

یک مجموعه انباشته‌ی قابل توضیح از مدل های یادگیری عمیق برای بهبود تشخیص سرطان پوست ملانوما

چکیده

ملانومای بدخیم یکی از وحشتناک ترین انواع سرطان پوست است که به دلیل رشد غیرطبیعی سلول های ملانوسیت ایجاد می شود. مدل های شبکه عصبی پیچیده‌ی عمیق (CNN) برای تشخیص خودکار ملانوما از تصاویر پوستی (درموسکوپی) برجسته می شوند. اگرچه به طور باورنکردنی دقیق است، اما ماهیت «جعبه سیاه» مدل های عمیق CNN به دلیل فقدان تفسیرپذیری مناسب، همچنان از استفاده گسترده آنها در محیط های بالینی جلوگیری می کند. این مقاله یک چارچوب مجموعه ای انباشته مبتنی بر CNN قابل شرح را برای تشخیص سرطان پوست ملانوما در مراحل اولیه پیشنهاد می کند. در چارچوب مجموعه ای انباشته، مفهوم یادگیری انتقال در جایی استفاده می شود که چندین مدل فرعی CNN که وظیفه طبقه بندی یکسانی را انجام می دهند، مجموع و مونتاژ می شوند. یک مدل جدیدی به نام فرا یادگیرنده از تمام پیش بینی های مدل های فرعی استفاده می کند و نتایج پیش بینی نهایی را تولید می کند. این مدل با استفاده از یک مجموعه داده دسترسی باز که شامل هر دو از تصاویر ملانومای خوش خیم و بدخیم است، ارزیابی می شود. یک روش توضیح پذیر با توضیحات تطبیقی شکل یافته برای تولید نقشه های حرارتی توسعه یافته است که مناطقی از تصاویر ملانوما را که بیشتر نشان دهنده بیماری هستند، تجسم می کند. این روش، تفسیرپذیری تصمیم مدل ما را به شیوه ای قابل درک برای متخصصان پوست فراهم می کند. نتایج ارزیابی اثربخشی مدل مجموعه ای ما را با درجه بالایی از دقت (95.76%)، حساسیت (96.67%) و AUC (0.975) نشان می دهد.

کلمات کلیدی: ملانوما · سرطان پوست · مدل مجموعه ای انباشته · یادگیری عمیق قابل توضیح · تصاویر درموسکوپی

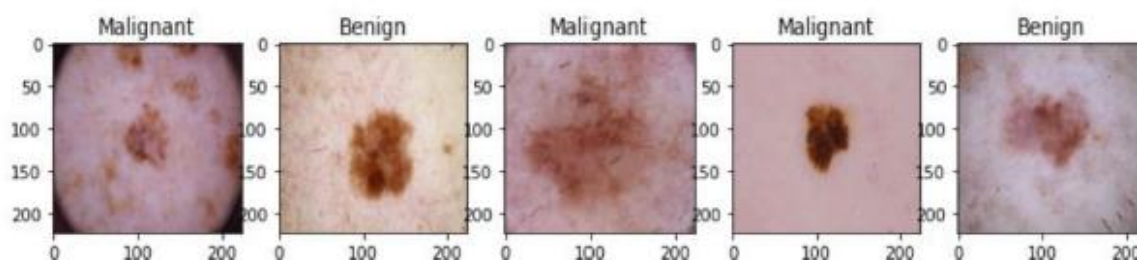
1. معرفی

ملانومای بدخیم که به عنوان ملانوما نیز شناخته می شود، یکی از مخوف ترین انواع سرطان پوست در نظر گرفته می شود. معمولاً در اثر قرار گرفتن طولانی مدت در معرض اشعه فرا بنفش (UV) ایجاد می شود، که باعث جهش در سلول های حاوی رنگدانه ملانین به نام ملانوسیت ها می شود [1]. هنگامی که اشعه فرا بنفش وارد ملانوسیت ها می شود، باعث ترشح بیش از حد میزان ملانین و در نتیجه ایجاد خال های تیره می شود. این خال ها یا ضایعات پوستی می توانند به مرور زمان غیر قابل کنترل شوند و می تواند شکل تومورهای سرطانی را به خود بگیرد و در نهایت می تواند به سرعت به سایر قسمت های بدن گسترش یابد. طبق گزارش انجمن سرطان آمریکا [2]، حدود 106110 مورد جدید سرطان پوست ملانوما در سال 2021 در ایالات متحده آمریکا (USA) تخمین زده شده است و انتظار می رود که حدود 7180 نفر در سال 2021 جان خود را از دست بدهند. آنها همچنین پی بردند که نرخ بروز فعلی حدود 23 در 100000 در ایالات متحده است. با این حال، تشخیص زودهنگام و درمان مناسب ملانوما می تواند این آمار طاقت فرسا را کاهش دهد. مطالعات تجربی نشان می دهد که ملانومای بدخیم در صورت تشخیص در مراحل اولیه قابل درمان است [3، 4]. بدین ترتیب، طبقه بندی ضایعات خوش خیم و بدخیم ملانوما برای تشخیص زودهنگام اینکه آیا تصویری از پوست آسیب دیده حاوی سرطان ملانوما است یا خیر، بسیار حیاتی است. برخی از نمونه ضایعات ملانومای خوش خیم و بدخیم در شکل 1 نشان داده شده است. در گذشته اخیر تلاش های تحقیقاتی زیادی برای توسعه تکنیک های کمک کننده ی کامپیوتری برای تجزیه و تحلیل ضایعات پوستی صورت گرفته است. برخی از تکنیک های پرکاربرد از قاعده ای استفاده می کنند که مبتنی بر عدم تقارن، بی نظمی مرزی، رنگی که یکنواخت نیست

و ساختارهای درماتوسکوپی (ABCD) است. متخصصان پوست معمولاً از این قانون برای تشخیص سرطان پوست استفاده می‌کنند.

در واقعیت، تصاویر ملانوما نشان‌دهنده‌ی کامل یک تغییر هستند که شامل نويز، تضاد پایین و بی‌نظمی مرزی است که تشخیص بیماری را از تمام تصاویر ملانوما چالش‌برانگیز می‌کند. به گفته بنیاد تحقیقاتی ملانوما، تشخیص دقیق ملانوما با استفاده از دستورالعمل‌های ABCD از تمام تصاویر همیشه امکان‌پذیر نیست.

علاوه بر این، این روش‌ها به متخصصان پوست با تجربه بالا نیاز دارند تا احتمال تشخیص اشتباه را کاهش دهند و در نتیجه موارد false positive و false negative زیاد باشد [5].



شکل 1: تصویر برخی از ضایعات خوش خیم و بدخیم

این محدودیت‌ها تقاضا برای کشف روش‌های قوی و هوشمند برای تشخیص خودکار ملانوما را افزایش می‌دهد که خدمات سریعی را به پزشکان خط مقدم که هم مبتدی و هم متخصص در زمینه پوست هستند، ارائه می‌کند. از این رو، الگوریتم‌هایی برای تشخیص خودکار ملانوما شروع به ظهور کردند. به طور خاص، الگوریتم‌هایی که می‌توانند ویژگی‌هایی را از تصاویر درماتوسکوپی (پوستی) استخراج کنند، با ارائه غربالگری خودکار، به طور فزاینده‌ای در حمایت از متخصصین بالینی در مرحله تشخیص محبوب شده‌اند. برای این منظور، دو الگوی یادگیری ماشین مشترک هستند. اول، تکنیک‌های طبقه‌بندی سنتی بر اساس ویژگی‌های استخراج شده از تصاویر درماتوسکوپی کار می‌کنند که اغلب به دست آوردن آنها دشوار است و به دانش دامنه نیاز دارند [6].

