، مفاهیم یادگیری ماشین<sup>۴</sup> و یادگیری عمیق<sup>۵</sup> به عنوان زیرشاخه‌های اصلی مطرح می‌شوند. یادگیری ماشین، زیرمجموعه‌ای از هوش مصنوعی است که با استفاده از داده‌ها، توانایی یادگیری خودکار را برای سیستم فراهم می‌کند. یادگیری عمیق نیز به عنوان پیشرفته‌ترین شاخه یادگیری ماشین از شبکه‌های عصبی مصنوعی با معماری عمیق برای پردازش و یادگیری از داده‌های پیچیده استفاده می‌نماید.

سلسله مراتب و رابطه بین مفاهیم هوش مصنوعی، یادگیری ماشین و یادگیری عمیق در شکل (۲) نشان داده شده است. این مفاهیم در ادامه به صورت جزئی‌تر توضیح داده خواهد شد.



شکل ۲ سلسله مراتب و رابطه بین هوش مصنوعی، یادگیری ماشین و یادگیری عمیق

سلول‌های سرطانی در گردش خون<sup>۱</sup> (CTCs) هستند که از تومور اصلی جدا شده و وارد جریان خون می‌شوند. تحقیقات پیشین نشان می‌دهد که CTCs می‌توانند به عنوان یک نشانگر زیستی<sup>۲</sup> موثر برای تشخیص زودهنگام سرطان، پیشرفت بیماری و ارزیابی اثربخشی درمان به کار روند [۱-۲]. در شکل (۱) نسبت اندازه ذرات مختلف خون و CTC به همراه تعداد به واحد حجم آنها نشان داده شده است. همان‌گونه که مشاهده می‌شود CTCها در میان انبوهی از گلبول‌های قرمز و سفید و سایر اجزای خون محصور شده‌اند. تعداد این سلول‌ها در مقایسه با سایر یاخته‌های خونی به طور قابل توجهی کمتر است. این مسئله شناسایی و شمارش آن‌ها را به یک چالش عمده تبدیل کرده است [۳-۴].



شکل ۱ نسبت اندازه ذرات مختلف اجزای موجود در خون و CTC به همراه تعداد به واحد حجم متناظر هر ذره [۵]

با توجه به اهمیت بالینی CTCs، مطالعات گسترده‌ای در این زمینه انجام و روش‌های متنوعی برای جداسازی آن‌ها توسعه داده شده است. این روش‌ها شامل استفاده از آنتی‌بادی‌های اختصاصی متصل‌شونده به CTCs [۶]، تفکیک بر اساس اندازه [۷]، چگالی [۸]، چسبندگی [۹]، خواص آکوستیک [۱۰] و اصول هیدرودینامیک [۱۱-۱۲] هستند. اخیراً هوش مصنوعی<sup>۳</sup> به عنوان یک فناوری نوظهور، توجه بسیاری از محققان را به خود جلب کرده است. این فناوری پتانسیل آن را دارد که شناسایی و شمارش CTCs را با دقت و سرعت قابل توجهی و به صورت کاملاً خودکار انجام دهد [۱۳-۱۵].

هدف از این مقاله ارائه یک مرور ساختاریافته بر مطالعاتی است که از هوش مصنوعی برای تشخیص و شمارش CTCs در

<sup>4</sup> Machine learning

<sup>5</sup> Deep learning

<sup>1</sup> Circulating Tumor Cells

<sup>2</sup> Biomarker

<sup>3</sup> Artificial Intelligence

## ۲-۱- یادگیری ماشین

یادگیری عمیق زیرمجموعه‌ای از یادگیری ماشین است که از شبکه عصبی<sup>۲</sup> چندلایه برای شناسایی خودکار الگو در داده‌ها استفاده می‌کند [۲۷-۲۸]. یکی از پرکاربردترین ساختارها در این حوزه، شبکه عصبی کانولوشنی<sup>۳</sup> (CNN) است که به‌طور خاص برای داده‌های تصویری طراحی شده است [۳۰-۳۳]. شبکه‌های عصبی معمولاً از یک لایه ورودی، یک یا چند لایه پنهان و یک لایه خروجی تشکیل شده‌اند. در لایه ورودی، داده‌ها به مدل ارائه می‌شوند. لایه‌های پنهان، شامل نورون‌هایی هستند که عملیات وزن‌دهی و محاسبات بر روی ورودی‌ها را انجام می‌دهند. در نهایت، لایه خروجی، نتیجه نهایی را برای دسته‌بندی تولید می‌کند. نمونه‌ای از یک ساختار CNN پنجاه لایه‌ای برای یادگیری انتقالی و طبقه‌بندی در شکل (۳) نشان داده شده است [۳۴].

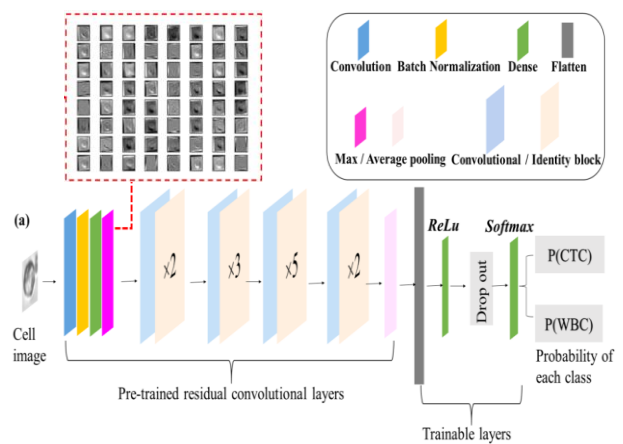
## ۲-۳- الگوریتم‌های هوش مصنوعی برای تشخیص CTC

به‌طور کلی هوش مصنوعی الگوریتم‌های زیاد و متنوعی دارد. در این بخش از مقاله صرفاً الگوریتم‌هایی معرفی می‌شوند که در پژوهش‌های پیشین برای تشخیص و شمارش CTC بکار رفته‌اند. الگوریتم‌های مختلفی به‌صورت یادگیری نظارت‌شده<sup>۴</sup>، یادگیری بدون نظارت<sup>۵</sup> و یادگیری تقویتی<sup>۶</sup> برای تشخیص CTC به‌کار رفته‌اند که در شکل (۴) نشان داده شده‌اند. در یادگیری نظارت‌شده، مدل بر اساس مجموعه‌ای از داده‌های برچسب‌دار که تحت نظارت مستقیم انسان آماده شده است، آموزش می‌بیند [۳۵] ولی در یادگیری بدون نظارت، مدل سعی می‌کند بدون داشتن برچسب‌های از پیش تعیین‌شده، الگوها و ساختارهای نهفته در داده‌های غیرساختاریافته را کشف کند [۳۶]. یادگیری نظارت‌شده به دو دسته اصلی طبقه‌بندی<sup>۷</sup> و رگرسیون<sup>۸</sup> تقسیم می‌شود.

