



Diagnosis of COVID-19 Disease Using Radiological Image Processing

Abbas Barzin¹, Mahmood Deypir², Jalil Mazloun³

Abstract

Background & Purpose: Due to the lack of definitive and effective treatment and the unknown side effects of Covid-19 disease, disruption of the transmission chain and the prevalence of this disease is of great importance. Providing intelligent and automated image analysis solutions is an effective step in controlling this disease that can reduce the workload in the field of health.

Methodology: This study introduced and examined the specific features of coronavirus in lung CT scan images, and processing and intelligent diagnostic methods (using machine learning method) to identify Covid-19 patients.

Findings: In this study, after receiving CT scan images of the lungs, first pre-processing is performed and then with the help of GLCM and HOG algorithms, feature extraction is performed from the images. Due to the large number of image extraction properties, the classification operation is not possible, so with the help of algorithm Chi-Square reduces the number of properties extracted and then uses the XGBoost algorithm to classify the extracted properties and distinguish between sick and healthy people.

Conclusion: Using a data set, 2484 CT scan images of healthy and sick people and using evaluation parameters, this study has achieved the results: 92.2% accuracy, 92.4% sensitivity and 92.3% F score.

Keywords: *Coronavirus, CT scan, machine learning, image processing*

1. Master student of Telecommunications (System) Shahid Sattari Aeronautical University of Science and Technology, Tehran, Iran.. **E-mail:** abbasbarzin99@gmail.com

2. Associate Professor, Department of Computer Engineering and Information Technology, Shahid Sattari Aeronautical University of Science and Technology, Tehran, Iran. **E-mail:** mdeypir@ssau.ac.ir

3. Associate Professor of Electrical Engineering, Shahid Sattari Aeronautical University of Science and Technology, Tehran, Iran.. **E-mail:** Jalil.mazloun@ssau.ac.ir



تشخیص بیماری کووید-۱۹ با استفاده از پردازش تصاویر رادیولوژی

عباس برزین^۱، محمود دی‌پیر^۲، جلیل مظلوم^۳

چکیده

زمینه و هدف: با توجه به نبود درمان قطعی و موثر و ناشناخته بودن عوارض جانبی بیماری کووید-۱۹، قطع زنجیره انتقال و شیوع این بیماری از اهمیت فراوانی برخوردار است. ارائه راه کارهای هوشمند و خودکار تحلیل تصاویر گامی موثر در کنترل این بیماری است که می‌تواند از فشار کاری حوزه سلامت کاسته شود. **روش شناسی:** این تحقیق به معرفی و بررسی شاخص‌های اختصاصی کروناویروس در تصاویر سی‌تی اسکن ریه، و روش‌های پردازشی و تشخیص هوشمند (با استفاده از روش یادگیری ماشین) برای شناسایی بیماراران کووید-۱۹ پرداخته است.

یافته‌ها: در این تحقیق بعد از دریافت تصاویر سی‌تی اسکن ریه‌ها، ابتدا پیش‌پردازش صورت می‌گیرد و سپس به کمک الگوریتم‌های GLCM و HOG کار استخراج ویژگی از تصاویر، صورت می‌پذیرد. بعلاوه تعداد ویژگی‌های استخراجی تصاویر، عمل طبقه‌بندی امکان‌پذیر نمی‌باشد بنابراین به کمک الگوریتم Chi-Square تعداد ویژگی‌های استخراج شده را کاهش داده و سپس با استفاده از الگوریتم XGBoost ویژگی‌های استخراجی را طبقه‌بندی کرده و افراد بیمار و افراد سالم را از یکدیگر تفکیک می‌کند. **نتیجه گیری:** این تحقیق با استفاده از مجموعه داده، ۲۴۸۴ تصویر سی‌تی اسکن افراد سالم و بیمار و استفاده از پارامترهای ارزیابی به نتایج: دقت ۹۲٫۲٪، حساسیت ۹۲٫۴٪ و امتیاز F_۱ ۹۲٫۳٪ دست پیدا نموده است.

کلیدواژه‌ها: کووید-۱۹، کروناویروس، سی‌تی اسکن، یادگیری ماشین، پردازش تصویر

۱. دانشجوی کارشناسی‌ارشد مخابرات (سیستم) دانشگاه علوم و فنون هوایی شهید ستاری - تهران- ایران. **رایانامه:**

abbasbarzin99@gmail.com

۲. دانشیار گروه مهندسی کامپیوتر و فناوری اطلاعات دانشگاه علوم و فنون ه ستاری-تهران-ایران. **رایانامه:**

mdeypir@ssau.ac.ir

۳. دانشیار گروه مهندسی برق دانشگاه علوم و فنون هوایی شهید ستاری -تهران-ایران. **رایانامه:**

Jalil.mazloum@ssau.ac.ir

تاریخ دریافت مقاله ۱۴۰۱/۰۲/۱۸

تاریخ پذیرش نهایی مقاله: ۱۴۰۱/۰۳/۰۹

نویسنده مسئول مقاله: محمود دی پیر

مقدمه

در دسامبر ۲۰۱۹، برای اولین بار در شهر ووهان استان هوبئی چین، پس از اینکه مردم بدون علت مشخصی دچار سینه پهلو شدند و واکسن‌ها و درمان‌های موجود مؤثر نبودند، نوع جدیدی از کروناویروس با همه‌گیری در انسان شناسایی شد. از میان نخستین افرادی که به این ویروس آلوده شده‌اند، مشخص شد که دو سوم آنها با بازار عمده‌فروشی غذاهای دریایی ملوانان، که در آن حیوانات زنده نیز به فروش می‌رسد، ارتباط داشته‌اند. تا ۳۰ بهمن، این کروناویروس جدید، جان بیش از ۲۷۰۹ نفر را گرفته و بیش از ۷۳،۳۳۲ نفر به این ویروس در ۸۰ کشور تأیید شده که از جمله در تایلند، کره جنوبی، ژاپن، تایوان، استرالیا، سنگاپور، نپال، ویتنام، اندونزی، آلمان، روسیه، فیجی، فرانسه، ایران و آمریکا مبتلا شدند. با عبور تعداد قربانیان ویروس کرونا از مرز ۱۰۰۰ نفر سازمان بهداشت جهانی برای بیماری ناشی آن نام رسمی انتخاب کرده‌است، کووید-۱۹ (COVID-19) که اشاره‌ای دارد به «کرونا»، «ویروس»، «بیماری» و سال ۲۰۱۹. این نوع کرونا ویروس قبل از این همه‌گیری در انسان دیده نشده بود. برآوردهای نخستین نشان می‌دهد که درصد مرگ و میر این ویروس بین ۲ تا ۳ درصد (یا بین ۱٪ و ۵٪) مبتلا شدگان است. علائم بیماری شامل: تب، سرفه‌های خشک و گاهی مشکلات تنفسی مانند تنگی نفس، تندنفسی و گلودرد و آبریزش بینی است. اولین مبتلایان این بیماری در ووهان در بازار خوراک دریایی کار می‌کردند یا در آنجا حضور داشتند؛ ولی با توجه به گسترش بیماری در افرادی که با حیوانات در تماس نبودند سازمان بهداشت جهانی انتقال انسان به انسان را نیز مدنظر قرار داده‌است.

بیماری کووید ۱۹ طیفی از علائم، از بی علامتی/قبل از بروز علائم (Asymptomatic/pre-symptomatic) الی بیماری بسیار شدید تظاهر می‌کند. خاطر نشان می‌شود که علائم ثابت نیست و در هر زمان ممکن است بیمار، وارد مرحله بعدی شود. دوران کمون بیماری کووید-۱۹ تا ۱۴-۳ روز می‌باشد و بطور متوسط در طی ۴-۵ روز پس از تماس، علائم آشکار می‌شود. طیفی از بیماری از بی علامت تا پنومونی شدید، سندروم دیسترس حاد تنفسی (ARDS)^۱ و مرگ ممکن است ایجاد شود. در حدود ۸۰٪ موارد بیماران مبتلا به کووید-۱۹ بصورت بی‌علامت، علائم خفیف تا متوسط دارند و در حدود ۱۵٪ موارد مبتلایان با علائم شدید و نیاز به بستری مراجعه می‌کنند. در ۵٪ موارد شرایط بیمار بحرانی شده و ممکن است نیازمند بستری در ICU^۲ و مراقبت‌های ویژه باشد. از دست دادن

^۱ Acute Respiratory Distress Syndrome

^۲ Intensive Care Unit

حس بویایی و نیز حس چشایی از جمله علائمی هست که در بسیاری از مبتلایان گزارش شده است. از علائم دیگر می توان به علائم گوارشی نظیر بی اشتها، ضعف بی حالی، خستگی زودرس، دل درد، تهوع، استفراغ و اسهال اشاره نمود.