دوم، تکنیک‌های یادگیری عمیق [7]، به ویژه شبکه‌های عصبی کانولوشن که می‌توانند به طور خودکار ویژگی‌ها را از تصاویر ورودی برای طبقه‌بندی بیماری استخراج کنند. در تلاش اخیر اشرف و همکاران. [8] یک سیستم مبتنی بر منطقه مورد علاقه (ROI) را برای تشخیص ملانوما با استفاده از مدل‌های CNN عمیق off-the-shelf پیشنهاد کرد. ROI از تصاویر درماتوسکوپی با استفاده از یک الگوریتم K-mean بهبود یافته استخراج شده است. نویسندگان از مدل AlexNet با تصاویر ROI افزوده شده برای شناسایی دقیق سرطان پوست ملانوما استفاده می‌کنند. نتایج ارزیابی، کارایی پیشرفته از مدل پیشنهادی را نشان داده است. برخی مطالعات دیگر [9-11] از تکنیک‌های یادگیری عمیق مبتنی بر تقسیم‌بندی برای طبقه‌بندی ضایعات پوستی ملانوما استفاده کرده‌اند. در حالی که رویکردهای یادگیری عمیق مزیت خود را نسبت به سایر تکنیک‌های تشخیص ملانوما نشان داده‌اند، اکثر این رویکردها یافته‌های خود را با تفسیرپذیری کافی از مدل‌های مرتبط با ویژگی‌های مرتبط از علائم آسیب‌شناسی در تصاویر ملانوما پشتیبانی نمی‌کنند. بنابراین، کارایی بالینی این تکنیک‌ها تا زمانی که مطالعات بیشتری برای تفسیر ویژگی‌های سطح بالا استخراج شده از این مدل‌ها انجام نشود، نامشخص است. بسیار بعید است که متخصصان پوست در زندگی واقعی یک مدل طبقه‌بندی جعبه سیاه را حتی با نتایج تجربی بسیار دقیق بپذیرند. علاوه بر این، روش‌های موجود که می‌توانند سرطان پوست ملانوما را تشخیص دهند، دقت بالایی ندارند. این مقاله یک چارچوب مجموعه‌ی انباشته یادگیری عمیق مبتنی بر CNN را برای تشخیص سرطان پوست ملانوما در مراحل اولیه پیشنهاد می‌کند. در چارچوب مجموعه انباشته، مفهوم یادگیری انتقال در جایی استفاده می‌شود که چندین مدل پایه CNN که وظیفه طبقه‌بندی یکسانی را انجام می‌دهند، مونتاژ می‌شوند. مدل جدیدی به نام فرا یادگیرنده (meta-learner) این پیش‌بینی‌ها از تمام مدل‌های

پایه را استفاده می‌کند و پیش‌بینی خود را تولید می‌کند. این فرآیند به عنوان تعمیم انباشته [12، 13] شناخته می‌شود، که در ابتدا برای به حداقل رساندن میزان خطای تعمیم برای تعمیم دهنده های متعدد پیشنهاد شد.

از این رو، مجموعه انباشته چند مدل پایه CNN (ترجیحاً عمیق و گسترده) از قدرت مدل‌های مختلف در یک وظیفه ی طبقه‌بندی استفاده می‌کند و نتایج پیش‌بینی پیشرفته‌تری نسبت به هر مدل شخصی مورد استفاده در مجموعه را تولید می‌کند. ما در ابتدا پنج مدل CNN آموزش دیده با وزن‌های تنظیم‌شده به نام‌های VGG-16، ResNet50، DenseNet121، Xception و EfficientNetB0 را داریم و سپس سه مدل از بهترین مدل‌های پایه را انتخاب کرده‌ایم که با استفاده از مدل فرایادگیرنده برای یافتن نتیجه طبقه‌بندی نهایی روی هم انباشته شده‌اند. ما از مجموعه داده‌های در دسترس عمومی Kaggle از ضایعات پوستی که از طریق بایگانی ISIC (مجموعه بین‌المللی تصاویر پوست) به دست آمده‌اند، برای آموزش و اعتبارسنجی مدل مجموعه انباشته خود استفاده کرده‌ایم [14]. مجموعه داده به ترتیب شامل 1497 و 1800 تصویر خال بدخیم و خوش خیم است. در نهایت، برای تولید نقشه‌های حرارتی که مناطقی از تصاویر ملانوما را که بیشتر نشان‌دهنده خال‌های بدخیم و خوش‌خیم هستند، تجسم می‌کند، یک روش قابل توضیح است که با استفاده از تکنیک توضیحات انطباقی شکل‌یافته (SHAP) توسعه یافته است. رویکرد ما می‌تواند ویژگی‌های ارزشمندی را از مجموعه داده‌های محدود یاد بگیرد و ابزارهای موثر و قابل توضیحی برای تشخیص سرطان پوست ملانوما برای مزایای بالینی بالقوه فراهم می‌کند. به طور خلاصه، این مقاله مشارکت‌های زیر را ارائه می‌دهد:

- ما یک چارچوب مجموعه ای انباشته مبتنی بر CNN را برای تشخیص سرطان پوست ملانوما در مراحل اولیه پیشنهاد می‌کنیم.
- ما از مفهوم یادگیری انتقال استفاده می‌کنیم که در آن چندین مدل فرعی CNN که وظیفه طبقه بندی یکسانی را انجام می‌دهند مونتاژ می‌شوند.
- یک روش توضیح‌پذیر با توضیحات انطباقی شکل‌یافته برای تولید نقشه‌های حرارتی توسعه یافته است که مناطقی از تصاویر ملانوما را که بیشتر نشان‌دهنده بیماری هستند، تجسم می‌کند.
- یک ارزیابی کارایی گسترده از مدل‌های مطالعه شده ارائه شده است. نتایج نشان می‌دهد که مدل مجموعه انباشته ما با درجه بالایی از دقت (95.76٪)، حساسیت (96.67٪) و AUC(0.957) است.

ساختار باقی مانده مقاله به شرح زیر است:

بخش 2 ادبیات مرتبط را در زمینه سرطان پوست ملانوما ارائه می‌دهد.

روش شناسی در **بخش 3** ارائه شده است.

بخش 4 نتایج ارزیابی دقیق و بحث را ارائه می‌دهد.

در نهایت، نتیجه گیری در **بخش 5** با کار احتمالی آینده ارائه شده است.

2 مطالعات ادبیات مقاله

در گذشته اخیر، تلاش‌های تحقیقاتی قابل توجهی برای توسعه سیستم‌های کمک کامپیوتری برای کاربردهای مراقبت‌های بهداشتی وجود داشته است [15-18]. مخصوصاً، پیشرفت‌ها در شبکه‌های عصبی پیچیده ی عمیق (CNN) اثربخشی خود را در کارهای طبقه‌بندی و تشخیص اشیا در برنامه‌های پردازش تصویر پزشکی ثابت کرده‌اند [19-21]. مطالعات مختلفی با هدف شناسایی سرطان پوست ملانوما با استفاده از تکنیک‌های یادگیری ماشینی و یادگیری عمیق در ادبیات این مقاله ارائه شده‌اند.

Kassani و همکاران [24] یک مطالعه مقایسه ای برای تشخیص ملانوما با استفاده از پنج مدل یادگیری انتقال انجام دادند. کارهای پیش پردازش مختلفی برای افزایش کیفیت تصویر و اطمینان از بهبود تعمیم پذیری مدل اعمال شد. علاوه بر این، نویسندگان به مشکل عدم تعادل کلاس در مجموعه داده با استفاده از تکنیک‌های مختلف افزایش داده دست یافتند. نتایج ارزیابی نشان می‌دهد که ResNet50 بالاترین عملکرد را در تشخیص ملانوما با دقت طبقه‌بندی و F1-score به ترتیب با 92/08 و 92/74 درصد را به دست می‌آورد. تشخیص خودکار بی نظمی در مرز ضایعه پوستی توسط Ali و همکاران پیشنهاد شد [25]. نویسندگان یک رویکرد تقسیم بندی برای استخراج ضایعه پوستی ایجاد کردند و سپس از آشکارساز لبه Canny برای تشخیص مرز آن استفاده کردند. یک بردار نشان دهنده بی نظمی مرز ضایعه پوستی است. در نهایت، آنها از CNN و مجموعه Gaussian naïve Bayes برای تشخیص خودکار ناهنجاری‌های مرزی ضایعات پوستی استفاده می‌کنند. نتایج ارزیابی با استفاده از مجموعه داده‌های همکاری بین‌المللی تصویربرداری پوست (ISIC2018) نشان می‌دهد که رویکرد پیشنهادی به دقت 93.6 درصد با دایره بالایی از حساسیت و ویژگی در تشخیص بی‌نظمی‌های مرزی ضایعه دست می‌یابد.

Tan و همکاران [26] یک الگوریتم اصلاح شده بهینه سازی ازدحام ذرات (PSO) به نام یادگیری ترکیبی (HLPSSO) را در تلاش برای تشخیص ملانوما با به کارگیری تقسیم بندی و طبقه بندی ضایعات پوستی پیشنهاد کردند. علاوه بر این، آنها تقسیم بندی ضایعه را با استفاده از الگوریتم خوشه بندی K-means بهبود بخشیدند. علاوه بر این، دو روش طبقه بندی بر اساس HLPSSO ابداع شد. هر دو روش نتایج برتری را در تقسیم بندی و طبقه بندی ضایعه نشان می‌دهند. نتایج عملکرد با استفاده از یک مجموعه داده ترکیبی نشان می‌دهد که روش پیشنهادی بهتر از تکنیک‌های مختلف PSO پیشرفته و جستجوی کلاسیک در ادبیات مقاله عمل می‌کند.