در طبقه‌بندی، ورودی‌ها به کلاس‌های از پیش تعیین‌شده تقسیم می‌شوند. الگوریتم‌های پرکاربرد در طبقه‌بندی عبارت از طبقه‌بند خطی<sup>۹</sup>، ماشین‌های بردار پشتیبان<sup>۱۰</sup>، درخت‌های تصمیم‌گیری<sup>۱۱</sup> و جنگل‌های تصادفی<sup>۱۲</sup> می‌باشند.

یادگیری ماشین زیرشاخه‌ای از هوش مصنوعی است که در آن، یک مدل کامپیوتری با استفاده از یک مجموعه داده آموزشی و به کارگیری الگوریتم‌های خاص، راهکار عملیاتی خود را تولید می‌کند [۱۵]. به‌عنوان مثال، با نمایش مجموعه‌ای از تصاویر برچسب‌دار<sup>۱</sup> (به عنوان سلول) به مدل، الگوریتم یاد می‌گیرد که الگوهای مرتبط با تشخیص کلاس «سلول» را شناسایی کند. سپس می‌تواند تصاویر جدید را بررسی کرده و پیش‌بینی کند که آیا در آن تصویر، سلولی وجود دارد یا خیر. معمولاً از ۸۰٪ داده‌های موجود برای آموزش مدل و از ۲۰٪ برای آزمون آن استفاده می‌شود. اگرچه فرآیند آموزش مدل، زمان‌بر است با این حال، مدل آموزش دیده، در مدت زمان کوتاهی قادر است داده‌های جدید را ارزیابی و دسته‌بندی کند. عملکرد مدل‌های موجود در تشخیص داده‌های درست متفاوت بوده و به عوامل مختلفی بستگی دارد که در ادامه توضیح داده خواهد شد.

## ۲-۲- یادگیری عمیق



شکل ۳ ساختار شبکه عمیق کانولوشنی پنجاه لایه‌ای برای تشخیص سلول‌های سرطانی در گردش و گلبول‌های سفید. لایه‌های مختلف که با رنگ‌های متفاوت مشخص شده‌اند، عملیات خاصی را انجام می‌دهند. در لایه‌های ابتدایی ویژگی‌های سلول استخراج می‌شود. هر لایه مربوط به بررسی یک ویژگی خاص است. در لایه‌های میانی شرایط CTC بودن سلول بررسی می‌شود. در لایه انتهایی احتمال CTC بودن سلول مشخص می‌شود (تصویر از مرجع [۳۴] با کمی تغییر، با مجوز بازنشر).

<sup>7</sup> Classification

<sup>8</sup> Regression

<sup>9</sup> Linear classifiers

<sup>10</sup> Support vector machines

<sup>11</sup> Decision trees

<sup>12</sup> Random forest

<sup>1</sup> Labeled

<sup>2</sup> Neural network

<sup>3</sup> Convolutional Neural Network

<sup>4</sup> Supervised learning

<sup>5</sup> Unsupervised learning

<sup>6</sup> Reinforcement learning

عملکرد مدل‌های مختلف معمولاً با پارامترهای دقت<sup>۴</sup>، حساسیت<sup>۵</sup> و ویژگی<sup>۶</sup> سنجیده می‌شود. نحوه محاسبه این پارامترها عبارت است از [۴۰]:

$$\text{دقت} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$\text{حساسیت} = \frac{TP}{TP + FN} \quad (2)$$

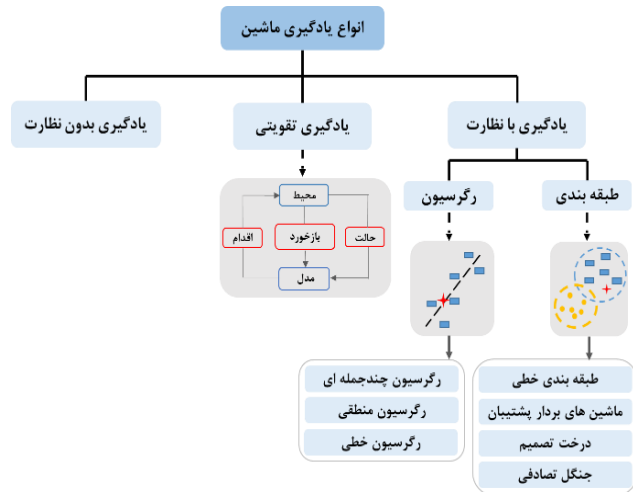
$$\text{ویژگی} = \frac{TN}{TN + FP} \quad (3)$$

در اینجا TP تعداد نمونه‌های مثبتی است که مدل به درستی آنها را مثبت تشخیص داده، TN تعداد نمونه‌های منفی است که مدل به درستی آنها را منفی شناسایی کرده، FP تعداد نمونه‌هایی است که مدل آنها را مثبت تشخیص داده اما در واقع منفی هستند، FN تعداد نمونه‌هایی است که مدل آنها را منفی تشخیص داده، اما در واقع مثبت هستند.