روش های آزمایشگاهی مختلفی، جهت تشخیص کروناویروس وجود دارد که بیشتر آنها، آزمایش های (۱) مولکولی (۲) سرولوژیکی (۳) سی تی اسکن سینه می باشند. در مسیر تشخیص کووید-۱۹، اگرچه روش 'RT-qPCR' اختصاصی است اما باید نتایج منفی کاذب را نیز در نظر داشت چراکه تشخیص نادرست این بیماری عواقب جدی را به دنبال خواهد داشت، و همچنین روش های سرولوژیکی، فاقد ارزش نتایج فوری (تهیه نتایج آزمایش در طی دو هفته)، می باشند از این رو بسیاری از پزشکان پیشنهاد کرده اند که انجام سی تی اسکن سینه به عنوان یک روش تشخیص کمی ضروری است زیرا یک روش بسیار حساس است و می تواند تصاویر مشخصی از وقوع پنومونی (ذات الریه) و میزان درگیری ریه های بیمار را به نمایش بگذارد. [1] از این رو در افراد مشکوک به عفونت SARS-CoV-2^۲ که نتایج تست RT-qPCR در آنها منفی گزارش می شود، تکرار مجدد تست RT-qPCR و سی تی اسکن سینه می تواند مفید باشد. انجام سی تی اسکن با وضوح بالا (High resolution CT, HRCT) در جهت بررسی ریه ها برای تشخیص بیماری در مراحل اولیه و ارزیابی شدت عفونت کووید-۱۹ در بیماران ضروری است. سی تی اسکن های معمولی در بیماران کووید-۱۹، تصاویری از شیشه مات در بافت پارانشیمی هر دو طرف ریه (Bilateral pulmonary parenchymal ground-glass) و کدورت های ریوی تثبیت شده (Consolidation) (Pulmonary opacities) را نشان می دهند و در بعضی از موارد به صورت اشکال دایره ای که به سمت اطراف خارجی بافت ریه توسعه پیدا کرده اند، دیده می شوند. سی تی اسکن سینه به عنوان یک روش تشخیصی بسیار باارزش در جهت تشخیص بیماری کووید-۱۹ در نظر گرفته می شود، به خصوص در مناطقی که ویروس SARS-CoV-2 شیوع بالایی دارد، در ابتدای تحقیقات، تشخیص افتراقی بیماری های سارس، مرس و کووید-۱۹ (و یا سایر بیماری های ویروسی ذات الریه)، از روی تصاویر سی تی اسکن سینه، امری غیرممکن بود ولیکن امروزه با بکارگیری هوش مصنوعی، این نقیصه برطرف گردیده است و بر محبوبیت و اختصاصیت این روش افزوده است. [7]

¹ Real-Time reverse transcription Polymerase Chain Reaction

² CT: Computed Tomography

³ Severe Acute Respiratory Syndrome coronaVirus 2

پیشینه پژوهش

امروزه روش‌های متعددی جهت تشخیص بیماری کووید-۱۹ مانند: علائم بالینی مرتبط با بیماری و روش‌های تشخیصی با دقت بالاتر، نظیر تصویربرداری‌های پزشکی استفاده می‌گردد. چندین روش تصویربرداری از جمله اشعه ایکس، سی‌تی اسکن، سونوگرافی و ... برای تشخیص بیماری COVID-19 می‌توان نام برد اما در حال حاضر، تصویربرداری از ریه با استفاده از سی‌تی اسکن، بهترین روش تشخیصی بیماری شناخته شده است. [4,7] چنانچه که آگاهی دارید تشخیص بیماری در تصویر ریه توسط پزشکان، کاری زمان‌بر است. بنابراین به منظور کاهش فشار کاری رادیولوژیست‌ها، ابزارهای تشخیصی مبتنی بر فناوری یادگیری عمیق یا یادگیری ماشین ایجاد شده‌اند که می‌توانند با دقت خوبی وجود و شدت بیماری را تشخیص دهند. پس از به‌دست آوردن یک تصویر، تقسیم‌بندی تصویر یک گام اساسی در تجزیه و تحلیل و پردازش تخمین COVID-19 می‌باشد.

بسیاری از مطالعات که اخیراً منتشر شده‌اند، نشان می‌دهد که در تصاویر سی‌تی اسکن ریه بیمار در کرونا، معمولاً کدرشدگی شیشه‌ای پس‌زمینه (GGO) دیده می‌شود. در نتیجه برای رادیولوژیست‌ها، یافتن مناطق غیرطبیعی مانند: کدرشدگی شیشه‌ای پس‌زمینه (GGO) الگوی سنگفرشی، به‌هم‌پیوستگی ریوی، سایه‌های تکه تکه دو طرفه، فیروز ریوی و ضایعات چندگانه در تصاویر سی‌تی برای تشخیص COVID-19، مهم است. تشخیص خودکار GGO می‌تواند به کاهش بار کاری پرسنل پزشکی کمک کند. از جمله روش‌های پیشنهادی استفاده از UNet ++ بوده که برای استخراج مناطق غیرطبیعی ریه در تصاویر سی‌تی استفاده شده است و توانسته است با حساسیت و دقت خوبی موارد ابتلا را در تصویر سی‌تی تشخیص دهد. [7]

راه‌کار دیگری که پیشنهاد شده است، استفاده از یک مدل هوش مصنوعی یا یادگیری ماشین، بوده که به‌جای این که فقط نواحی غیرعادی یا عفونی را تفکیک کند، از روش تجزیه و تحلیل رادیومیک استفاده کرده است. ویژگی‌های رادیومیک به‌صورت دستی از ROI مشخص شده، استخراج می‌شوند. ROI منطقه مورد علاقه ما است که آسیب‌هایی مانند: ضایعات، لوب‌ها، ریه‌های آلوده را در تصویر نشان می‌دهد. از ROI تقسیم‌بندی شده برای استخراج ویژگی‌ها و تجزیه و تحلیل اهداف تشخیصی، مورد استفاده قرار می‌گیرد. سپس از روش‌های خوشه‌بندی توافقی: با نظارت، بدون نظارت و یا نیمه نظارت شده برای

¹ Opacity Ground-Glass

² Region Of Interest

انتخاب ویژگی‌های مهم مرتبط با COVID-19 استفاده می‌شوند و همچنین در نهایت می‌تواند از ماشین بردار پشتیبانی (SVM)، برای طبقه‌بندی ویژگی‌ها و تشخیص بیماری، استفاده شود. [7]

روش دیگری که دنبال شده است، این بوده که چون تفسیر ROIهای بیماران مختلف وقت گیر است، یک مدل یادگیری عمیق برای شناسایی خودکار مناطق عفونی در تصاویر سی تی پیشنهاد شده است و سپس از معماری‌هایی هم‌چون: ResNet50, DResNet3, Ncovnet, VGG-19 و غیره ... برای شناسایی COVID-19 در مناطق عفونی استفاده شده است. مناطق عفونی ابتدا با استفاده از یک مدل یادگیری عمیق سه بعدی از یک مجموعه تصاویر سی تی اسکن ریوی انتخاب می‌شوند. این تصاویر به سه گروه بیماران COVID-19، عفونت ویروسی آنفلوانزا و سالم دسته‌بندی شده و در نهایت، نوع عفونت و درجه اطمینان کل با تابع بیزین محاسبه می‌شود. [4] جهت اشاره به نتایج کیفی تجربیات به دست آمده در سال- های اخیر در این زمینه، می‌توان به موارد ذیل اشاره کرد.

در تحقیق Xiaowei Xu و همکارانش [1] از یک شبکه عصبی کانولوشن با معماری VNet برای پردازش تصاویر سی تی اسکن استفاده کردند. جهت ارزیابی از دیتاست سی تی اسکن ریه افراد مبتلا به کووید-۱۹، شامل ۶۱۸ تصویر سی تی اسکن (۲۱۹ عکس افراد مبتلا به کووید-۱۹ و ۳۹۹ افراد سالم) استفاده کردند و به دقت ۸۶٫۷۰٪ در طبقه‌بندی باینری به دو کلاس COVID و Covid Non دست یافتند. و یا کار تحقیقاتی Jun Chen و همکارانش [1] با استفاده از معماری UNet++ و دیتاست ۴۶۰۹۶ عکس سی تی اسکن ریه، به دقت ۹۵٫۲۴٪ رسیدند.

در تحقیقاتی دیگر، از یک سیستم هوش مصنوعی^۱ برای تقسیم‌بندی و طبقه‌بندی ضایعات در COVID-19، استفاده شده است. در این روش، ابتدا مناطق ریه با استفاده از-3D UNet استخراج شده و سپس، ضایعات خاص در مناطق ریه تقسیم‌بندی شده‌اند و نهایتاً یک طبقه‌بندی کننده مبتنی بر^۲ CNN ضایعات را تفکیک نموده است. [4]

به‌طور مثال در تحقیق wang و همکارانش [4] با دیتاست ۱۲۴۸ تصویر سی تی اسکن (۴۵ عکس افراد مبتلا به کووید-۱۹ و ۱۲۰۳ افراد سالم) و استفاده از معماری CNN، دقت ۸۳٫۵۰٪ را در پی داشت و یا می‌توان اشاره به کار تحقیقاتی آقای Zheng و همکارانش [4]

¹ Support Vector Machines

² Artificial Intelligence

³ Convolutional Neural Networks

با مجموعه داده ۵۴۲ تصویر سی‌تی اسکن (۳۱۳ عکس افراد مبتلا به کووید-۱۹ و ۲۲۹ افراد سالم) و استفاده از معماری‌های CNN و UNet، کرد که دقت ۹۵٫۹۰٪ را به‌دنبال داشت.