Viriri و Adegun [27] از یک شبکه رمزگذار-رمزگشا برای استخراج ویژگی‌های کارآمد و طبقه بندی ضایعات پوستی ملانوما و غیر ملانوما استفاده کردند. شبکه پیشنهادی یک رویکرد چند مرحله‌ای را برای انجام طبقه‌بندی ضایعات ملانوما در سطح پیکسل به کار می‌گیرد. این روش برای تجزیه و تحلیل ضایعات پوستی که چالش برانگیز هستند و دارای مرزهای فازی و بافت‌های ناهمگن هستند، مفید است. دو مجموعه داده درموسکوپ (پوستی) مختلف در دسترس عموم، سمپوزیوم بین‌المللی در تصویربرداری زیست پزشکی (ISBI 2017) و بیمارستان پدرو هیسپانو (PH2)، برای ارزیابی عملکرد روش پیشنهادی استفاده شده است. نویسندگان بالاترین دقت و ضریب تاس را به ترتیب 95% و 93% بدست آوردند.

Sarker و همکاران [28] یک شبکه پیچیده باقیمانده قابل تفکیک عمیق را برای تشخیص ملانوما از تصاویر درموسکوپ پیشنهاد کردند. کارهای پیش پردازش گسترده ای مانند اعمال یک فیلتر میانگین غیر محلی و یکسان سازی هیستوگرام تطبیقی محدود کنتراست بر روی تصاویر برای حذف نویز و بهبود تصاویر انجام شدند. نویسندگان از نقشه فعال سازی کلاس با وزن گردان برای فراهم کردن توضیح پذیری مدل، استفاده کرده‌اند. نتایج ارزیابی با استفاده از مجموعه داده‌های چندگانه، اثربخشی روش پیشنهادی را نشان می‌دهد که به حداکثر دقت 99.5 درصد در مجموعه داده ISIC2018 دست می‌یابد.

Banarjee و همکاران [29] یک روش مبتنی بر یادگیری عمیق را با استفاده از الگوریتم You Only Look Once (YOLO) برای تشخیص ملانوما از تصاویر دیجیتالی و درموسکوپ پیشنهاد کردند. سهم قابل توجه این مطالعه از معرفی تقسیم بندی دو فازی با استفاده از نظریه گراف و تقریب مبتنی بر اعداد فازی ناشی می‌شود. سه مجموعه داده مختلف در دسترس عموم برای ارزیابی مدل پیشنهادی استفاده شده است. امتیاز Jaccard 79.84% در مجموعه داده ISIC2019 به عنوان عملکرد تقسیم‌بندی به دست آمده است. نویسندگان همچنین موفق به دستیابی به دقت و ویژگی به ترتیب 97.11% و 98.13% برای طبقه بندی تصاویر ملانوما با استفاده از مجموعه داده یکسان شدند.

Filali و همکاران [30] در تلاشی نسبتاً اخیر، یک تکنیک ترکیبی را برای تشخیص سرطان پوست ملانوما معرفی کردند. آنها ویژگی‌های دست ساز و ویژگی‌های به دست آمده از شبکه‌های عصبی پیچیده از قبل آموزش دیده را با یکدیگر ترکیب کردند. سپس از مهندسی ویژگی برای کنار گذاشتن ویژگی‌های نامربوط استفاده شده است. نویسندگان مدل ترکیبی خود را با استفاده از مجموعه داده‌های کوچک و بزرگ ارزیابی کردند و دقت 87.8 و 98 درصد را در مجموعه داده‌های PH2 و ISIC2019 به دست آوردند.

Pham و همکاران [9] تشخیص خودکار ملانوما و تقسیم بندی ضایعه پوستی را با استفاده از شبکه CNN عمیق و XGBoost پیشنهاد کردند. آنها DenseNet-161، ResNet-152 و Inception-v4 را برای طبقه بندی ملانوما آموزش

داده اند و از U-Net و U-Net با VGG-166 برای تولید ماسک های تقسیم بندی استفاده کرده اند. نتایج ارزیابی با استفاده از مجموعه داده ISIC2017 نشان می دهد که DenseNet-161 در طبقه بندی ملانوما با توجه به (0.867) AUC و دقت (0.863) از سایر CNN ها بهتر عمل می کند.

نویسندگان در [31] یک روش ترکیبی برای طبقه بندی ضایعات پوستی ملانوما با استفاده از ویژگی های دست ساز و یادگیری عمیق ارائه کردند. سپس ویژگی های ترکیب شده به چندین طبقه بندی کننده ها مانند رگرسیون خطی (LR)، ماشین های بردار پشتیبان (SVM) و ماشین های بردار مربوطه (RVM) برای طبقه بندی نهایی ضایعات پوستی داده می شوند. نویسندگان چارچوب پیشنهادی را با استفاده از مجموعه داده ISIC 2018 در دسترس عموم آزمایش کردند و در مقایسه با بسیاری از مطالعات موجود در ادبیات مقالات، به عملکرد بهبود یافته ای دست یافتند.

Daghrir و همکاران [32] یک رویکرد ترکیبی را برای تشخیص سرطان پوست ملانوما با استفاده از یک شبکه عصبی پیچیده و دو طبقه بندی کننده یادگیری ماشین پیشنهاد کردند. این مدل ها با انواع مختلفی از ویژگی ها مانند رنگ، بافت و شکل ضایعات پوستی آموزش دیده اند. در پایان، برای دستیابی به عملکرد بهتر از یک روش مجموعه ای مبتنی بر رای اکثریت استفاده می شود.

در تلاشی دیگر، Hossin و همکاران [33] یک رویکرد مبتنی بر CNN با تنظیم کننده پیشرفته برای تشخیص ملانومای بدخیم پیشنهاد کردند. نویسندگان از یک معماری چندلایه CNN از ابتدا با تکنیک های منظم سازی متعدد مانند حذف و نرمال سازی دسته ای استفاده کردند. نتایج عملکرد با استفاده از مجموعه داده ای متشکل از 3297 تصویر درموسکوپی نشان داده است که روش پیشنهادی به دقت 93.5 درصد در تشخیص ملانومای بدخیم دست یافته است.

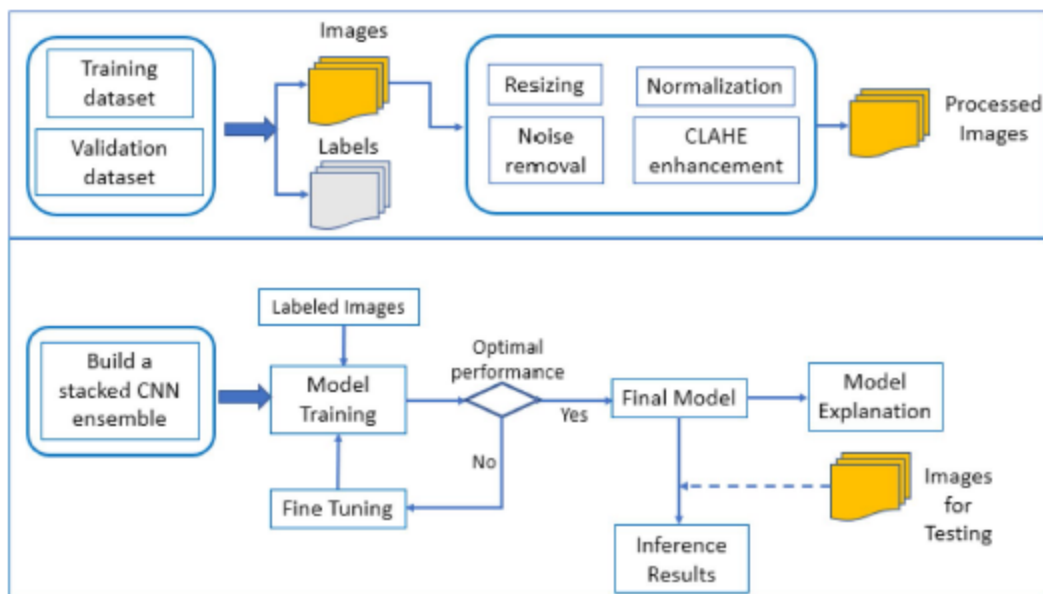
Bisla و همکاران [34] یک تکنیک خالص سازی و تقویت داده را برای شناسایی ملانوما با استفاده از تکنیک یادگیری عمیق ارائه کردند. تکنیک پیشنهادی بر محدودیت مطالعات یادگیری عمیق موجود که امکان اجرایی شدن آن به شدت توسط مجموعه داده های کوچک و بسیار نامتعادل به خطر افتاده است، غلبه می کند. نویسندگان از یک پیش پردازش و یک واحد تولید داده برای حذف انسداد تصویر و تولید تصاویر برای کلاس های ضایعه کمیاب استفاده کردند. نتایج تجربی نشان داد که ادغام این دو واحد تشخیص ملانوما را با عملکرد برتری در مقایسه با روش های خط پایه موجود تقویت می کند.