دقت معیاری برای بیان تعداد مثبت‌ها و منفی‌هایی که مدل به درستی تشخیص داده تقسیم به کل داده‌های موجود است. حساسیت نسبت تعداد موارد مثبت تشخیص داده شده توسط مدل تقسیم به کل مثبت‌های موجود است. به عبارتی حساسیت بیان می‌کند که مدل چند درصد از موارد مثبت را توانسته شناسایی کند. ویژگی بیانگر درصد نمونه‌های منفی واقعی است که مدل به درستی تشخیص داده است. معیار دیگر برای ارزیابی عملکرد مدل‌های طبقه‌بندی، منحنی ROC و مساحت زیر آن AUC<sup>۷</sup> است [۴۱]. این معیار، توانایی مدل در تمایز بین کلاس‌ها بدون در نظر گرفتن توزیع آنها نشان می‌دهد.

### ۳- پژوهش‌های مرتبط با هوش مصنوعی در تشخیص CTC

تا کنون مقالات متنوعی در مورد استفاده از هوش مصنوعی در تشخیص انواع سرطان‌ها با استفاده از داده‌های بالینی یا تصاویر پزشکی بافت ارائه شده است که به موضوع مقاله حاضر مرتبط نمی‌شود. زیرا تمرکز این مقاله بر کاربرد هوش مصنوعی در تشخیص CTCs است. به‌طور کلی محققان هوش مصنوعی برای شناسایی CTCs از دو دسته اطلاعات شامل تصاویر میکروسکوپی یا داده‌های ژنومیک استفاده کرده‌اند [۴۲-۴۳].



شکل ۴ انواع یادگیری ماشین و زیر مجموعه‌های آن که برای تشخیص CTC به کار رفته‌اند.

طبقه‌بند خطی، الگوریتم‌های ساده‌ای هستند که با ترسیم یک مرز تصمیم‌گیری خطی یا صفحه‌ای، داده‌ها را دسته‌بندی می‌کنند. روش ماشین‌های بردار پشتیبان مشابه طبقه‌بند خطی است، اما با این تفاوت که به حداکثر کردن حاشیه اطمینان<sup>۱</sup> بین دو کلاس می‌پردازد تا مدل در برابر داده‌های جدید مقاوم‌تر باشد. درخت‌های تصمیم‌گیری با استفاده از یک ساختار درختی و پرسیدن سؤالات سلسله‌مراتبی، داده‌ها را تقسیم‌بندی می‌کند. جنگل‌های تصادفی یک روش یادگیری گروهی<sup>۲</sup> است که از ترکیب چندین درخت تصمیم‌گیری برای بهبود دقت و کاهش overfitting استفاده می‌کند [۳۷].

در رگرسیون از رابطه بین متغیرهای وابسته و مستقل برای تقسیم‌بندی استفاده می‌شود. الگوریتم‌های پرکاربرد آن عبارت از رگرسیون خطی، رگرسیون منطقی<sup>۳</sup> و رگرسیون چند جمله‌ای می‌باشد [۳۸].

رگرسیون خطی برای مدل‌سازی از یک رابطه خطی بین یک متغیر وابسته و یک یا چند متغیر مستقل استفاده می‌کند. در رگرسیون منطقی از یک تابع منطقی برای تخمین احتمال تعلق به یک کلاس خاص استفاده می‌شود. در رگرسیون چند جمله‌ای از روابط غیرخطی برای دسته‌بندی بین متغیرها استفاده می‌شود. الگوریتم‌های دیگری نیز در یادگیری ماشین استفاده می‌شوند که از ترکیب الگوریتم‌های موجود یا تقویت آنها با استفاده از بازخورد سیستم برای بهبود عملکرد خود استفاده می‌کنند و به یادگیری تقویتی معروف هستند [۳۹].

<sup>5</sup> Sensitivity

<sup>6</sup> Specificity

<sup>7</sup> Area Under the Curve

<sup>1</sup> Margin

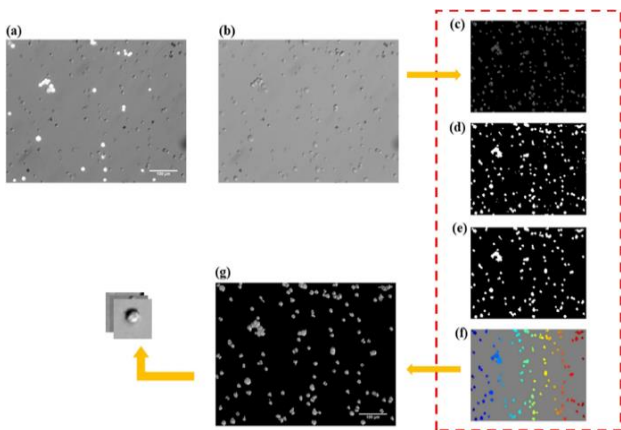
<sup>2</sup> Ensemble

<sup>3</sup> Logistic regression

<sup>4</sup> Accuracy

شد. نتایج مدل ترکیبی با حساسیت ۹۲/۸۷٪ و ویژگی ۹۹/۹۸٪ عملکرد بهتری نسبت به مدل‌های جداگانه داشت.

زیون و همکاران (۲۰۲۰) یک مدل CNN ترکیبی با روش‌های تجسم ارائه کردند که سرعت بالایی در تجزیه و تحلیل داده‌های بزرگ داشت [۴۷]. آنها توانستند ۲۳/۳۷۷ تصویر تک‌سلولی را در کمتر از ۳ دقیقه بر روی یک رایانه معمولی در پنج گروه مختلف CTCs، وژیکول‌های خارج سلولی، گلبول‌های سفید، هسته‌های برهنه و ذرات دیگر دسته‌بندی کنند. حساسیت و ویژگی آن مدل به ترتیب ۹۷/۳٪ و ۹۸/۸٪ گزارش شد. داده‌های آنها از ۴۹۹ بیمار سرطانی متاستاتیک استخراج شده بود [۴۸]. ونگ و همکاران (۲۰۲۰) در مدل خود تعداد لایه‌های شبکه عصبی را افزایش داده و یک شبکه عصبی ۵۰ لایه مدل ResNet را ایجاد کردند [۳۴]. آنها برای تهیه داده‌ها از میکروسکوپ میدان روشن با بزرگنمایی ۲۰ برابر بزرگ‌تر استفاده کردند. مراحل آماده‌سازی تصاویر آن در شکل (۵) نشان داده شده است. در مرحله پردازش اولیه، تصاویر فیلتر شدند. سپس فضاهای خالی را پر و بر اساس مشخصات مورفولوژی سلولی، نواحی غیرمرتبط را حذف کردند. در ادامه تصاویر را برش دادند به گونه‌ای که در هر تصویر فقط یک سلول مجزا وجود داشته باشد. در آن پژوهش همچنین از تکنیک چرخش و انعکاس تصاویر برای تولید داده مصنوعی استفاده شد. نمونه خونی از بیماران دارای سرطان کلیه تهیه شده بود. این مدل برای نمونه خون دستکاری شده با حذف گلبول‌های قرمز با دقت ۹۷٪ شناسایی را انجام داد. با این حال برای خون کامل با دقت ۸۸/۶٪ عملکرد مطلوبی در تشخیص CTCها نداشت.