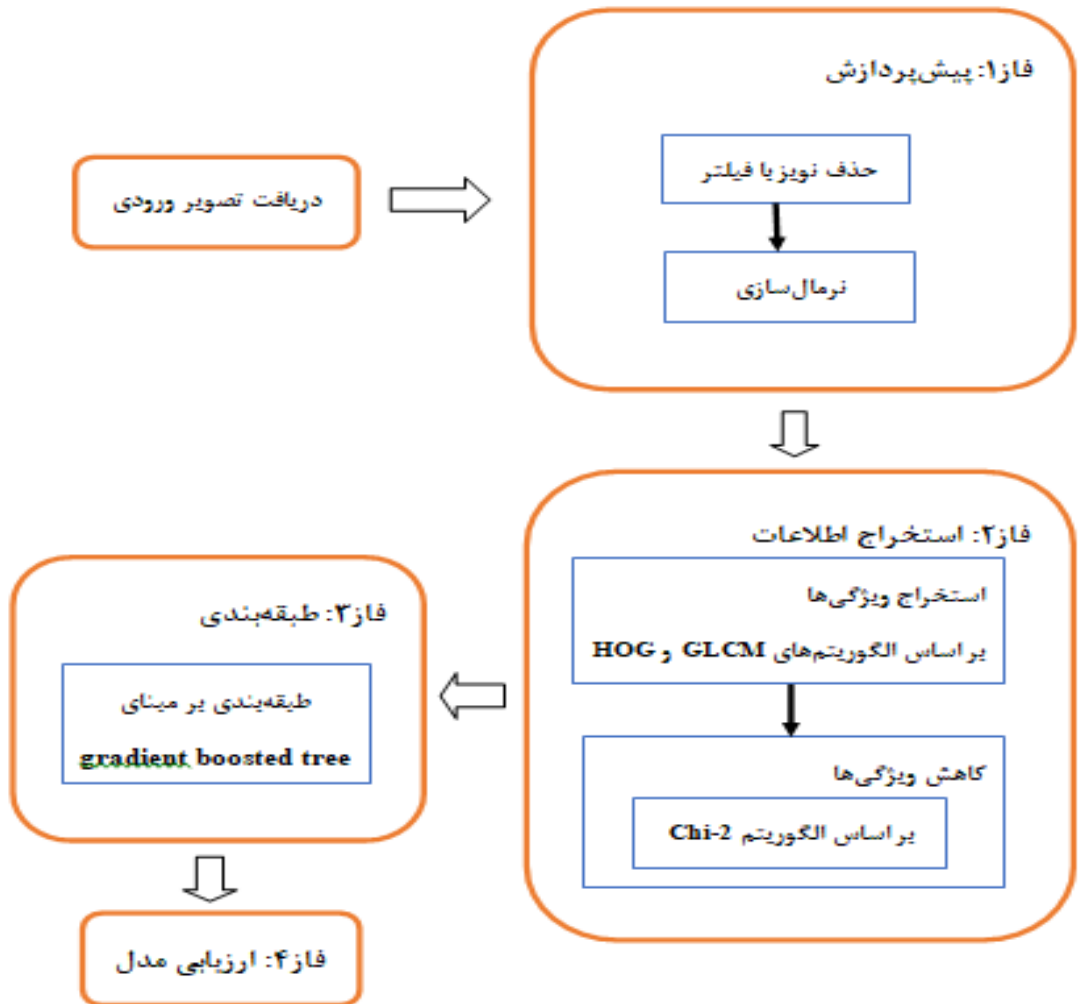
با توجه به مطالب پیش‌گفته و نتایج امیدوار کننده در مطالعات اخیر، بسیاری از

مدل‌های هوش مصنوعی در مجموعه داده‌های کوچک آزمایش شده‌اند. دقت بالای مدل‌های هوش مصنوعی در مجموعه داده‌های کوچک، ممکن است ناشی از پیش‌پردازش تصاویر باشد. مطالعاتی که از مجموعه داده‌های کوچک (مثلاً ۳۰۰ بیمار) استفاده می‌کنند، اغلب دقت بالایی از خود نشان می‌دهند. با این حال، مطالعات نشان داده است که راندمان عملکرد با مجموعه داده بزرگتر، کاهش می‌یابد. مطالعات مقیاس بزرگتر با دقت کمتر ممکن است نشان‌دهنده توانایی واقعی هوش مصنوعی در تشخیص کرونا باشد. از سوی دیگر، مطالعات بسیار کمی از مجموعه داده اعتبار مستقل و خارجی برای ارزیابی توانایی تعمیم مدل‌های یادگیری عمیق استفاده کرده‌اند. [1]

روش‌شناسی پژوهش

هنگامی که تصویر از ورودی دریافت می‌شود در ابتدا باید مراحل را طی کند تا بتوانیم روش‌های موردنظر را بر روی آن اعمال کنیم. بعد از اینکه تصویر ورودی خوانده شد حال باید پیش‌پردازش روی آن اعمال شود. بعد از آن با استفاده از ویژگی‌های استخراج شده و کاهش ویژگی به‌منظور طبقه‌بندی ضایعات شناسایی شده، از الگوریتم درخت گرادیان تقویتی^۱ استفاده می‌کنیم به‌منظور سرعت بخشیدن در کار شناسایی و تشخیص ضایعات از شبکه‌های عصبی استفاده می‌کنیم. در شکل ۱ چارت روش انجام تحقیق مشخص شده است.

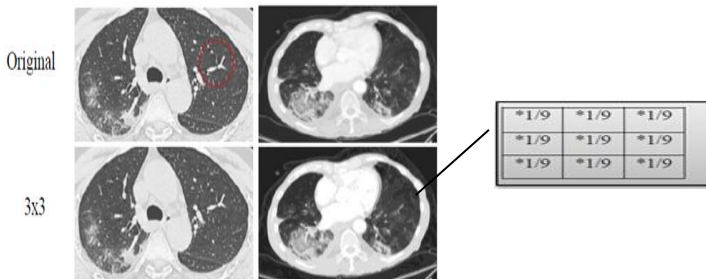
¹ gradient boosted tree



شکل ۱. شماتیک روش انجام تحقیق مدل پیشنهادی.

فاز ۱: پیش‌پردازش

در این مرحله الگوریتم‌های پیش‌پردازش ارائه شده است. این الگوریتم‌ها شامل نرمال‌سازی اندازه، ابعاد تصویر و حذف نویز می‌باشند که جهت بهبود کیفیت تصاویر استفاده می‌شوند تا اعتبار بیشتری پیدا کنند. اولین گام پیش‌پردازش حذف نویز است. هدف این گام کاهش نویز و جداکردن سیگنال از نویز است در این مرحله با حذف نویز، تصویر برای مرحله بعدی آماده می‌شود. برای حذف نویز از فیلتر میانگین‌گیر که از فیلترهای پرکاربرد است استفاده خواهد شد. ابتدا یک کرنر از تصویر مشخص می‌شود که اندازه آن بستگی به میزان تارشدگی تصویری که مدنظر است دارد. در این مدل از یک کرنر 3×3 استفاده کرده‌ایم برای این کار ابتدا کرنر را روی تصویر اولیه قرار می‌دهیم وقتی این ماسک روی تصویر اولیه قرار می‌گیرد.



شکل ۲. CT اسکن تصاویر پس از کاهش نویز با استفاده از فیلتر میانی 3×3 .

سطح روشنایی تمام پیکسل‌هایی که در اطراف آن پیکسلی که به‌عنوان مرکز انتخاب شده باهم جمع می‌کند و حاصل این جمع شدن بر عدد ۹ تقسیم می‌کند (چون کرنر ما از ۹ پیکسل تشکیل شده است). نتیجه حاصل میانگین این ۹ پیکسل است که وقتی به پیکسل متناظرش در ماتریس نتیجه اعمال شود یک شدت رنگ متناسب با این پیکسل و همسایه‌های آن خواهد بود. روش آسان‌تر آن ضرب تک تک پیکسل‌ها در $1/9$ است سپس جمع آنها با یکدیگر انجام می‌شود.

گام بعدی پیش‌پردازش نرمال‌سازی اندازه و ابعاد تصویر می‌باشد. ابتدا باید تصاویر به تاثیر عواملی مثل فاصله خطوط، اندازه‌ها نرمال‌سازی شوند تا بلوک یکنواختی از تصویر حاصل شود تصاویر ورودی هر کدام ابعاد و اندازه متفاوتی برای خود دارند تعیین ابعاد امری ضروری در روش‌های هوش مصنوعی است چون تصویر به ماتریس‌های کاملاً یکسانی تبدیل می‌شود. دو تصویر با اندازه‌های متفاوت را می‌توانیم با دستور `imresize` در نرم‌افزارهای متلب و پایتون به تصاویر با اندازه‌های یکسان تبدیل کنیم.