علاوه بر این مقالات، برخی تلاش های دیگر با استفاده از مدل های CNN سفارشی بهینه سازی شده به مشکل تشخیص سرطان ملانوما پرداختند [35، 36].

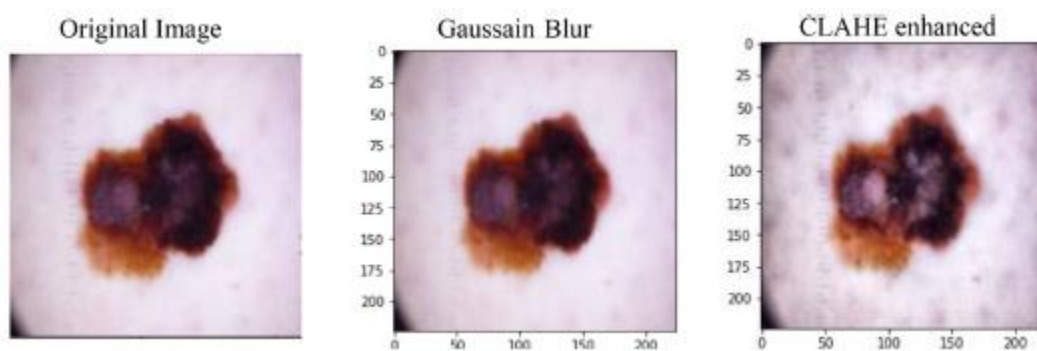
3 اصول و روش ها

این بخش سیستم پیشنهادی را تشریح می کند. شکل 2 یک بررسی اجمالی مختصر از روش پیشنهادی را برای تشخیص سرطان پوست ملانوما نشان می دهد. به عنوان گام اولیه، تصاویر خال های پوستی ملانوما از قبل پردازش می شوند تا کیفیت آنها با استفاده از تکنیک های حذف نویز و افزایش کنتراست بهبود یابد. بعد از آن، ویژگی های یادگیری عمیق از تصاویر پردازش شده با استفاده از مدل های پایه انتخاب شده CNN که بر روی مجموعه داده ImageNet از قبل آموزش داده شده اند استخراج می شوند. نتایج پیش بینی از معماری های پایه CNN هستند، پس برای طبقه بندی نهایی تصاویر ملانوما به یک فرایادگیرنده داده می شود. شکل 3 نمای کلی معماری شبکه مجموعه انباشته پیشنهادی را نشان می دهد. طبقه بندی کننده های پایه و فرایادگیرنده بر روی مجموعه داده های ISIC متشکل از تصاویر ضایعات پوستی خوش خیم و بدخیم آموزش دیده اند.

در نهایت، مدل گروه انباشته آموزش دیده چند سری برای پیش بینی ملانومای خوش خیم یا بدخیم از تصاویر دیده نشده استفاده شده است. بقیه بخش به تفصیل هر مرحله از سیستم مجموعه پیشنهادی را توضیح می دهد.



شکل 2: روند کار روش پیشنهادی برای پیش پردازش تصویر (بالا) و آموزش مدل و اعتبارسنجی (پایین)



شکل 3: مراحل پیش پردازش یک تصویر نمونه

3.1 شرح مجموعه داده و پیش پردازش

این مطالعه از مجموعه داده‌های عمومی در دسترس Kaggle از ضایعات پوستی به دست آمده از آرشیو ISIC (مجموعه بین‌المللی تصاویر پوست) برای آموزش و اعتبارسنجی مدل مجموعه انباشته ما استفاده کرده است. مجموعه داده شامل 1497 و 1800 تصویر خال بدخیم و خوش خیم است [14]. تصاویر جمع آوری شده ضایعات پوستی دارای نویزها و مصنوعات هستند.

بنابراین، ما برخی از وظایف استاندارد پیش پردازش مانند تغییر اندازه، عادی سازی، حذف نویز و افزایش کنتراست را اعمال کرده ایم. همه تصاویر به ابعاد یکنواخت 224×224 تغییر مقیاس داده می شوند و برای تغییر دامنه مقادیر شدت

پیکسل در $[0,1]$ نرمال می شوند. ما از روش تاری گاوسی برای کاهش میزان نویز در تصاویر ضایعه پوستی ورودی استفاده کردیم. این را می توان با پیچیدگی تصویر با یک فیلتر گاوسی انجام داد که به عنوان یک فیلتر پایین گذر برای حذف اجزای فرکانس بالا از تصویر عمل می کند. فیلتر گاوسی بر اساس تابع گاوسی در فضاها یک و دو بعدی به شرح زیر عمل می کند:

$$G_{1D}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\left(\frac{x^2}{2\sigma^2}\right)}, \quad (1)$$

$$G_{2D}(x, y) = \frac{1}{2\pi\sigma^2} e^{-\left(\frac{x^2+y^2}{2\sigma^2}\right)}, \quad (2)$$

در معادله های بالا که σ انحراف معیار تابع گاوسی است. درجه هموارسازی تصویر متناسب با مقدار σ است. مختصات دکارتی یک تصویر با (x,y) نشان داده می شود که بعد پنجره را مرحله بندی می کند. بنابراین، واضح است که تابع گاوسی به صورت دو بعدی قابل تفکیک است، که آن را به طور مقایسه ای سریعتر از سایر تکنیک های محو کردن می کند.

ما از یکسان سازی هیستوگرام تطبیقی با کنتراست محدود استفاده کرده ایم که به اختصار (CLAHE) نامیده میشود، تا کنتراست تصویر را با کشش توزیع پیکسل ها بهبود دهیم. CLAHE یک نوع تطبیقی از یکسان سازی هیستوگرام استاندارد (HE) است. HE معمولی کنتراست سراسری یک تصویر را در نظر می گیرد و از تبدیل یکسانی برای تبدیل تمام پیکسل های تصویر استفاده می کند. فرض کنید m تصویری باشد که با پیکسل هایی با مقادیر شدت از رنج 0 تا 255 نمایش داده می شود. فرض کنید h_i نشانگر هیستوگرام نرمال شده m باشد که حاوی یک bin با شدت هر پیکسل به صورت i است. علاوه بر این، فرض کنید n_i و N تعداد پیکسل های با شدت i و تعداد کل پیکسل های تصویر را نشان دهند، ما h_i را به صورت زیر به دست می آوریم:

$$h_i = \frac{n_i}{N}, i = 0, 1, 2, \dots, 255. \quad (3)$$

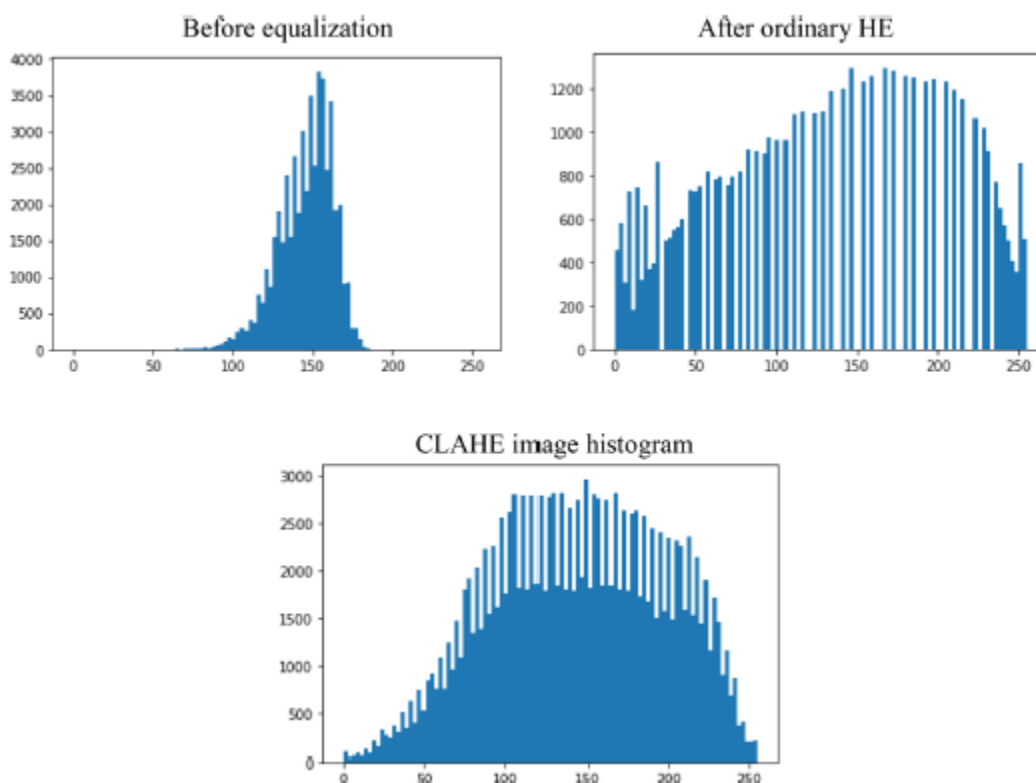
حال، تصویر با تساوی هیستوگرام m_{he} ، می تواند به صورت زیر تعریف شود:

$$m_{he}(r, c) = \left\lfloor 255 * \sum_{i=0}^{m_{r,c}} h_i \right\rfloor. \quad (4)$$