شکل ۵ مراحل پیش‌پردازش تصویر داده‌های خام؛ (a) تصویر فلورسنت برچسب‌گذاری شده، (b) تصویر میدان روشن، (c) فیلتر کردن بر اساس مرز سلول، (d) عملیات انباشتن سیلابی تصویر، (e) اعمال فیلتر مورفولوژیکی برای شناسایی سلول و حذف نقاط غیرمرتبط، (f) تبدیل واترشد برای بخش‌بندی و (g) تصویر پردازش شده که برای تهیه تصاویر سلول‌های مجزا بکار می‌رود، با مجوز بازنشر [۳۴].

تمرکز این مقاله بر پژوهش‌های مرتبط به داده‌های تصویری است. تصاویر مورد استفاده به دو صورت تصاویر خون کامل دستکاری‌نشده (حاوی تمامی اجزای خون) و یا تصاویر خون دستکاری‌شده (مانند CTCهای کشت‌شده در محیط مصنوعی یا نمونه‌های خونی که گلبول‌های قرمز یا سفید آنها حذف شده‌اند) به کار گرفته شده‌اند. در این مرور، تنها مقالاتی که از تصاویر خون کامل دستکاری‌نشده استفاده کرده‌اند مورد بررسی قرار می‌گیرند.

شولتنز و همکاران (۲۰۱۲) را می‌توان از پیشگامان این عرصه دانست. آنها برای اولین بار با به کارگیری الگوریتم جنگل تصادفی در نرم‌افزار متلب، یک روش خودکار برای شمارش CTCs بر اساس تصاویر سیتومتری ارائه دادند [۴۴]. این مدل با استخراج ویژگی‌های بافت و مورفولوژی سلول، قادر به شناسایی و دسته‌بندی CTCها، CTCهای آپوپتوز، بقایای سلولی، لکوسیت‌ها و ذرات غیرمرتبط بود. داده‌های این مطالعه از ۶۸ بیمار سرطانی و ۹ فرد سالم جمع‌آوری شده بود. نمونه‌ها با استفاده از غنی‌سازی مغناطیسی و رنگ‌آمیزی فلورسنت (DAPI برای هسته، آنتی‌بادی سیتوکراتین برای CTCها و CD45 برای لکوسیت‌ها) آماده شدند. معیارهای شناسایی شامل قطر حداقل ۴ میکرومتر، مورفولوژی گرد تا بیضی، وجود هسته مشخص، مثبت بودن سیتوکراتین و منفی بودن CD45 بود.

موکان و همکاران (۲۰۱۸) با ارائه یک مدل U-Net اصلاح‌شده متشکل از یک شبکه کانولوشنی هفت لایه، دقت در تشخیص CTCs را افزایش دادند [۴۵]. آنها با استفاده از تکنیک‌های تقویت داده مانند جابجایی، مقیاس‌گذاری، چرخش و ورق زدن تصاویر، تعداد داده‌ها را از ۱۲۰ تصویر به ۵۶،۰۰۰ تصویر تک‌سلولی افزایش دادند. بدین ترتیب با افزایش تعداد داده‌های آموزشی مدل موفق به دستیابی به دقت قابل توجه ۹۹/۸٪ در تشخیص شد. این داده‌ها از نمونه خون سالم رنگ‌آمیزی‌نشده حاوی رده سلولی سرطان روده بزرگ با استفاده از میکروسکوب میدان تاریک تهیه شده بود.

سیورته و همکاران (۲۰۱۸) روش‌های تقویت‌کننده در CNN و الگوریتم Adaboost به همراه روش فیلتر تصاویر (بر اساس ویژگی سلول) را بکار بردند [۴۶]. آنها رده‌های سلول سرطان پستان Hs 578T و MCF7 و رده سلول سرطان روده بزرگ DLD-1 را به خون انسان سالم اضافه کردند. با استفاده از میکروسکوپ میدان تاریک ۲۶۳ تصویر تهیه کردند. تصاویر این نوع میکروسکوب شرایط ثبت ویژگی و مشخصه‌های سلول را فراهم می‌کند. آنها چهار مدل را با داده‌های مختلف و هدفمند آموزش دادند. سه مدل به‌طور مستقل با داده‌های مجزای سه رده سلولی آموزش دیدند. برای آموزش مدل چهارم، به صورت ترکیبی از تمام داده‌های سه رده سلولی به‌طور همزمان استفاده

نمی‌توان عملکرد آنها را به راحتی با یکدیگر مقایسه کرد. با این حال تمامی مدل‌های ارائه شده در مقایسه با روش‌های سنتی شمارش و دسته‌بندی CTC، سرعت و دقت بیشتری نشان دادند. به طور کلی افزودن مراحل پیش پرداز تصاویر، حذف نویزها و نواحی غیرمرتبط بر اساس ویژگی‌های مورفولوژیک و بافتی سلول، سرعت مدل‌ها را افزایش می‌دهد.