فاز ۲: استخراج اطلاعات (استخراج ویژگی‌ها براساس هیستوگرام)

هدف از استخراج ویژگی به دست آوردن داده‌های خام به شکل قابل استفاده‌تری به منظور پردازش‌های آماری می‌باشد. در اغلب روش‌های موجود برای ارزیابی، شناسایی و بهبود کیفیت تصاویر، استخراج ویژگی از مراحل ویژه به حساب می‌آید. درحقیقت، وقتی حجم داده ورودی به یک الگوریتم زیاد است و پردازش آن مدت زیادی طول می‌کشد، استخراج ویژگی انجام می‌شود تا از داده‌هایی که متناسب با حجم زیادشان اطلاعات ندارند، یک مجموعه ویژگی به دست آید که نمایش مختصرتری از داده است که برای نمایش یک کلاس یا یک مجموعه نمونه باید شناخته شده باشند. انتخاب ویژگی نویسه، یا یک نمونه اصلی سعی می‌کند تا بهترین ویژگی‌ها را که بیشترین اطلاعات متمایزکننده را در خود دارند، ضمن رعایت محدود بودن تعداد ویژگی از ویژگی‌های استخراج شده، به دست آورد. مکان نسبی نقاط انتهایی، نقاط تقاطع و گوشه‌ها از ویژگی‌های محلی و تعداد حفره‌ها در یک نویسه یا تعداد تقعرها در مرز بیرونی از ویژگی‌های کلی هستند. یکی از روش‌های متداول برای استخراج اطلاعات، استخراج ویژگی براساس هیستوگرام می‌باشد که در این مدل از آن استفاده شده است.

هیستوگرام تصویر نموداری است که توسط آن تعداد پیکسل‌های هر سطح روشنایی در تصویر ورودی مشخص می‌شود. فرض کنید تصویر ورودی یک تصویر مقیاس خاکستری^۱ با ۲۵۶ سطح روشنایی باشد، بنابراین هر یک از پیکسل‌های تصویر مقداری در بازه ۰ تا ۲۵۵ می‌توانند داشته باشند. بنابراین با محاسبه هیستوگرام، شما برای هر تصویر یک آرایه به طول ۲۵۶ دارید که یک بردار ویژگی تصویر مورد نظر شما است. یکی از راه‌های موثر در استخراج ویژگی نمایش روش هیستوگرام طیفی برای عمل تشخیص و دسته‌بندی می‌باشد که دارای ویژگی تبدیل - ثابت است. این نمایش با تقسیم حوزه فرکانس به نواحی کوچک مستقل از هم به دست می‌آید.

این کار منجر به تولید یک مجموعه از فیلترها و یک نمایش که شامل توزیع‌های حاشیه‌ای و پاسخ‌های آن فیلترها می‌باشد، می‌شود. به این ترتیب حوزه فرکانس به نواحی کوچک تقسیم خواهد شد و فیلترهای مکانی مربوطه به دست می‌آید و کانون‌شن تصویر داده شده با هر فیلتری محاسبه شده و توزیع حاشیه‌ای تصویر فیلتر شده حاصل را به دست خواهد آورد و آنگاه، توزیع‌های حاشیه‌ای با هم ترکیب خواهند شد. بنابراین هر تصویر به وسیله یک

¹ Grayscale

بردار شامل توزیع‌های حاشیه‌ای نمایش داده می‌شوند که نمایش هیستوگرام طیفی تصویر نام دارد. منظور از روش هیستوگرام برای استخراج ویژگی، استفاده از روش‌های اصطلاحاً "هیستوگرام استاتیک"^۱ است. در این روش برای استخراج ویژگی از تصاویر با استفاده از مشخصه‌های آماری مرتبه‌ی اول (براساس هیستوگرام شدت روشنایی پیکسل‌ها) ویژگی‌هایی از تصویر استخراج شده است و تعدادی از این ویژگی‌ها شامل: میانگین، چولگی، انحراف مطلق، واریانس، کشیدگی و انحراف معیار می‌باشند. به این علت اسم هیستوگرام را دارد که در آنها براساس روشنایی تک پیکسل (نه براساس بافت تصویر و همسایگی تعداد پیکسل) ویژگی استخراج می‌شود (ویژگی‌های استاتیک). با استفاده از روش‌های آماری مراتب بالاتر هم استخراج ویژگی‌های بافتی نتیجه می‌شود که ویژگی‌های دیگری هستند و بر اساس بافت تصویر از تصویر حاصل می‌شوند و با این ویژگی‌ها متفاوتند که مستلزم استفاده از قوانین و تکنیک‌های GLCM، GLRLM و ... است. برای طبقه‌بندی و کلاس‌بندی داده‌ها نیاز به یک سری دیتا داریم. وقتی که بر روی تصاویر کار می‌کنیم و این دیتاها در اختیارمان نیست، نیاز هست که یک سری ویژگی‌ها را از تصویر استخراج کنیم که یکی از این روش‌ها استفاده از ویژگی‌های بافت تصویر است. همان‌طور که در مطالب بالاتر توضیح داده شد می‌توان از اطلاعات هر پیکسل و بر اساس مشخصه‌های آماری مرتبه‌ی اول این ویژگی‌ها را استخراج کرد که استخراج ویژگی از روش‌های استاتیک نامیده می‌شود اما می‌توان علاوه بر آن از مشخصه‌های آماری مرتبه‌ی دوم یا همان GLCM و یا مشخصه‌های آماری مراتب بالاتر مانند: SURF، SIFT، HOG، GLRLM و... استفاده کرد. برخی روش هیستوگرام استاتیک / استاتیک را هم جزو ویژگی‌های بافتی می‌دانند (چون علاوه بر اینکه از پیکسل‌های تصویر استفاده می‌کند از اطلاعات تک پیکسل هم استفاده می‌کند) اما GLCM و GLRLM براساس ویژگی‌های بافتی و بر اساس اطلاعات پیکسل‌های همسایگی هر پیکسل این ویژگی‌ها را نتیجه می‌دهند (نه یک پیکسل تک و خاص در مواردی که مثلاً دقت تصویر ما پایین هست). مثل تصاویر MRI^۲ که حاوی نویز هستند نیاز است که بر اساس اطلاعات همسایگی پیکسل‌ها این ویژگی‌ها استخراج شود. پس این روش‌ها، روش‌های خوبی برای این نوع تصاویر هستند. روش GLCM از ماتریس‌های هم‌رخداد که اولین بار توسط هارالیک معرفی شده، مشخصه‌ها و ویژگی‌های هارالیک را استخراج کرده است. این روش بر اساس ماتریس‌هایی کار می‌کند که در جهت و زوایای مختلف می‌توانند

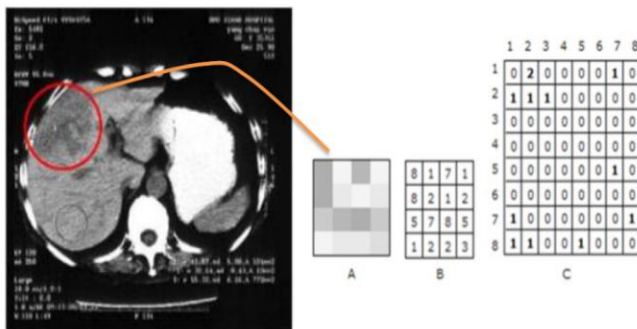
^۱ static histogram

^۲ Magnetic Resonance Imaging

ویژگی‌هایی از بافت را نتیجه دهند و در واقع بر اساس احتمال اتصال هر جفت پیکسل در کنار هم عمل می‌کنند.

• استخراج ویژگی براساس GLCM (مبتنی بر سطح خاکستری).

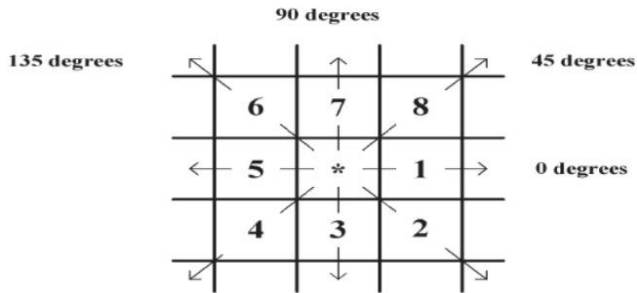
همان‌طور که گفته شد در GLCM^۱ که مشخصه آماری مرتبه دو است، ویژگی‌هایی را از ماتریس هم‌رخداد استخراج کرده است. ویژگی مبتنی بر سطح خاکستری بر اساس تفاوت بین سطح خاکستری در پیکسل موردنظر و ارزش آماری یک نماینده از محیط اطراف استخراج می‌شود. به عنوان مثال غده‌های سرطانی دارای رنگ تضاد نسبت به اطراف خود هستند این ویژگی در توصیف تنوع سطح خاکستری در اطراف پیکسل موردنظر انتخاب خوبی است. به عنوان مثال ماتریس (C) شکل ۳ یک ماتریس ۸*۸ را نشان می‌دهد. اگر ماتریس سمت چپ (B) را ماتریس تصویر اصلی فرض کنیم مشاهده می‌کنیم که حاوی عددهایی است، که این اعداد مربوط به سطح خاکستری هست که هر پیکسل از تصویر می‌تواند داشته باشد. اما در ماتریس سمت راست (C) سطر و ستون آن شامل بزرگترین عدد سطح خاکستری تصویر اصلی است. در ماتریس B احتمال وقوع اینکه عدد ۱ در کنار عدد ۱ قرار گرفته باشد، چقدر است: می‌بینیم که در ماتریس B، ۱ در هیچ حالتی کنار ۱ قرار نگرفته، پس در ماتریس نتیجه یعنی سمت راستی هم عدد صفر در خانه‌ی ۱ و ۱ قرار می‌گیرد. و یا مثلاً ۸ و ۱ در ماتریس B یکبار کنار هم قرار گرفتن پس در ماتریس سمت راست در خانه‌ی ۱ و ۸ عدد ۱ قرار می‌گیرد.