این تبدیل را می توان معادل تبدیل شدت های پیکسلی i از m توسط تابع زیر در نظر گرفت:

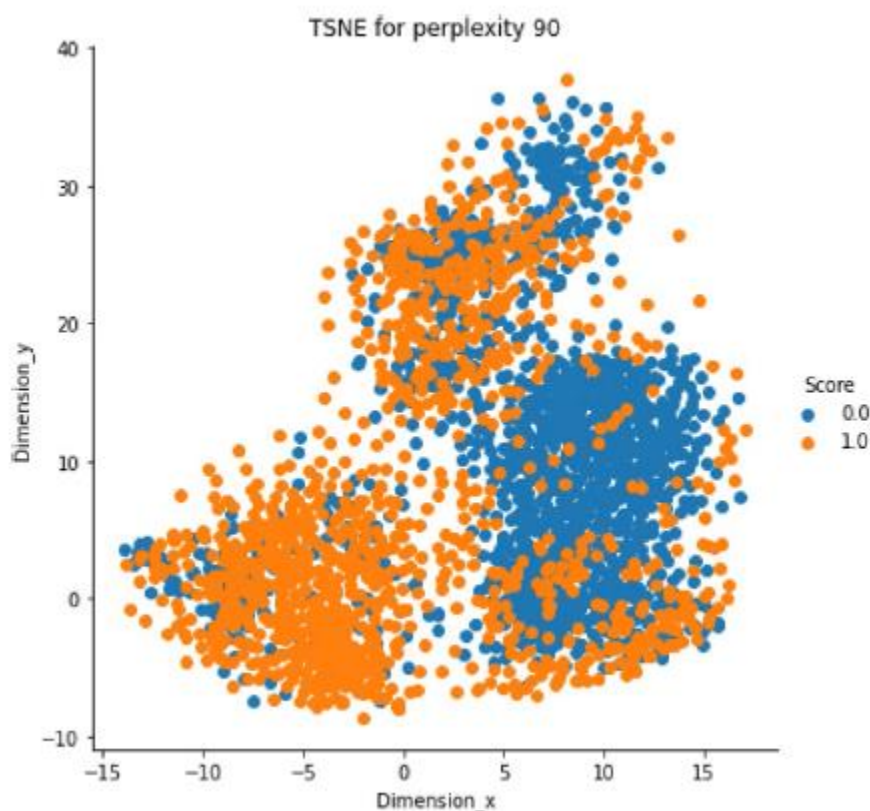
$$F(i) = \left\lfloor 255 * \sum_{j=0}^i h_j \right\rfloor. \quad (5)$$

توجه به این نکته مهم است که HE معمولی زمانی به خوبی کار می کند که کل تصویر دارای مقادیر پیکسل توزیع مشابه باشد. با این حال، اگر تصویر تغییر روشنایی قابل توجهی را در بخش های مختلف آن نشان دهد، احتمالاً این تکنیک به طور مؤثر کار نخواهد کرد. یکسان سازی هیستوگرام تطبیقی (AHE) این مشکل را با تولید هیستوگرام های متعدد که زیر هیستوگرام نیز نامیده می شوند، برای اهمیت دادن بر افزایش کنتراست محلی، برطرف می کند. AHE از همان اشتقاق توابع تبدیل برای هیستوگرام ها به عنوان HE معمولی پیروی می کند. با این حال، استفاده از AHE محدود است زیرا کنتراست را برای یک منطقه تقریباً ثابت در تصویر بیش از حد تقویت می کند، در نتیجه نویز را برای آن منطقه خاص تقویت می کند. CLAHE نشان دهنده گونه ای از AHE است که تقویت نویز را کنترل می کند. تصویر را به قسمت های کوچکتر تقسیم می کند و تقویت کنتراست محدود در هر یک از این مناطق انجام می شود و از افزایش نویز جلوگیری می کند. سپس این مناطق کوچکتر با استفاده از درون یابی دوخطی برای حذف مرزهای منطقه ادغام می شوند. شکل 3 تصویر اصلی و تصویر مربوطه را پس از اعمال Gaussian blur و CLAHE نشان می دهد. همچنین، هیستوگرام یک تصویر اصلی قبل و بعد از افزایش کنتراست در شکل 4 نشان داده شده است.



شکل 4: هیستوگرام تصویر اصلی قبل و بعد از افزایش

ما همچنین از روش t-SNE (t-Distributed Stochastic Neighbor Embedding) که مربوط به مقاله ی [37] است، برای تجسم تصاویر ضایعه پوستی به صورت دوبعدی استفاده می کنیم تا درک کنیم که جداسازی آنها به کلاس های ملانوما خوش خیم و بدخیم مربوطه چقدر دشوار است برای دستیابی به آن، تصاویر را صاف می کنیم تا یک نمایش ویژگی پیدا کنیم. پس از تنظیم هایپر پارامتر گجی، مشاهده می کنیم که طبقات تصویر بر اساس ویژگی های استخراج شده قابل تفکیک هستند، همانطور که در شکل 5 نشان داده شده است.

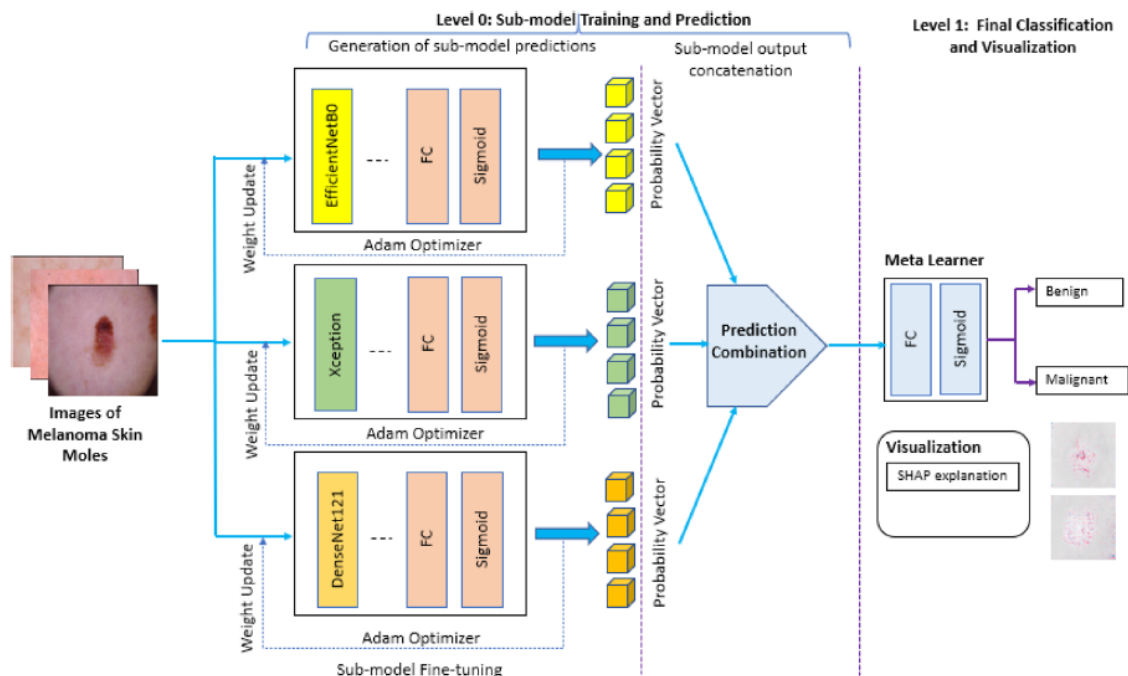


شکل 5: تجسم (t-SNE) کلاس های ملانومای خوش خیم و بدخیم

3.2 مدل مجموعه انباشته پیشنهادی

شکل 6 معماری مجموعه انباشته پیشنهادی از مدل های CNN را برای تشخیص خودکار ملانومای بدخیم نشان می دهد. در ابتدا، احتمالات پیش بینی از تصاویر ضایعه پوستی ملانوما با استفاده از سه مدل فرعی CNN با تنظیم دقیق مختلف به نام های EfficientNetB0 [22]، DenseNet121 [38] و Xception [39] تولید می شوند که در سطح 0 روی هم انباشته شده اند.