جدول ۱ خلاصه مطالعات پیشین مربوط به استفاده از هوش مصنوعی در تشخیص CTC شامل نوع مدل، عملکرد مدل و نمونه خونی

منبع	مدل	عملکرد مدل	نمونه خونی
۲۰۱۲ [۴۴]	جنگل تصادفی	دقت ۸۹٪	۶۸ بیمار و ۹ فرد سالم غنی‌سازی مغناطیسی رنگ‌آمیزی فلورسنت تصاویر سیتومتری
۲۰۱۸ [۴۵]	U-Net اصلاح شده با CNN هفت لایه	دقت ۹۹٪	۱۲۰ تصویر خون سالم رده سلولی سرطان روده میکروسکوپ میدان تاریک
۲۰۱۸ [۴۶]	CNN و الگوریتم Adaboost	حساسیت ۹۲٪ ویژگی ۹۹٪	۲۶۳ تصویر رده MCF7.Hs 578T و DLD-1 خون انسان سالم میکروسکوپ میدان
۲۰۲۰ [۴۷]	CNN ترکیبی با روش‌های تجسم	حساسیت ۹۷٪ ویژگی ۹۸٪	۲۳،۳۷۷ تصویر تک سلولی ۴۹۹ بیمار سرطانی متاستاتیک
۲۰۲۰ [۳۴]	CNN، ۵۰ لایه با ResNet	دقت ۹۷٪ خون دستکاری شده، دقت ۸۸٪ خون کامل	بیماران سرطان کلیه میکروسکوپ میدان روشن با بزرگنمایی ۲۰ برابر
۲۰۲۰ [۴۹]	CNN	دقت ۹۲٪ حساسیت ۹۳٪	۶۰۰ بیمار با انواع مختلف سرطان رنگ‌آمیزی سلولی
۲۰۲۱ [۴۰]	ماشین‌های بردار	دقت ۸۲٪ در داده‌های آموزشی دقت ۹۰٪ در داده‌های کور	۴۱ بیمار سرطانی روده بزرگ و ۴۱ فرد سالم
۲۰۲۲ [۵۰]	VGG16، VGG19، ResNet18، ResNet50 و AlexNet	دقت ۸۲٪ در VGG16 ۹۴٪ و حساسیت ۹۷٪ بهترین نتیجه	۶۹۴ تصویر دارای CTC و ۱۳۴۷۲ بدون CTC از ۷۷۶ بیمار و فرد سالم

هی و همکاران (۲۰۲۰) برای افزایش دقت مدل CNN تعداد نمونه‌های بیشتر و متنوع‌تری را به کار گرفتند [۴۹]. آنها مدل خود را با انواع مختلف سلول‌های سرطانی که از نمونه خونی ۶۰۰ بیمار بدست آمده بود، آموزش دادند. بعلاوه برای بهبود تشخیص مدل، آنها قسمت‌های مختلف سلول‌ها (هسته و سیتوپلاست) را با رنگ‌های متمایز، رنگ‌آمیزی کردند. با این حال مدل آنها با حساسیت ۹۳٪، توانست به دقت ۹۲٪ دست یابد. این مدل بهترین عملکرد را در تشخیص سلول‌های ریوی داشت که بیشترین تعداد در داده‌های آموزشی را داشتند.

هاتزیداکا و همکاران (۲۰۲۱) یک مدل ماشین‌های بردار پشتیبان را ایجاد دادند [۴۰]. نمونه‌های خونی را از ۴۱ بیمار سرطانی روده بزرگ و ۴۱ فرد سالم تهیه کردند. آنها ۳۱ نمونه را برای آموزش مدل و ۱۰ نمونه دیگر را به صورت داده کور بکار بردند. داده کور تصاویری هستند که در مرحله آموزش به ماشین نمایش داده نمی‌شوند و ماشین بر اساس داده برداری از نمونه‌های آموزشی، نمونه‌های کور را که قبلاً ندیده است، ارزیابی می‌کند. دقت آن مدل برای داده‌های آموزشی ۸۲٪ و برای داده‌های کور ۹۰٪ گزارش شد.

گوا و همکاران (۲۰۲۲) در یک پژوهش قابلیت مدل‌های مختلف AlexNet، ResNet50، ResNet18، VGG19، VGG16 را با یکدیگر مقایسه کردند [۵۰]. نمونه‌ها شامل ۶۹۴ تصویر دارای CTC و ۱۳۴۷۲ بدون CTC بود که از ۷۷۶ بیمار و فرد سالم بدست آمده بود. آنها تکنیک‌های افزایش و تنوع تعداد تصاویر آموزشی، حذف نویز، برش و افزایش کیفیت تصاویر و اعمال ویژگی‌های سلول را به کار بردند. در نهایت از بین این مدل‌ها، VGG16 با دقت ۹۴٪ و حساسیت ۹۷٪ بهترین نتیجه را داشت.

خلاصه مطالعات پیشین شامل نوع مدل، عملکرد مدل و نمونه خونی به ترتیب سال در جدول (۱) نشان داده شده است. بررسی مطالعات پیشین نشان می‌دهد که عملکرد همگی مدل‌های هوش مصنوعی به تعداد و کیفیت تصاویر آموزشی بستگی دارد. هر چه تعداد داده‌های آموزشی بیشتر باشد، حساسیت و دقت مدل بهتر می‌شود. محققان برای افزایش تعداد داده‌های آموزشی از دو راهکار افزایش تعداد بیماران برای نمونه برداری و یا روش‌های تولید داده مصنوعی مانند جابجایی، مقیاس گذاری، چرخش و ورق زدن تصادفی عکس‌ها استفاده می‌کنند. در تصاویر با کیفیت، تفکیک و مرزبندی انواع گونه‌ها برای ماشین راحت‌تر و سریع‌تر است. از آنجایی که مدل‌های ارائه شده در تحقیقات پیشین بر روی داده‌های مختلف با شرایط آماده‌سازی متفاوت و انواع سلول‌های متنوع انجام شده‌اند،

استفاده از مدل‌های سریع در پردازش تصاویر پیشنهاد می‌شود. علاوه بر این، چالش ارتقاء عملکرد مدل‌ها در تشخیص داده‌های کور همچنان نیاز به تحقیقات بیشتر برای کاربردی شدن این مدل‌ها در کلینیک دارد.