شکل ۳. فرایند نمایش ماتریس هم‌رخداد با استفاده از ویژگی مبتنی بر سطح خاکستری.

¹ Gray-Level Co-occurrence Matrix

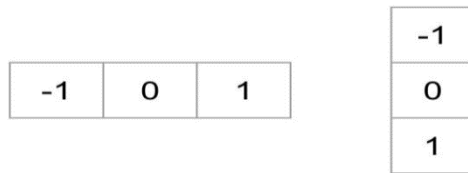
این مثال بر اساس جهت صفر درجه، این ماتریس را بررسی کرده است. همان‌طور که در تصویر زیر نشان داده شده است می‌توانیم از زوایای ۴۵، ۹۰ و ۱۳۵ درجه هم برای استخراج این ویژگی‌ها استفاده کنیم.



شکل ۴. نمایش استخراج ویژگی از زوایای مختلف

• استخراج ویژگی بر اساس HOG.

الگوریتم HOG^1 یک روش برای استخراج ویژگی‌های محلی در تصویر از طریق محاسبه هیستوگرام می‌باشد که در سال ۲۰۰۵ توسط Navneen Dalal ابداع گردید. برای محاسبه هیستوگرام شیب‌های جهت‌دار (HOG) ابتدا باید گرادیان‌های افقی و عمودی را محاسبه کرد. این کار به وسیله فیلتر کردن تصویر توسط کرنل‌های زیر انجام خواهد شد.

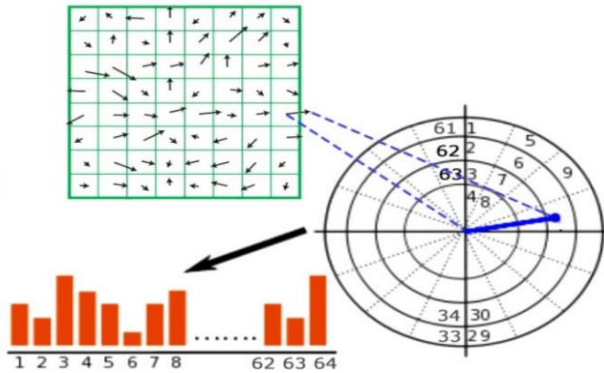


سپس با استفاده از فرمول زیر می‌توان اندازه و جهت شیب را محاسبه کرد.

$$g = \sqrt{g_x^2 + g_y^2} \quad \theta = \tan^{-1} \frac{g_y}{g_x} \quad (۱.۲)$$

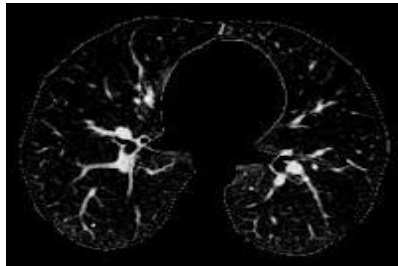
¹ Histogram of Oriented Gradients

هیستوگرام اندازه جهت شیب‌دار، را می‌توان بصورت شماتیک در شکل زیر مشاهده نمود.



شکل ۵. نمایش استخراج ویژگی با استفاده از هیستوگرام الگوریتم HOG.

استفاده از این توصیفگر همانند فیلترهای مشتق‌گیر چون اپراتورهای سوبل می‌تواند جهت آشکارسازی لبه‌ها نتایج یکسانی داشته باشند که به‌طور مثال تصویر زیر بعد از اعمال توصیفگر HOG به‌دست آمده است.



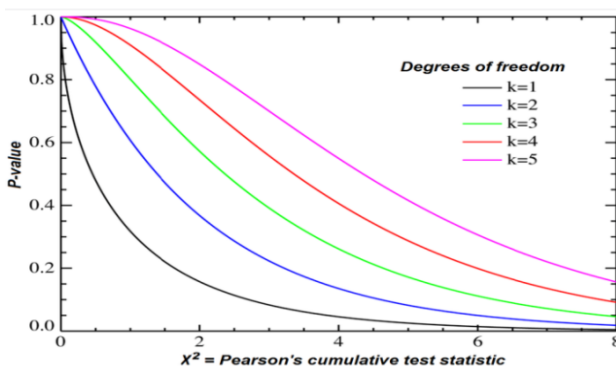
شکل ۶. نمایش استخراج ویژگی بعد از اعمال توصیفگر HOG.

• کاهش ویژگی‌ها (براساس الگوریتم Chi-square)

پیشرفت‌های به‌وجود آمده در جمع‌آوری ویژگی و قابلیت‌های ذخیره‌سازی در طی دهه‌های اخیر باعث شده در بسیاری از علوم با حجم بزرگی از ویژگی‌ها و اطلاعات روبرو شویم. در مقایسه با بسترهای داده‌ای قدیمی و کوچکتر، بسترهای داده‌ای امروزی چالش‌های جدیدی در تحلیل داده‌ها بوجود آورده‌اند. روش‌های آماری سنتی به دو دلیل امروزه کارایی خود را از

دست داده‌اند: علت اول افزایش تعداد مشاهدات است، و علت دوم که از اهمیت بالاتری برخوردار، افزایش تعداد متغیرهای مربوط به یک مشاهده می‌باشد، که چالش‌های محاسباتی زیادی را ایجاد می‌کند. یکی از مشکلات داده‌ها با ویژگی زیاد این است که در بیشتر مواقع تمام ویژگی‌های داده‌ها برای یافتن دانشی که در داده‌ها نهفته است مهم و حیاتی نیستند. به همین دلیل در بسیاری از زمینه‌ها کاهش ویژگی یکی از مباحث قابل توجه باقی مانده است. در مدل پیشنهادی از الگوریتم Chi-square، جهت کاهش ویژگی، استفاده شده است.

آزمون توزیع نرمال یا آزمون مربع کای یا آزمون χ^2 (به صورت χ^2 به انگلیسی: Chi-squared test) نوشته می‌شود از آزمون‌های آماری است و برای ارزیابی میزان ارتباط متغیرهای اسمی نسبت به هم به کار می‌رود. برای تعیین اینکه بدانیم آیا از نظر آماری اختلاف معنی‌داری بین مقادیر فراوانی مشاهدات و فراوانی مورد انتظار در یک یا چند گروه از جدول توافقی (دوطرفه) وجود دارد، از آزمون کای- χ^2 پیرسون استفاده می‌شود. در کاربردهای معمول این آزمون، مشاهدات به کلاس‌های جدا از هم تقسیم می‌شوند. اگر فرضیه صفر مبنی بر اینکه هیچ تفاوتی بین طبقات در جمعیت وجود ندارد درست باشد، آمار آزمون محاسبه شده مشاهدات توزیع فراوانی χ^2 را دنبال می‌کند. هدف از این آزمون ارزیابی چگونگی احتمال فرضیه صفر در فراوانی مشاهدات است.



شکل ۷. نمودار کای- χ^2 .

مقدار کای- χ^2 طبق رابطه زیر محاسبه می‌شود

$$\chi^2 = \sum_{t=1}^m \frac{(\theta_t - E_t)^2}{E_t} \quad (3)$$

که در آن:

θ_t = مقدار مشاهده شده (فراوانی هر داده)

$E_t =$ فراوانی‌های مورد انتظار

پس از محاسبه جدول فراوانی‌های مورد انتظار و محاسبه مقدار درجه آزادی با مقایسه مقادیرهای به دست آمده با جدول توزیع مربع کای می‌توان در مورد اینکه دو متغیر با هم ارتباط دارند یا خیر نتیجه‌گیری کرد. این آزمون تنها راه حل موجود برای آزمون همگنی در مورد متغیرهای مقیاس اسمی با بیش از دو مقوله است؛ بنابراین، کاربرد خیلی زیادتری نسبت به آزمون‌های دیگر دارد. این آزمون نسبت به حجم نمونه حساس است. آزمون مربع کای برای تعیین تفاوت‌ها میان چند چیز هم بکار می‌رود. که یکی از اساسی‌ترین کاربردهای آن در علوم کامپیوتر به خصوص مبحث یادگیری ماشین و یادگیری عمیق است.

فاز ۳: طبقه‌بندی بر مبنای الگوریتم XGBoost

در فاز سوم، ویژگی‌های تصاویر بعد از استخراج و کاهش در فاز دوم، دسته‌بندی می‌شوند. الگوریتم مورد استفاده در طبقه‌بندی ویژگی‌ها، XGBoost می‌باشد که جزء الگوریتم‌های گرادیان تقویتی می‌باشد. در این روش از الگوریتم‌های چندگانه ضعیف (ترکیبی از الگوریتم‌های خطی (liner) و درختی (tree)) برای ایجاد یک الگوریتم دقیق‌تر استفاده می‌کند. به جای استفاده از یک تخمین زنده، چندین الگوریتم پایدار و قوی‌تر خواهیم داشت. گرادیان تقویتی با ساختن یک درخت برای تلاش برای متناسب کردن داده‌ها شروع به کار می‌کند، و درختان بعدی با هدف کاهش باقیمانده‌ها (خطا) ساخته می‌شوند. این کار را با تمرکز بر مناطقی انجام می‌دهد که یادگیرندگان موجود عملکرد ضعیفی دارند.