سپس، پیش بینی های تولید شده از این شبکه ها را ترکیب می کنیم و آنها را برای طبقه بندی نهایی تصاویر ملانوما به یک فرا یادگیرنده در سطح 1 می دهیم. در نهایت، توضیح پذیری مدل با تولید تجسم های نقشه حرارتی ارائه می شود که مناطقی از تصاویر ضایعات پوستی را که بیشتر نشان دهنده انواع ملانوما هستند، شناسایی می کند.



شکل 6: معماری مجموعه انباشته پیشنهادی از مدل های CNN

چارچوب مجموعه ی ما از سه مدل CNN از پیش آموزش دیده تنظیم شده مختلف (بر روی مجموعه داده ImageNet) برای استخراج ویژگی از تصاویر فوندوس (fundus) ورودی استفاده کرده است.

در ابتدا، Efficient-NetB0 [22] مورد استفاده قرار می گیرد که یک روش مقیاس بندی مدل جدید را برای افزایش مقیاس CNN با استفاده از یک ضریب ترکیب موثر ارائه می دهد. این روش به جای سایر رویکردهای سنتی که به طور دلخواه هر بعد شبکه را مقیاس بندی می کنند، هر بعد شبکه را با مجموعه ای ثابت از ضرایب مقیاس بندی به طور یکنواخت مقیاس می دهد. معماری به دست آمده از پیچیدگی گلوگاه معکوس موبایل (MBConv) مشابه MobileNetV2 استفاده می کند، اما دارای اندازه FLOP کمی افزایش یافته است.

دومین مدل CNN مورد استفاده در سیستم ما DenseNet121 است که توسط Szegedy و همکاران پیشنهاد شده است [38]. این مدل به طور متراکم متصل است و در مقایسه با CNN سنتی به پارامترهای کمتری نیاز دارد. DenseNets لایه های به شدت فشرده ای دارد و تنها از 12 فیلتر به همراه نقشه های ویژگی بسیار کمی استفاده می کند. DenseNet121 می تواند مقادیر گرادین را مستقیماً از تابع ضرر پیدا کند، که زمان مورد نیاز برای آموزش را بهبود می بخشد. این به طور قابل توجهی هزینه محاسبات را کاهش می دهد و این را به یک گزینه برتر تبدیل می کند. در نهایت، ما از مدل Xception به بهترین شکل توصیف شده توسط یک نسخه توسعه یافته از ماژول Inception [23] استفاده کرده ایم. با پردازش مستقل هر کانال خروجی با کمک نگاشت همبستگی های فضایی کار می کند. یک مرحله اضافی در پایان اضافه می شود تا همبستگی بین کانالی را از طریق پیچیدگی های 1×1 ثبت کند. در پایان، ما بخش طبقه بندی کننده CNN های ستون فقرات را با لایه پیش بینی سفارشی خود جایگزین می کنیم، که از یک لایه حذف و یک لایه متراکم متشکل از یک نورون منفرد تشکیل شده است. ما از بهینه ساز Adam با mini-batch استفاده کرده ایم تا کل مدل انباشته را آموزش دهیم. به طور خلاصه، مدل پشته ای ما به عنوان یک مدل یادگیری عمیق منفرد کار می کند که برای پذیرش تصاویر ورودی یکسان از داده های آموزشی چند سری است. بردارهای احتمال پیش بینی متوسط تولید شده از مدل های فرعی CNN سطح اول ترکیب می شوند و از طریق یک فرایادگیرنده (در سطح دوم) برای طبقه بندی نهایی تصاویر ملانوما تغذیه می شوند.

آموزش کامل و اعتبارسنجی مدل انباشته با استفاده از مجموعه داده ضایعات پوستی ملانوما انجام می شود. الگوریتم 1 مکانیسم مجموعه انباشته کامل را خلاصه می کند.

Algorithm 1: Stacked ensemble network for the screening of melanoma skin lesion images

Input: Training data $D = \{x_i, y_i\}_{1 \leq i \leq N}$, Test data D_{test} , CNN sub-models

Output: Classification results from the stacked ensemble network

Split D into D^{train} , D^{valid}

Generate probability scores from predictions made by CNN sub-models (S):

for $s = 1$ to S **do**

 Get predictions $P^{(s)}$ based on D^{train} , D^{valid}

end for

$P^{(st)} = \text{Concatenation}([P^{(1)}, P^{(2)}, \dots, P^{(S)}])$

Create a new dataset, D_{cv} containing the probability scores and target labels:

for $i = 1$ to N **do**

$D_{cv} = \{P_i^{(st)}, y_i\}$

end for

Train a meta-learner, M_{meta} with the newly created dataset, D_{cv}

Validate M_{meta} with D^{valid}

result = classify (M_{meta} , D^{valid})

return result

4 ارزیابی عملکرد

برای نشان دادن اثربخشی مدل گروه انباشته ما در غربالگری بیماری سرطان پوست ملانوما، ما به طور گسترده نتایج عملکرد به دست آمده با استفاده از پیکربندی های مدل مختلف را ارزیابی و مقایسه می کنیم. ما با تنظیمات تجربی، معیارهای ارزیابی، تجزیه و تحلیل نتایج از دیدگاه های کمی و کیفی و بحث ادامه خواهیم داد.

4.1 تنظیمات آزمایشی

جدول 1 توزیع تصاویر را در هر کلاس در کل مجموعه داده بر حسب تعداد نشان می دهد. کل مجموعه داده به مجموعه های آموزشی، اعتبارسنجی و تست با نسبت 60:20:20 برای هر مجموعه تقسیم می شود. سپس، مدل مجموعه انباشته و مدل های فرعی شخصی توسط تصاویر موجود در مجموعه داده با نسبت ذکر شده در جدول بالا، آموزش داده می شوند و اعتبارسنجی می شوند. ما از تقویت تصویر برای مقابله با مشکل مجموعه داده های کوچک و افزایش کارایی آموزش و در عین حال جلوگیری از برآزش بیش از حد (overfitting) مدل ها استفاده می کنیم. همچنین اعتقاد بر این است که تقویت تصویر قابلیت تعمیم پذیری مدل را افزایش می دهد. به طور خاص، زمانی که اندازه مجموعه داده محدود باشد، مفیدتر است و به طور مصنوعی اندازه مجموعه داده آموزشی را افزایش می دهد. خلاصه ای از ویژگی های تقویتی مورد استفاده برای آماده سازی آموزش مجموعه داده و سایر مقادیر فرارامتر در جدول 2 آورده شده است.

Table 1 Distribution of samples in the dataset from both categories of melanoma images

Class	No. of samples			
	Training (60%)	Validation (20%)	Testing (20%)	Total (100%)
Benign	1080	360	360	1800
Malignant	897	300	300	1497

Table 2 Model configuration and augmentation features

Parameters	Values
Max epochs	30
Size of batch	16
Optimizer	Adam
Loss function	Binary cross-entropy
Learning rate (initial)	0.0001
Range of rotation	Random (with factor 0.5)
Shuffling	Yes
Flip	Random
Fill mode	Nearest
Re-scaling	1/255

ما انتخاب می کنیم که از API کاربردی TensorFlow و Keras برای پیاده سازی آنها استفاده کنیم. ما مدل های خود را با استفاده از Google Colab آموزش می دهیم و اعتبارسنجی می کنیم، که دسترسی رایگان به GPU را فراهم می کند. در حال حاضر، پردازنده گرافیکی NVIDIA Tesla P100 را با 16 گیگابایت رم و با Keras API و بکند TensorFlow ارائه می دهد. ما از بهینه ساز Adam با momentum برای آموزش و اعتبارسنجی مدل های خود استفاده می کنیم. ما یادگیری اولیه 0.0001 را برای بهینه ساز Adam تنظیم کردیم. علاوه بر این، ما از تابع از ضرر آنتروپی متقاطع باینری برای آموزش و اعتبارسنجی مدل استفاده کرده ایم. برای یک کار طبقه بندی باینری مانند شناسایی ضایعات بدخیم و خوش خیم پوست، تابع ضرر آنتروپی متقاطع باینری یک انتخاب طبیعی است که همگرایی مدل را سریعتر تسهیل می کند. ما همچنین از ModelCheckpoint و Reduce on Loss Plateau Decay (ReduceLROnPlateau) در Keras استفاده می کنیم. ModelCheckpoint معیارهای عملکرد را نظارت می کند و به صورت دوره ای مدل را بر اساس معیار نظارتی مانند از ضرر اعتبارسنجی (validation loss) ذخیره می کند. ReduceLROnPlateau نرخ یادگیری را کاهش می دهد اگر از ضرر اعتبار سنجی در طول آموزش برای تعداد خاصی از دوره ها (به نام «patience»). بهبود نیابد.

ما عملکرد مدل های مورد مطالعه را با دقت (accuracy)، درستی (precision)، حساسیت (sensitivity)، F1-score و AUC (ناحیه زیر منحنی) اندازه گیری کرده ایم.