#### ۵- مراجع

- [1] R. Lawrence, M. Watters, C. R. Davies, K. Pantel, and Y.-J. Lu, "Circulating tumour cells for early detection of clinically relevant cancer," *Nature Reviews Clinical Oncology*, vol. 20, no. 7, pp. 487-500, 2023, <https://doi.org/10.1038/s41571-023>.
- [2] H. Yao, L. Wen, Z. Li, and C. Xia, "Analysis of diagnostic value of CTC and CTDNA in early lung cancer," *Cellular and Molecular Biology*, vol. 69, no. 6, pp. 57-62, 2023, <https://doi.org/10.14715/cmb>.
- [3] A. Strati, A. Markou, E. Kyriakopoulou, and E. Lianidou, "Detection and molecular characterization of circulating tumour cells: challenges for the clinical setting," *Cancers*, vol. 15, no. 7, p. 2185, 2023, <https://doi.org/10.3390/cancers15072185>.
- [4] M. Vidlarova *et al.*, "Recent advances in methods for circulating tumor cell detection," *International Journal of Molecular Sciences*, vol. 24, no. 4, p. 3902, 2023, <https://doi.org/10.3390/ijms24043902>.
- [5] Galanzha, E.I.; Zharov, V.P. Circulating Tumor Cell Detection and Capture by Photoacoustic Flow Cytometry in Vivo and ex Vivo. *Cancers*, 5, 1691-1738, 2013, <https://doi.org/10.3390/cancers5041691>.
- [6] T. N. A. Nguyen, P.-S. Huang, P.-Y. Chu, C.-H. Hsieh, and M.-H. Wu, "Recent progress in enhanced cancer diagnosis, prognosis, and monitoring using a combined analysis of the number of circulating tumor cells (CTCs) and other clinical parameters," *Cancers*, vol. 15, no. 22, p. 5372, 2023, <https://doi.org/10.3390/cancers15225372>.
- [7] M. Li, J. Shi, Y. Zhang, S. Cui, L. Zhang, and Q. Shen, "ECL cytosensor for sensitive and label-free detection of circulating tumor cells based on hierarchical flower-like gold microstructures," *Analytica Chimica Acta*, vol. 1303, p. 342505, 2024, <https://doi.org/10.1016/j.aca.2024.342505>.
- [8] J. Bialek, A. Muthe, S. Yankulov, F. Kawan, G. Gakis, and G. Theil, "Optimizing CTC isolation techniques for molecular characterization of circulating tumor cells in clear cell renal cell carcinoma: A comparative study of EpCAM-based and density-based methods," *Cancer Research*, vol. 84, no. 6\_Supplement, pp. 3689-3689, 2024, <https://doi.org/10.1158/1538-7445.AM2024-3689>.

تحقیقات نشان می‌دهد که با حذف گلبول‌های قرمز و سفید از نمونه خونی، شانس شناسایی سلول‌های سرطانی برای مدل هوش مصنوعی افزایش می‌یابد. این درحالی است که استفاده از لایزکننده گلبول قرمز و فرآیند جداسازی گلبول‌های سفید می‌تواند باعث ترکیدن و از بین رفتن CTC‌های موجود در نمونه شوند. در این حالت تعداد CTC‌های شمارش شده با تعداد CTC‌های واقعی موجود در نمونه اصلی، مطابقت نخواهد داشت. در واقع دستکاری نمونه‌های خونی منجر به خطا در نتایج می‌شود. با توجه به تعداد کم CTC‌ها اعمال روش‌هایی که منجر به حذف ناخواسته آنها می‌شود قابل توجه نمی‌باشد. آنچه مطلوب است ارائه روشی سریع و دقیق برای شناسایی CTC در خون واقعی بدون حذف اجزای مختلف خون و رنگ‌آمیزی است. حضور اجزای مختلف خون در تصاویر، شناسایی CTC را چالشی می‌کند. ذرات اضافی باعث ایجاد نویز در تصاویر می‌شوند. همچنین دانسیته بالای ذرات خونی شناسایی مرز CTC را سخت می‌کند. به نظر می‌رسد که استفاده از یک سیال رقیق‌کننده بتواند مشکل تعداد ذرات بر واحد سطح تصویر را کاهش دهد. افزودن رقیق‌کننده به نمونه خونی منجر به فاصله گرفتن ذرات از یکدیگر می‌شود و از این رو شناسایی دقیق‌تر مرز ذرات راحت‌تر می‌شود. البته این پیشنهاد منجر به افزایش حجم نمونه مورد آزمایش و افزایش تعداد تصاویر می‌شود. این روش برای مدل‌های سریع که مشکلی با پردازش داده‌های حجیم ندارند، مناسب است.

#### ۴- نتیجه‌گیری

در این مقاله کاربردهای روش یادگیری ماشین و یادگیری عمیق در تشخیص سلول‌های سرطانی در گردش در خون بررسی شد. هرکدام از این روش‌ها با استفاده از الگوریتم‌های متنوع قابلیت‌های متفاوتی در تشخیص سلول‌های هدف نشان دادند. با این حال همگی آنها در مقایسه با روش‌های سنتی، عملکرد بهتری داشتند. به‌طور کلی دقت این روش‌ها با افزایش تعداد داده‌های آموزشی و ارتقاء کیفیت تصاویر افزایش می‌یابد. همچنین اعمال پیش‌پردازش و حذف داده‌های اضافی نقش به‌سزایی در افزایش سرعت و دقت مدل‌ها دارد. از میان انواع الگوریتم‌هایی که در این حوزه بکار گرفته شدند، مدل‌های تقویت‌شده و مدل‌های ترکیبی نتایج بهتری را ارائه دادند. یافته‌ها نشان می‌دهد که هوش مصنوعی می‌تواند نقش مؤثری در حل چالش‌های تشخیص و شمارش CTCs داشته باشد. به‌عنوان تحقیقات آتی طراحی مدل‌های که با داده‌های محدود نیز قادر به یادگیری باشند و یا