این الگوریتم به خاطر داشتن دقت بالا و عملکردی بسیار قدرتمند، امروزه یکی از محبوب‌ترین و پرکاربردترین الگوریتم‌ها، محسوب می‌گردد به طوری که نقش تعیین‌کننده‌ای برای برنده و بازنده شدن در رقابت‌های Kaggle را ایفاء می‌کند. این الگوریتم که اخیراً در حوزه یادگیری ماشین بکار گرفته می‌شود، به تازگی برای سرعت و کارایی بالا طراحی شده است. الگوریتم XGBoost قدرت پیش‌بینی فوق‌العاده بالایی دارد که آن را به بهترین انتخاب برای داشتن بالاترین صحت در مسابقات مبدل می‌سازد، زیرا هم از مدل‌های خطی و هم از الگوریتم یادگیری درختی استفاده می‌کند و دستکم ۱۰ برابر سریع‌تر از سایر روش‌های گرادیان تقویتی می‌باشد.

از دیگر مزایای این الگوریتم می‌توان به انعطاف‌پذیری (پشتیبانی از مسائل کلاسه‌بندی، رگرسیون و رتبه‌بندی)، دسترسی (در دسترس برای زبان‌های برنامه‌نویسی مانند R، Python، Java، Julia و Scala) و ذخیره و بارگذاری مجدد (برای تجدید و محاسبات پیچیده)،

اشاره نمود. یکی از جالب‌ترین موارد پیرامون الگوریتم XGBoost، روش تقویتی با قاعده (Regularized Boosting Technique) می‌باشد. این روش به کاهش بیش-برازش (Overfit) مدل کمک کرده و پشتیبانی وسیعی را برای طیف گسترده‌ای از زبان‌ها مانند: اسکالا (Scala)، جاوا (Java)، آر (R)، پایتون (Python)، جولیا (Julia) و سی‌پلاس‌پلاس (C++) فراهم می‌کند. الگوریتم XGBoost، آموزش توزیع‌شده و گسترده‌ای را در بسیاری از ماشین‌ها فراهم می‌کند و شامل خوشه‌هایی مانند AWS، GCE، Azure و Yarm می‌شود. XGBoost قابل یکپارچه‌سازی با اسپارک (Spark)، فلینک (Flink) و دیگر سیستم‌های جریان داده با یک اعتبارسنجی متقابل (Cross validation) در هر تکرار از فرآیند شتاب‌دهی است. موارد پیش‌گفته، بخشی از دلایلی برای استفاده از این الگوریتم می‌باشد.

فاز ۴: ارزیابی روش کار مدل پیشنهادی

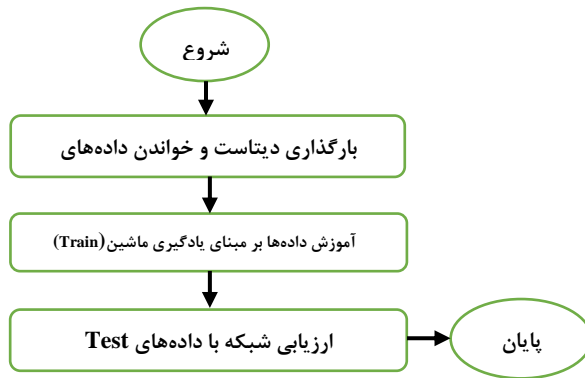
در بخش روند کاری نحوه ارزیابی مدل را توضیح خواهیم داد.

مراحل روند کاری طرح پیشنهادی

این تحقیق در ابتدا دیتاست موردنظر ما که SARSCoV-2 CT Scan Dataset است، بارگذاری می‌شود. در این دیتاست دو نوع داده وجود دارد COVID و Non-COVID، در مرحله اول، داده‌ها خوانده شده، تغییر اندازه و نرمال‌سازی می‌شوند سپس با توجه به داده‌ها دو کلاس اصلی، سالم و کووید-۱۹ ایجاد می‌شود. همچنین تصاویر را به دو دسته آزمایش (Test) و آموزش (Train) تقسیم می‌کنیم و با طرح پیشنهادی یادگیری ماشین، دسته‌ی Train را آموزش می‌دهیم، مراحل آموزش را آن‌قدر انجام می‌دهیم که میزان خطا کاهش یافته و بتوانیم مدل‌های آموزشی مناسبی را درخور این طرح پیدا نماییم. هنگامی که آموزش شبکه با داده‌های Train، به پایان رسید، نوبت به اعتبارسنجی شبکه روی داده‌های Test موجود در دیتاست، است و نتیجه به‌صورت طبقه‌بندی به دو حالت کووید-۱۹ و سالم روی داده‌های تست نمایش داده می‌شود.

این تحقیق ۳ مرحله اساسی دارد:

- ۱ مرحله بارگذاری دیتاست
- ۲ مرحله آموزش
- ۳ طبقه‌بندی و ارزیابی



شکل ۸. مراحل روندکاری طرح پیشنهادی

• مرحله بارگذاری دیتاست

در این تحقیق ما از دیتاست عکس‌های سی‌تی اسکن ریه بیماران در SARS-CoV-2 CT Scan Dataset [21] استفاده کردیم. که شامل ۲۴۸۲ تصویر سی‌تی اسکن ریه است، در این دیتاست ۲ نوع پوشه با نام‌های COVID و Non-COVID که شامل: ۱۲۵۲ عکس سی‌تی اسکن ریه افراد مبتلا به بیماری کووید-۱۹ و ۱۲۳۰ عکس سی‌تی اسکن ریه متعلق به افراد سالم می‌باشد، وجود دارد. در این مجموعه ۱۹۸۵ تصویر سی‌تی اسکن ریه جهت Train (۸۰٪) و ۴۹۷ تصویر سی‌تی اسکن جهت Test (۲۰٪) در نظر گرفته می‌شود. در این مرحله دیتاست معتبر SARS-CoV-2 که شامل مجموعه‌ای از داده‌ها از تصاویر سی‌تی اسکن ریه است. در شبیه‌ساز بارگذاری می‌شود.

• مرحله آموزش

مرحله آموزش خود به چند بخش تقسیم می‌شود:

۱. تغییر اندازه همه تصاویر

در این مرحله کلیه تصاویر با سایزهای مختلف خوانده شده و همه به سایز ۱۰۰*۱۰۰ تغییر پیدا کرده و نرمال‌سازی می‌شوند.

۲. ایجاد دو کلاس اصلی

در این مرحله با توجه به داده‌های موجود در دیتاست، دو کلاس سالم و کووید-۱۹ ایجاد می‌شود.

۳. تقسیم داده‌ها

در این مرحله داده تقسیم‌بندی شده و ۸۰٪ برای آموزش و ۲۰٪ تصویر برای اعتبارسنجی استفاده می‌شود که ۱۹۸۵ تصویر سی تی اسکن ریه جهت آموزش و ۴۹۷ تصویر دیگر جهت اعتبارسنجی به کار می‌رود.

۴. ایجاد مدل‌های اصلی

در این مرحله ۳ مدل اصلی آموزش دیده: مدل برای انتخاب بهترین ویژگی‌ها یا کاهش ویژگی‌ها (fs)، مدل برای نرمال‌سازی ویژگی‌ها (trans) و مدل جهت طبقه‌بندی و محاسبه احتمال بیماری کووید-۱۹ (model)، با توجه به ۱۹۸۵ تصویر مورد نظر ایجاد می‌گردد.

۵. آموزش داده‌ها (Train)

در این مرحله، بعد از استخراج ویژگی‌های هر تصویر، ابتدا نیاز است با استفاده از مدل MXGB01-fs.h5 بهترین این ویژگی‌ها انتخاب شوند و بعد توسط مدل MXGB01.trans.h5 نرمال‌سازی می‌شوند، و در آخر این ویژگی‌های نرمال شده به‌عنوان ورودی به مدل XGB (MXGB01.h5) داده شده و احتمال داشتن کووید-۱۹ محاسبه می‌شود.

• مرحله طبقه‌بندی و ارزیابی مدل با داده‌های Test

در این طرح با توجه به نوع تصاویر دیتاست، ۲ نوع کلاس طبقه‌بندی (کووید-۱۹ و سالم) وجود دارد که به‌عنوان خروجی مدل پیشنهادی محسوب می‌گردد. جهت ارزیابی این مدل از شاخص‌های ارزیابی ذیل استفاده می‌گردد. شکل ۹ طرح کامل شبکه یادگیری ماشین پیشنهادی را نشان می‌دهد.

شاخص‌های ارزیابی به شرح زیر است:

TP_ موارد مثبت که به‌درستی طبقه‌بندی شده‌اند.

TN_ موارد منفی که به‌درستی طبقه‌بندی شده‌اند.

FN_ موارد مثبت که به‌طور نادرست طبقه‌بندی شده‌اند.

FP_ موارد منفی که به‌طور نادرست طبقه‌بندی شده‌اند

دقت = این معیار بیانگر دقت کلی طبقه‌بندی می‌باشد و بیانگر نرخ طبقه‌بندی صحیح می‌-

باشد. همانطور که در رابطه (۴) مشاهده می‌نمایید.