4.2 تجزیه و تحلیل نتایج

همانطور که در جدول 3 نشان داده شده است، ما نتایج عملکرد به دست آمده از تمام مدل های CNN و مدل مجموعه انباشته پیشنهادی را ارائه می کنیم.

Model	Acc	Prec	Rec	Spec	F1-score	AUC
ResNet50	88.79	93.33	85.56	92.67	89.28	0.891
VGG-16	90.30	95.68	86.11	95.33	90.64	0.907
Xception	92.42	93.30	92.78	92.00	93.04	0.924
DenseNet121	92.27	91.87	94.17	90.00	93.00	0.921
EfficientNetB0	92.30	94.02	91.67	93.00	92.83	0.923
Stacked model	95.76	95.60	96.67	94.67	96.13	0.957

جدول 3: نتایج عملکرد به دست آمده از CNN و مدل مجموعه ای انباشته پیشنهادی

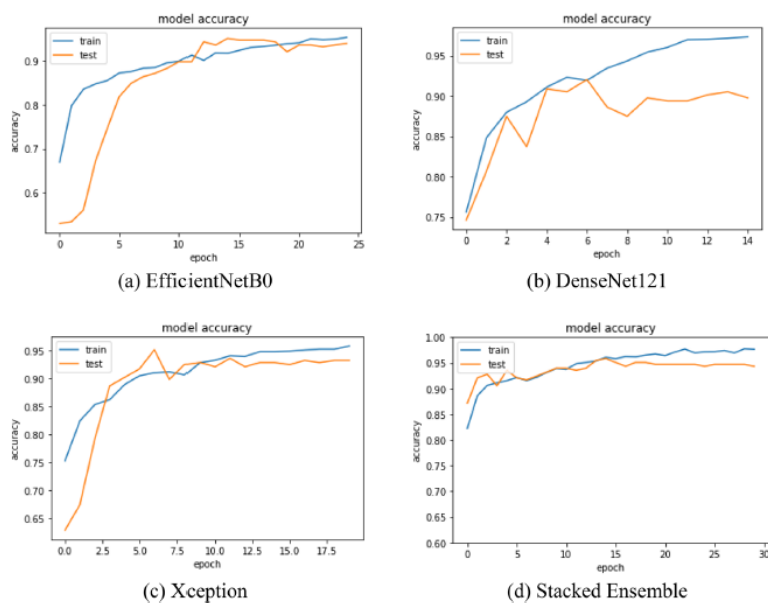
توجه شده است که مدل مجموعه ای همواره نتایج بهتری نسبت به سایر مدل های مجموعه ای در معیارهای کارایی نشان می دهد که شامل دقت (95.76%)، حساسیت (96.67%)، F1-score (96.13%) و AUC (95.7%) در مجموعه ی اعتبارسنجی می باشد. مشخصاً مزیت شبکه مجموعه ای را در طبقه بندی تصاویر ملانوما نسبت به سایر روش های CNN نشان می دهد. نتیجه حساسیت دلالت بر این است که مدل مجموعه ای می تواند به درستی خال های ملانوما را با دقت 96.67 درصد از همه مواردی که هر یک از علائم ملانوما را دارند طبقه بندی کند. تأثیر مدل مجموعه انباشته نیز با نتایج ویژه کلاس نشان داده شده در جدول 4 روشن شده است.

Model	Class	Accuracy	Precision	Recall	F1-score
EfficientNetB0	Benign	91.66	94.0	92.0	93.0
	Malignant	93.00	90.0	93.0	92.0
Xception	Benign	92.77	93.0	93.0	93.0
	Malignant	92.00	91.0	92.0	92.0
DenseNet121	Benign	94.16	92.0	94.0	93.0
	Malignant	90.00	93.0	90.0	91.0
Stacked Model	Benign	96.67	96.0	97.0	96.0
	Malignant	94.67	96.0	95.0	95.0

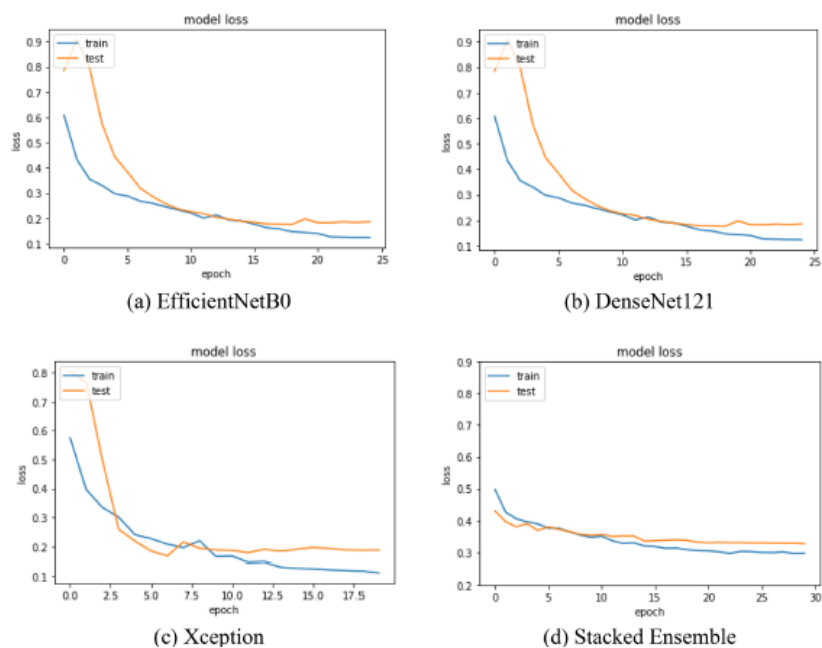
جدول 4: نتایج عملکرد از نظر کلاسی برای سه مدل از بهترین مدل های فرعی CNN و مدل مجموعه انباشته

به نظر می رسد همه مدل ها نتایجی با کارایی بالایی در شناسایی موارد خوش خیم بدون نشانه ای از ملانوما بدخیم نشان می دهند در حالی که عملکرد نسبتاً پایین تری در طبقه بندی خال های بدخیم نشان می دهند.

این نتیجه ظاهراً قابل توجه است زیرا طبقه بندی دقیق تصاویر ملانوما برای هر دو دسته برای یک سیستم تشخیصی مؤثر بسیار مهم است. همچنین منحنی های یادگیری را برای تمامی مدل های مورد مطالعه مشاهده می کنیم. بدیهی است که مدل ها پیشرفت یادگیری متوسطی را در طول فرآیند آموزش نشان می دهند (همانطور که در شکل های 7 و 8 نشان داده شده است) در حالی که کاهش نسبتاً پایداری در تلفات اعتبارسنجی نشان می دهند.

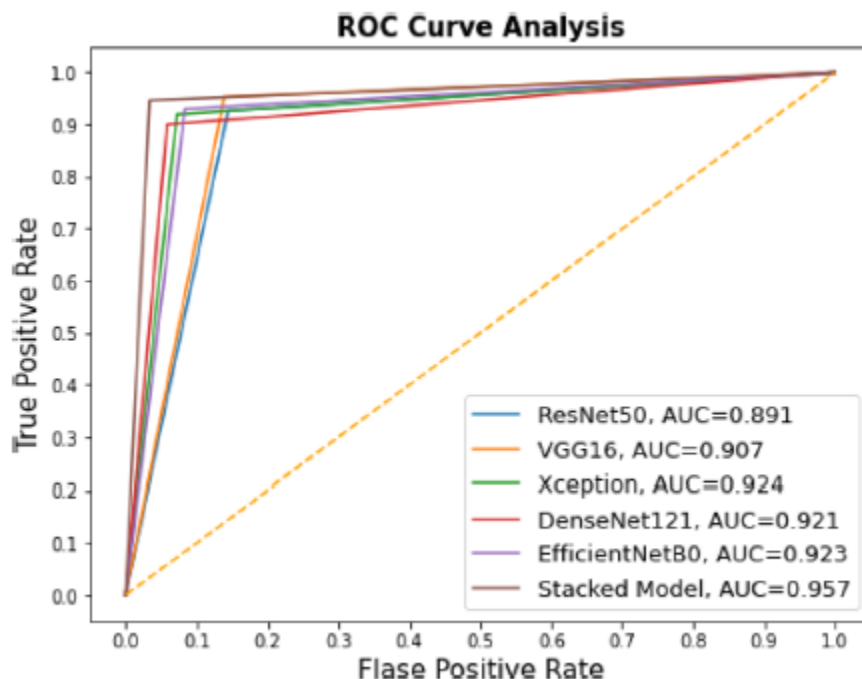


شکل 7: دقت آموزش و اعتبارسنجی با استفاده از سه مدل فرعی و مدل مجموعه انباشته



شکل 8: ضرر آموزش و اعتبارسنجی با استفاده از سه مدل فرعی و مدل مجموعه انباشته

علاوه بر این، به نظر می رسد تلفات آموزشی و اعتبارسنجی در مدل مجموعه انباشته، همانطور که در شکل 8 نشان داده شده است، بسیار بهتر از مدل های فرعی CNN همگرا هستند. به طور کلی، منحنی های یادگیری تمایلی به برآزش بیش از حد ندارند، اگرچه مجموعه داده ما حاوی حداقل نمونه است. این عمدتاً با استفاده از تکنیک های افزایش داده و منظم سازی حذف در مدل مجموعه های انباشته به دست آمده است. برای درک بهتر تفکیک پذیری طبقات مدل های مجموعه مورد مطالعه، از منحنی های مشخصه عملیاتی گیرنده (ROC) همانطور که در شکل 9 نشان داده شده است استفاده می کنیم.