- [20] B. S. Abunasser, M. R. J. AL-Hiealy, I. S. Zaqout, and S. S. Abu-Naser, "Literature review of breast cancer detection using machine learning algorithms," in *AIP Conference Proceedings*, 2023, vol. 2808, no. 1: AIP Publishing, <https://doi.org/10.1063/5.01336>.
- [21] B. Y. Kasula, "Machine Learning Applications in Diabetic Healthcare: A Comprehensive Analysis and Predictive Modeling," *International Numeric Journal of Machine Learning and Robots*, vol. 7, no. 7, 2023, [https://researchgate.net/Machine\\_Learning](https://researchgate.net/Machine_Learning).
- [22] E. K. Oikonomou and R. Khera, "Machine learning in precision diabetes care and cardiovascular risk prediction," *Cardiovascular Diabetology*, vol. 22, no. 1, p. 259, 2023, <https://10.1186/s12933-023>.
- [23] S. N. A. Shah and R. Parveen, "An extensive review on lung cancer diagnosis using machine learning techniques on radiological data: state-of-the-art and perspectives," *Archives of Computational Methods in Engineering*, vol. 30, no. 8, pp. 4917-4930, 2023, <https://10.1007/s11831-023>.
- [24] Q. Gao, L. Yang, M. Lu, R. Jin, H. Ye, and T. Ma, "The artificial intelligence and machine learning in lung cancer immunotherapy," *Journal of Hematology & Oncology*, vol. 16, no. 1, p. 55, 2023, <https://10.1186/s13045-023-01456-y>.
- [25] L. Chen *et al.*, "Machine Learning Predicts Oxaliplatin Benefit in Early Colon Cancer," *Journal of Clinical Oncology*, vol. 42, no. 13, pp. 1520-1530, 2024, <https://doi.org/10.1200/JCO.23.01080>.
- [26] M. A. Islam, M. Z. H. Majumder, and M. A. Hussein, "Chronic kidney disease prediction based on machine learning algorithms," *Journal of pathology informatics*, vol. 14, p. 100189, 2023, <https://doi.org/10.1016/j.jpi.2023.100189>.
- [27] S. F. Ahmed *et al.*, "Deep learning modelling techniques: current progress, applications, advantages, and challenges," *Artificial Intelligence Review*, vol. 56, no. 11, pp. 13521-13617, 2023, <https://10.1007/S10462-023-10466-8>.
- [28] C. Shen *et al.*, "Automatic detection of circulating tumor cells and cancer associated fibroblasts using deep learning," *Scientific reports*, vol. 13, no. 1, p. 5708, 2023, <https://s41598-023-32955>.
- [29] Y. Kumar, R. Singh, M. R. Moudgil, and Kamini, "A systematic review of different categories of plant disease detection using deep learning-based approaches," *Archives of Computational Methods in Engineering*, vol. 30, no. 8, pp. 4757-4779, 2023, <https://10.1007/s11831-023-09958-1>.
- [30] W. Duan and B. Ren, "Application Effects of NNN-link Care Model in Patients with Coronary Heart
- [9] X. Ye *et al.*, "An adhesion-based method for rapid and low-cost isolation of circulating tumor cells," *Clinica Chimica Acta*, vol. 547, p. 117421, 2023, <https://doi.org/10.1016/j.cca.2023.117421>.
- [10] Y. Zhang, Z. Zhang, D. Zheng, T. Huang, Q. Fu, and Y. Liu, "Label-free separation of circulating tumor cells and clusters by alternating frequency acoustic field in a microfluidic chip," *International Journal of Molecular Sciences*, vol. 24, no. 4, p. 3338, 2023, <https://doi.org/10.1016/j.cca.2023.117421>.
- [11] Y. Li *et al.*, "Label-free detection and simultaneous viability determination of CTCs by lens-free imaging cytometry," *Analytical and Bioanalytical Chemistry*, pp. 1-13, 2024, DOI: [10.1007/s00216-024-05624-y](https://doi.org/10.1007/s00216-024-05624-y).
- [12] S. Zhu *et al.*, "Liquid Biopsy Instrument for Ultra-Fast and Label-Free Detection of Circulating Tumor Cells," *Research*, vol. 7, p. 0431, 2024, <https://doi.org/10.34133/research.0431>.
- [13] M. J. Poellmann *et al.*, "Nanotechnology and machine learning enable circulating tumor cells as a reliable biomarker for radiotherapy responses of gastrointestinal cancer patients," *Biosensors and Bioelectronics*, vol. 226, p. 115117, 2023, <https://doi.org/10.1016/j.bios.2023.115117>.
- [14] K. Pastuszak *et al.*, "Detection of circulating tumor cells by means of machine learning using Smart-Seq2 sequencing," *Scientific Reports*, vol. 14, no. 1, p. 11057, 2024, <http://dx.doi.org/10.1038/s41598-024>.
- [15] J. Ma *et al.*, "Artificial intelligence based on blood biomarkers including CTCs predicts outcomes in epithelial ovarian cancer: A prospective study," *OncoTargets and therapy*, pp. 3267-3280, 2021, <https://doi.org/10.2147/OTT.S307546>.
- [16] K. Sharifani and M. Amini, "Machine learning and deep learning: A review of methods and applications," *World Information Technology and Engineering Journal*, vol. 10, no. 07, pp. 3897-3904, 2023, [https://papers.ssrn.com/sol3/abs\\_id=4458723](https://papers.ssrn.com/sol3/abs_id=4458723).
- [17] M. M. Taye, "Understanding of machine learning with deep learning: architectures, workflow, applications and future directions," *Computers*, vol. 12, no. 5, p. 91, 2023, <https://doi.org/10.3390/compu>.
- [18] D. Sengupta and L. Kumar, "Machine learning approaches in cancer detection and treatment," *IIIT-Delhi*, 2023, <http://repository.iiitd.edu.in/xmlui/han>.
- [19] R. Rejuan *et al.*, "Validation of a Microfluidic Device Prototype for Cancer Detection and Identification: Circulating Tumor Cells Classification Based on Cell Trajectory Analysis Leveraging Cell-Based Modeling and Machine Learning," *bioRxiv*, 2024, <https://doi.org/10.1002/cnm.70037>.