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \quad (۴)$$

حساسیت = این معیار بیانگر حساسیت کلی طبقه‌بندی می‌باشد و حاصل تقسیم موارد مثبت واقعی به حاصل جمع موارد مثبت واقعی و موارد منفی کاذب است. همان‌طور که در رابطه (۵) مشاهده می‌نمایید.

$$\text{Sensitivity} = \frac{TP}{TP+FN} \quad (۵)$$

خاصیت = خاصیت حاصل تقسیم موارد منفی واقعی به حاصل جمع موارد منفی واقعی و مثبت کاذب است. همان‌طور که در رابطه (۶) مشاهده می‌نمایید.

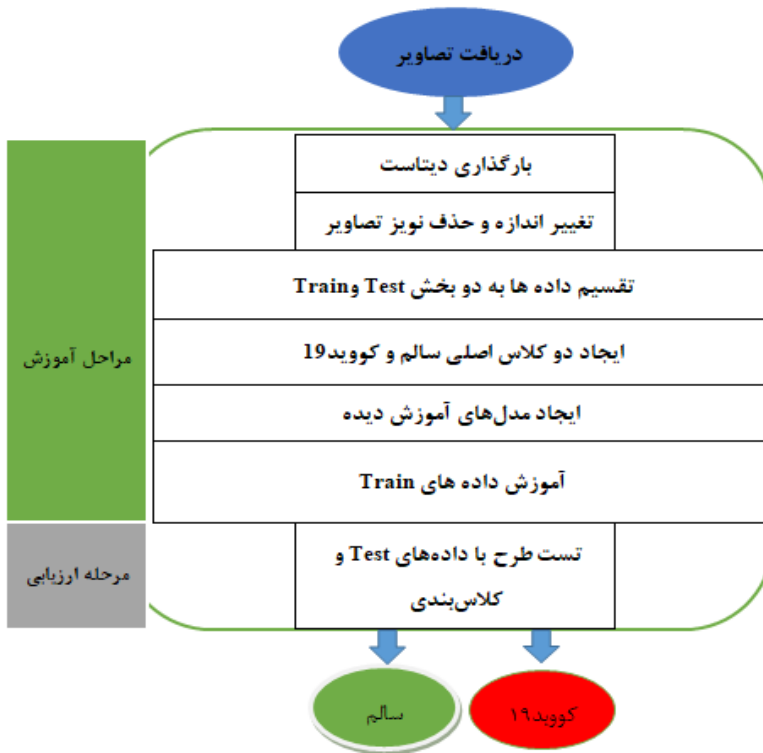
$$\text{Specificity} = \frac{TN}{TN+FP} \quad (۶)$$

درستی = درستی حاصل تقسیم موارد مثبت واقعی به حاصل جمع موارد مثبت واقعی و مثبت کاذب است. همان‌طور که در رابطه (۷) مشاهده می‌نمایید.

$$\text{Precision} = \frac{TP}{TP+FP} \quad (۷)$$

امتیاز = امتیاز F حاصل تقسیم ۲ برابر موارد مثبت واقعی به حاصل جمع ۲ برابر موارد مثبت واقعی و منفی کاذب و همچنین مثبت کاذب است. همان‌طور که در رابطه (۸) مشاهده می‌نمایید.

$$\text{F-score} = \frac{2TP}{2TP+FP+FN} \quad (۸)$$



شکل ۹. شماتیک کامل روش کار طرح پیشنهادی.

یافته‌های پژوهش

در این بخش، به یافته‌ها و ارزیابی طرح پیشنهادی، یعنی تشخیص بیماری کووید-۱۹ با استفاده از پردازش تصاویر سی‌تی اسکن ریه، توسط الگوریتم‌های یادگیری ماشین، می‌پردازیم، در ابتدا نیاز است، مدل پیشنهادی را در نرم‌افزار پایتون پیاده‌سازی و شبیه‌سازی کنیم بنابراین لازم است، فایل test covid.ipynb را توسط محیط اجرایی jupyter نرم افزار آنالوندا باز کنیم و تمام بلاک‌های آن را اجرا نماییم.

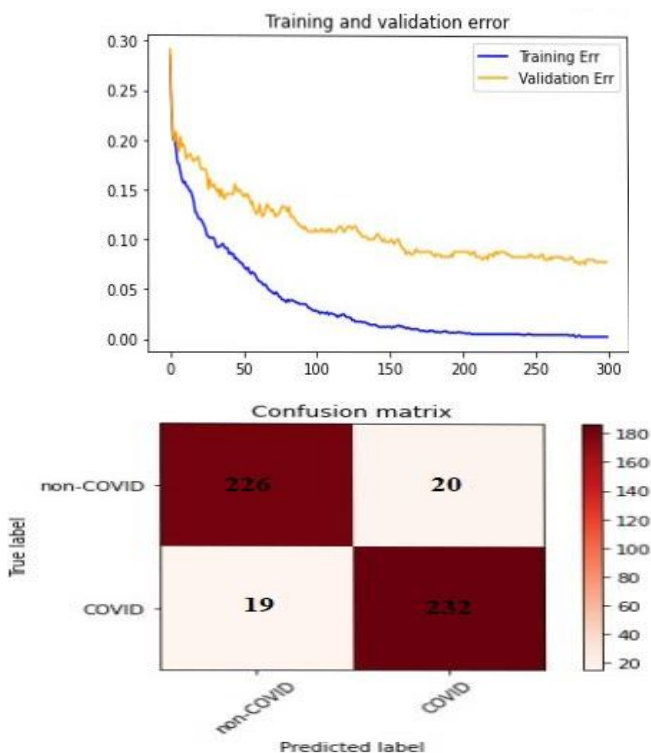
جهت اعتبارسنجی، ارزیابی و به‌دست آوردن پارامترهای دقت، حساسیت و تشخیص مدل پیشنهادی، از ماتریس درهم‌ریختگی استفاده می‌شود. این ماتریس اطلاعاتی را در مورد صحت طبقه‌بندی شامل دو کلاس Covid و Non-Covid را در دوبخش واقعی و پیش‌بینی شده به ما ارائه می‌نماید. جدول درهم‌ریختگی و ماتریس درهم‌ریختگی را در ذیل می‌توان مشاهده نمایید.

جدول ۱. جدول درهم‌ریختگی

پیش‌بینی شده		کلاس‌بندی
سالم	بیمار	دیتاست (واقعی)
۱۹ (FN)	۲۳۲ (TP)	بیمار
۲۲۶ (TN)	۲۰ (FP)	سالم

شکل ۱۰. ماتریس درهم‌ریختگی و کلاس‌بندی.

نتایج حاصل از ماتریس درهم‌ریختگی دقت ۹۲٫۲٪، حساسیت ۹۲٫۴٪، تشخیص ۹۱٫۸٪ و امتیاز F_1 ۹۲٫۳ درصد را در اجرای طرح پیشنهادی، به‌دنبال داشت. همچنین درصد خطاهای مراحل آموزش و ارزیابی را می‌توان در دیاگرام دقت زیر مشاهده نمود.



شکل ۱۱. درصد خطاهای مراحل آموزش و ارزیابی

نتیجه، چالش‌ها و پیشنهادات کارهای آینده

با توجه به اینکه بیماری کووید-۱۹ تبدیل به یکی از عوامل اساسی مرگ‌ومیر در سال‌های جاری، شده است به‌همین اساس یک شبکه تخصصی می‌تواند روش تشخیص مناسبی جهت تشخیص زودهنگام بیماری محسوب شود. در این تحقیق با استفاده از شبکه یادگیری ماشین (استفاده از الگوریتم‌های HOG، GLCM و Chi2) و قدرت آن در کلاس‌بندی (استفاده از الگوریتم XGBoost) و تشخیص بیماری‌ها از ۲۴۸۲ تصویر سی‌تی‌اسکن موجود در دیتاست SARS-CoV-2 به صورت ۸۰ درصد جهت آموزش و ۲۰ درصد جهت تست و ارزشگذاری استفاده شد، که در نهایت دقت ۹۲٫۲٪، حساسیت ۹۲٫۴٪، تشخیص ۹۱٫۸٪ و امتیاز F₁ ۹۲٫۳٪ را به‌دنبال داشت و نشان داد استفاده از این مدل، روشی مناسب جهت تشخیص زودهنگام بیماری کووید-۱۹ است. بنابراین این مدل پیشنهادی قابل‌گسترش و بهبود است و می‌تواند به‌عنوان نرم‌افزاری ساخته شده و در اختیار پزشکان و کادر حوزه سلامت جهت تشخیص بیماری کووید-۱۹ قرار گیرد.

با توجه به بحران بی‌سابقه جهانی بهداشت در سال‌های اخیر (شیوع بیماری‌های مرس، سارس، کووید-۱۹ و توالی ژنومی این بیماری و اخیراً بیماری آبله میمون) و شاهد بودن پدیده^۱ HAARP و گوشزد فیلم‌های هالیوودی (مانند: فیلم شیوع ۲۰۱۱ و ...)، هشدار و زنگ خطری برای سیاست‌گذاران دولت (بخصوص بخش دفاعی)، ایجاد می‌کند. نگاهی گذرا به آمار تعداد قربانیان در اندک سنوات از شیوع این بیماری‌ها بخصوص کووید-۱۹ (آمار رسمی ۱۴۰۱/۲/۲۲، در ایران: ۱۴۱،۱۹۲ نفر و در جهان: ۶،۲۸۶،۷۳۴ نفر) و مقایسه آن با تعداد کشته‌شدگان جنگ تحمیلی (تقریباً ۲۰۰ هزار نفر)، می‌تواند ارزیابی خوبی جهت درک هر چه بهتر جنگ‌های نوین باشد. هر چند جنگ‌های نوین و نوپدید بخصوص جنگ بیولوژیک، دیرینه یا سابقه تاریخی دارند و اکثر کشورهای قدرتمند تعهدنامه‌هایی در رابطه با عدم استفاده از سلاح‌های بیولوژیک امضاء نموده‌اند ولیکن همیشه خط‌هایی بصورت آشکارا و پنهان صورت گرفته است. جنگ‌های بیولوژیک امروزی، با ساختار شکنی در ژنوم DNA و RNA، همانند جنگ‌های هسته‌ای می‌تواند دامنه و وسعت جهانی در پی داشته و خطری برای بقا زیست در این کره محسوب گردد.

فن‌آوری‌های دیجیتال برای حل این مهم کمک‌های قابل‌توجهی به مراقبت‌های بهداشتی و نگرانی‌های درمانی مرتبط نموده‌اند. اکثریت این فن‌آوری‌های جدید، بلاک‌چین^۲،

^۱ High Frequency Active Auroral Research Program

^۲ Blockchain

اینترنت اشیاء^۱ و هوش مصنوعی (یادگیری ماشین و یادگیری عمیق) موفق به پیاده سازی تجزیه و تحلیل داده‌های بزرگ در سراسر تمام بخش‌های زندگی بشر و ارائه راه‌حل‌های بهینه، شده‌اند.

هوش مصنوعی پتانسیل بسیار زیادی دارد آینده هوش مصنوعی در توسعه روبات‌ها و چت‌بات‌ها و ماشین‌های خودمختار برای ضد عفونی، کار مراقبت‌های بهداشتی در بیمارستان‌ها، مناطق صعب‌العبور، ارائه داروها و ملزومات بیماران و همچنین ارائه مراقبت‌های شخصی از آنها، نهفته است. با توجه به چالش‌ها و مسائل مرتبط با مطالعات موجود مانند حریم خصوصی داده‌ها، تنوع الگوی شیوع، تنظیم و شفافیت، و تمایز بین علائم بیماری‌ها، هنوز به زمان و تلاش قابل توجهی جهت درمان بیماری‌های اپیدمی، نیاز است. ما معتقدیم با ترکیب علوم پزشکی، علوم داده‌ها و علوم ارتباطات سیار (مانند: ترکیب سیستم اطلاعات جغرافیایی^۲ با هوش مصنوعی) و همچنین همکاری نزدیک بین گروه‌های مختلف از دولت، صنعت و دانشگاه می‌توانیم پیش‌بینی شیوع، ردیابی، تشخیص و درمان گسترش ویروس، کشف واکسن و دارو را برای این بیماری‌ها را انجام دهیم، همان‌طوری که نمونه کاربردی این تحقیقات را در تشخیص و نظارت خودکار بیماران کووید-۱۹ در کشورهای چین، کانادا و کره جنوبی با استفاده از تکنیک‌های یادگیری عمیق مبتنی بر پردازش تصاویر پزشکی، در سال اخیر شاهد بودیم. امیدواریم که کار ما کاری هرچند کوچک، ولی مقدمه‌ای بر بسیاری از مطالعات جدید در حوزه هوش مصنوعی و پردازش تصویر پزشکی در نبرد با جنگ بیولوژیک محسوب گردد.

¹ Internet of Things

² Geographical Information System

منابع

- [1] Bhargava, Anuja & Bansal, Atul (2021). Novel coronavirus (COVID-19) diagnosis using computer vision and artificial intelligence techniques: a review. <https://doi.org/10.1007/s11042-021-10714-5>
- [2] Ai, T & Yang, Z & Hou, H & Zhan, C & Chen, C & Lv, W & Tao, Q & Sun, Z & Xia, L (2020). Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases. *Radiology* 26: 200642
- [3] Barstugan M, Ozkaya U, Ozturk S (2020). Coronavirus (COVID-19) Classification using CT Images by Machine Learning Methods. arXiv preprint arXiv:2003.09424
- [4] MOHAMMED, AHMED& WANG, CONGCONG& ZHAO, MENG& ULLAH, MOHIB& NASEEM, RABIA& WANG, HAO& PEDERSEN, MARIUS& ALAYA CHEIKH, FAOUZI (2020). Semi-supervised Network for Detection of COVID-19 in Chest CT Scans.
Citation information:DOI 0.1109/ACCESS.2020.3018498, IEEE Access.
- [5] Johns Hopkins University, “Coronavirus COVID-19 global cases by the center for systems science and engineering(CSSE),”<https://coronavirus.jhu.edu/map.html>, 2020, Online. Visited April 2, 2020.
- [6] Bullock, Joseph & Hoffmann Pham, Katherine & Sin Nga Lam, Cynthia & Luengo-Oroz, Miguel& et al (2020)., “Mapping the landscape of artificial intelligence applications against COVID-19,” arXiv preprint arXiv:2003.11336, 2020.
- [7] Bhattacharya, Sweta & Kumar Reddy Maddikunta, Praveen & Pham, Quoc-Viet & Reddy Gadekallu, Thippa & Rama, Krishnan Siva & Lal Chowdhary, Chiranji & Alazab, Mamoun & Md. Piran, Jalil (2020). Sustainable Cities and Society. Deep learning and medical image processing for coronavirus (COVID-19) pandemic: A survey . journal homepage: www.elsevier.com/locate/scs
- [8] Amoff, Abr` & M. D. & Lou, Y& Erginay, A& Clarida, W& Amelon, R& Folk, J. C& et al (2016). Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. *Investigative Ophthalmology & Visual Science*, 57(13), 5200–5206.
- [9] Ai, T & Yang, Z & Hou, H & Zhan, C& Chen, C& Lv, W& et al (2020). Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: A report of 1014 cases. *Radiology*, 200642.
- [10] Asaad Zebari, Dilovan & Mohsin Abdulazeez, Adnan & Qader Zeebaree, Diyar & Shamal Salih, Merdin (2020).A Fusion Scheme of Texture Features for COVID-19 Detection of CT Scan Images. 2020 International Conference on Advanced Science and Engineering (ICOASE) | 978-1-6654-1579-8/20/\$31.00 ©2020 IEEE DOI: 10.1109/ICOASE51841.2020.9436538.
- [11] Barstugan, Mucahid & Ozkaya, Umut & Ozturk, Saban (2020). Coronavirus (COVID-19) Classification using CT Images by Machine Learning Methods .
Corresponding Author: mbarstugan@ktun.edu.tr
- [12] Barstugana, Mucahid & Özkayaa, Umut & Öztürkb, Şaban(2021).Coronavirus (Covid-19) Classification Using CT Images by Machine Learning Methods. Proceedings of RTA-CSIT 2021, May 2021, Tirana, Albania EMAIL: mbarstugan@ktun.edu.tr (A. 1); uozkaya@ktun.edu.tr A. 2); saban.ozturk@amasya.edu.tr (A. 3)

- [13] Morgan Jams, Ronalus & sunyoto, Andi (2021). Ditection of CT-Scan Lungs COVID-19 Image Using Convolutional Neuval Network And CLAHE. 2020 3rd International Conference on Information and Communications Technology (ICOIACT) |978-1-7281-7356-6/20/\$31.00 ©2020 IEEE |DOI: 10.1109/ICOIACT50329.2020.9332069
- [14] Mohammad-Rahimi, Hossein & Nadimi, Mohadeseh & Ghalyanchi-Langeroudi, Azadeh & Taheri, Mohammad & Ghafouri-Fard, Soudeh (2021). Application of Machine Learning in Diagnosis of COVID-19 Through X-Ray and CT Images: A Scoping review. published: 25 March 2021 doi: 10.3389/fcvm.2021.638011
- [15] J. Zhao & Y. Zhang & X. He and P. Xie (2020). "Covid-ct-dataset: a ct scan dataset about covid-19", "arXiv preprint arXiv:2003.13865, 2020.
- [16] M. Loey & F. Smarandache and N. E. M. Khalifa (2020). "A deep transfer learning model with classical data augmentation and cgan to detect covid-19 from chest ct radiography digital images", Preprints, 2020.
- [17] O. Gozes & M. F.-Adar & H. Greenspan & P. D. Browning and H. Zhang & et al (2020). "Rapid AI development cycle for the coronavirus (COVID-19) pandemic: initial results for automated Detection & patient monitoring using deep learning CT image analysis", arXivpreprint arXiv:2003.05037, 2020.
- [18] Y.-H. Wu & S.-H. Gao & J. Mei & J. Xu, and D.-P. Fan & et al (2020). "JCS: An explainable COVID-19 diagnosis system by joint classification and segmentation", arXiv:2004.07054v1, April 2020 .
- [19] <https://www.kaggle.com/c/utkml-covid-detection-from-ct-scans/data>.