- on circulating tumor cells and machine learning," *Entropy*, vol. 23, no. 10, p. 1248, 2021, <https://doi.org/10.3390/s22103776>.
- [41] H. Li *et al.*, "Digital Quantitative Detection for Heterogeneous Protein and mRNA Expression Patterns in Circulating Tumor Cells," *Advanced Science*, p. 2410120, 2025, <https://doi.org/10.1002>.
- [42] P. Chen *et al.*, "Detection of circulating plasma cells in peripheral blood using deep learning-based morphological analysis," *Cancer*, vol. 130, no. 10, pp. 1884-1893, 2024, <https://doi.org/10.1002/cncr.352>.
- [43] L. Gerratana *et al.*, "Integrating machine learning-predicted circulating tumor cells (CTCs) and circulating tumor DNA (ctDNA) in metastatic breast cancer: A proof of principle study on endocrine resistance profiling," *Cancer Letters*, vol. 609, p. 217325, 2025, <https://doi.org/10.1016/j.canlet.2024>.
- [44] T. M. Scholtens, F. Schreuder, S. T. Ligthart, J. F. Swennenhuis, J. Greve, and L. W. Terstappen, "Automated identification of circulating tumor cells by image cytometry," *Cytometry Part A*, vol. 81, no. 2, pp. 138-148, 2012, <https://doi.org/10.1002/cyto.a>.
- [45] I. Mocan, R. Itu, A. Ciurte, R. Danescu, and R. Buiga, "Automatic Detection of Tumor Cells in Microscopic Images of Unstained Blood using Convolutional Neural Networks," in *2018 IEEE 14th International Conference on Intelligent Computer Communication and Processing (ICCP)*, 2018: IEEE, pp. 319-324, [10.1109/ICCP.2018.8516638](https://doi.org/10.1109/ICCP.2018.8516638).
- [46] A. Ciurte, C. Selicean, O. Soritau, and R. Buiga, "Automatic detection of circulating tumor cells in darkfield microscopic images of unstained blood using boosting techniques," *PloS one*, vol. 13, no. 12, p. e0208385, 2018, <https://doi.org/10.1371/journal>.
- [47] L. L. Zeune *et al.*, "Deep learning of circulating tumour cells," *Nature Machine Intelligence*, vol. 2, no. 2, pp. 124-133, 2020, <https://doi.org/10.1038/s42256-020-0153>.
- [48] L. Zeune *et al.*, "Quantifying HER-2 expression on circulating tumor cells by ACCEPT," *PloS one*, vol. 12, no. 10, p. e0186562, 2017, <https://doi.org/10.1371/journal.pone.0186562>.
- [49] B. He *et al.*, "A new method for CTC images recognition based on machine learning," *Frontiers in Bioengineering and Biotechnology*, vol. 8, p. 897, 2020, <https://www.frontiersin.org/journals/bioeng>.
- [50] Z. Guo, X. Lin, Y. Hui, J. Wang, Q. Zhang, and F. Kong, "Circulating tumor cell identification based on deep learning," *Frontiers in Oncology*, vol. 12, p. 843879, 2022, <https://doi.org/10.3389/fonc.2022.843879>.
- Disease," in *The Heart Surgery Forum*, 2023, vol. 26, no. 5, pp. E592-E599, <https://doi.org/10.59958/hsf.58>.
- [31] A. K. Swain, A. Swetapadma, J. K. Rout, and B. K. Balabantaray, "A Review on Lung Cancer Detection and Classification using Shallow Learning and Deep Learning," in *2023 OITS International Conference on Information Technology (OCIT)*, 2023: IEEE, pp. 570-573, [10.1109/OCIT59427.2023.10431005](https://doi.org/10.1109/OCIT59427.2023.10431005).
- [32] M. T. Ahad, Y. Li, B. Song, and T. Bhuiyan, "Comparison of CNN-based deep learning architectures for rice diseases classification," *Artificial Intelligence in Agriculture*, vol. 9, pp. 22-35, 2023, <https://doi.org/10.1016/j.aiaa.2023.07.001>.
- [33] G. E. Rao, B. Rajitha, P. N. Srinivasu, M. F. Ijaz, and M. Woźniak, "Hybrid framework for respiratory lung diseases detection based on classical CNN and quantum classifiers from chest X-rays," *Biomedical Signal Processing and Control*, vol. 88, p. 105567, 2024, <https://doi.org/10.1016/j.bspc.2023.105567>.
- [34] S. Wang, Y. Zhou, X. Qin, S. Nair, X. Huang, and Y. Liu, "Label-free detection of rare circulating tumor cells by image analysis and machine learning," *Scientific reports*, vol. 10, no. 1, p. 12226, 2020, <https://doi.org/10.1038/s41598-020-69056>.
- [35] J. A. Pruneski *et al.*, "Supervised machine learning and associated algorithms: applications in orthopedic surgery," *Knee Surgery, Sports Traumatology, Arthroscopy*, vol. 31, no. 4, pp. 1196-1202, 2023, <https://doi.org/10.1007/s00167-022>.
- [36] S. Naeem, A. Ali, S. Anam, and M. M. Ahmed, "An unsupervised machine learning algorithms: Comprehensive review," *International Journal of Computing and Digital Systems*, 2023, <https://www.researchgate.net/publication/368145444>.
- [37] D. Valkenborg, M. Geubbelmans, A.-J. Rousseau, and T. Burzykowski, "Supervised learning," *American Journal of Orthodontics and Dentofacial Orthopedics*, vol. 164, no. 1, pp. 146-149, 2023, <https://doi.org/10.1007/978-3-540>.
- [38] S. Patil and S. Patil, "Linear with polynomial regression: Overview," *Int J Appl Res*, vol. 7, pp. 273-275, 2021, <https://doi.org/10.1007/978-3-540>.
- [39] J. Isabona, A. L. Imoize, and Y. Kim, "Machine learning-based boosted regression ensemble combined with hyperparameter tuning for optimal adaptive learning," *Sensors*, vol. 22, no. 10, p. 3776, 2022, <https://doi.org/10.3390/s22103776>.
- [40] E. Hatzidaki, A. Iliopoulos, and I. Papisotiriou, "A novel method for colorectal cancer screening based