شکل 9: منحنی های ROC برای مدل مجموعه های انباشته و مدل های فرعی مختلف CNN

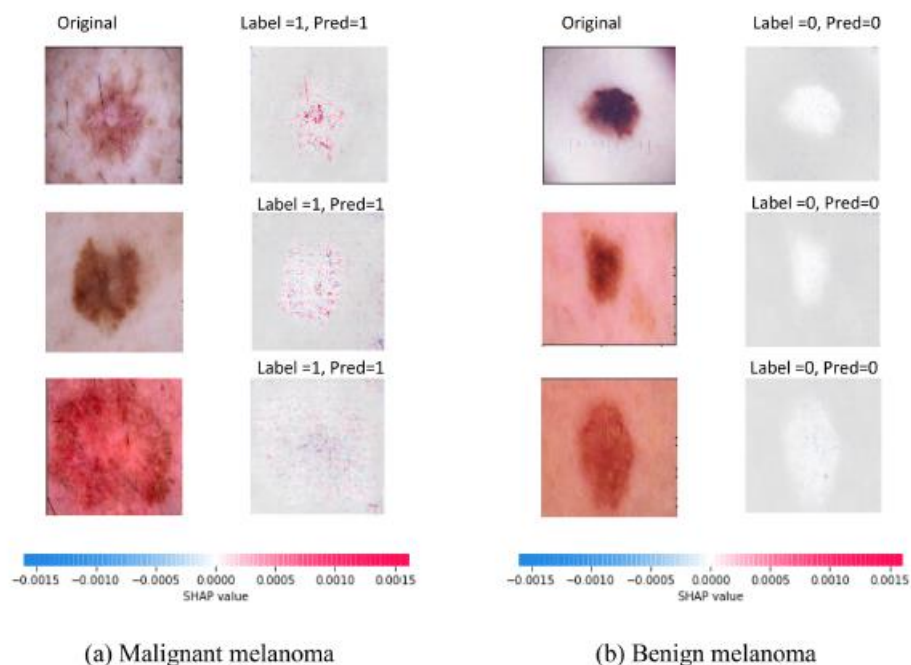
در یک منحنی ROC، true positive rate (TPR) در برابر false positive rate (FPR) با استفاده از مقادیر آستانه متفاوت در نتایج احتمالی مدل های یادگیری عمیق ترسیم می شود. TPR احتمال تشخیص تصاویر ملانوما بدخیم را تعیین می کند. از سوی دیگر، FPR احتمال هشدار کاذب را نشان می دهد که در آن یک تصویر خوش خیم به عنوان دارنده بعضی از علائم ملانوما طبقه بندی می شود. منحنی ROC پایداری مدل مجموعه انباشته را نشان می دهد، جایی که مشاهده می کنیم که مدل به میانگین امتیاز AUC (همانطور که در جدول 3 نشان داده شده است) 0.957 می رسد. مقادیر AUC برای سایر منحنی های ROC به ترتیب 0.924، 0.921 و 0.923 برای Xception، DenseNet121 و Efficient-NetB0 هستند. در میان این مدل های فرعی، Xception توانایی تشخیص برتری را برای عوارض ملانوما نشان می دهد. این امر با این واقعیت اثبات می شود که منحنی Xception کمی بالاتر از سایر مدل های فرعی است.

از جدول 5 مشاهده می شود که مدل مجموعه انباشته، موارد false positive (FP) و false negative (FN) فوق العاده کمی را برخلاف زیرمدل های CNN شخصی ایجاد می کند. کاهش تعداد FP نشان می دهد که تعداد برخوردهای مثبت نادرست شناسایی شده کمتر است و دقت و میزان ویژگی را بهبود می بخشد. در عمل، پایین نگه داشتن تعداد موارد FN نیز مهم است، زیرا شناسایی نادرست بیمار ملانوما به عنوان یک فرد سالم به شدت مانع درمان مناسب بیمار می شود. نتیجه می گیریم که مدل مجموعه انباشته پیشنهادی به عنوان بهترین انجام دهنده در بین همه مدل های مورد مطالعه از نتایج ارزیابی کلی ظاهر می شود. شکل 10 تجسم نمونه ای از مقادیر SHAP را برای مدل فرعی Xception نشان می دهد.

رنگ قرمز به ویژگی‌هایی اشاره دارد که مسئول افزایش مقدار خروجی ملانومای بدخیم هستند، در حالی که رنگ آبی نشان دهنده ویژگی‌هایی است که مقدار خروجی را برای همان ملانوما کاهش می‌دهد. برای یک طبقه خاص، برجستگی منطقه مورد نظر با شدت کل ویژگی‌ها نشان داده شده است. بنابراین، تکنیک تفسیرپذیری، استحکام رویکرد مجموعه عمیق پیشنهادی ما را برای تشخیص خودکار ملانومای بدخیم توضیح می‌دهد. ما همچنین بهترین عملکرد کارایی مدل مجموعه انباشته خود را با چند مطالعه اخیر در مورد تشخیص ضایعات پوستی ملانومای ارائه شده در بخش بررسی ادبیات مقایسه کرده ایم. با توجه به این واقعیت که همه مطالعات از یک مجموعه داده و معیارهای کارایی استفاده نکرده اند، مقایسه واقعی کارایی آنها دشوار خواهد بود. جدول 6 مقایسه کارایی مدل ما را با چندین رویکرد پیشرفته (state-of-the-art) با استفاده از مجموعه داده‌های محبوب ارائه می‌دهد. ما متوجه شدیم که مدل ما عملکرد بهتری نسبت به بسیاری از روش‌های موجود با توجه به معیارهای مختلف کارایی نشان می‌دهد.

5 نتیجه‌گیری و کار آینده

این مقاله یک چارچوب مجموعه‌ای انباشته یادگیری عمیق مبتنی بر CNN را با استفاده از مفهوم یادگیری انتقال برای تشخیص سرطان پوست ملانوما معرفی می‌کند. مطالعه جاری به دنبال دستیابی به این هدف با توسعه یک شبکه مجموعه‌ای است که در آن نتایج پیش‌بینی از چندین مدل فرعی CNN ترکیب شده و به یک فرا یادگیرنده برای پیش‌بینی نهایی خال‌های ملانومای بدخیم داده می‌شود. این مدل بر روی مجموعه داده‌های تصویر ملانومای عمومی آموزش داده شده است. ما از سه مدل فرعی CNN با وزن‌های تنظیم شده به نام‌های DenseNet121، EfficientNetB0 و Xception استفاده کرده ایم. بهترین عملکرد مدل مجموعه انباشته به یک مقدار بالا از دقت (95,76%) و AUC (0.978) در طبقه‌بندی تصاویر ملانوما با درجه بالایی از حساسیت و ویژگی دست می‌یابد. نتایج تجربی نیز ثابت می‌کند که مدل ما با شناسایی مؤثر علائم مختلف ملانومای خوش‌خیم و بدخیم، ویژگی‌های قابل توضیح مطلوبی را نشان می‌دهد. در آینده، ما همچنین قصد داریم با استفاده از تصاویر ملانومای بخش‌بندی شده از طریق شبکه‌های بخش‌بندی پیشرفته، به نتایج پیش‌بینی بهتری دست یابیم.



شکل 10: تجزیه و تحلیل SHAP نمونه‌هایی از تصاویر ملانومای بدخیم و خوش‌خیم با پیش‌بینی‌های واقعی با استفاده از بهترین اجرای مدل فرعی Xception

Model	TP	FP	TN	FN
EfficientNetB0	279	30	330	21
Xception	276	26	334	24
DenseNet121	270	21	339	30
Stacked model	284	12	348	16

جدول 5: ماتریس سردرگمی برای مدل های مختلف ارزیابی شده

Authors	Model	Dataset	Performance results
Kassani and hosseinzadeh [24]	Transfer learning models (ResNet50, AlexNet, Xception, VGGNet16 and VGGNet19)	International Skin Imaging Collaboration (ISIC2018)	ResNet50 performs the best: Accuracy 92.08%, Precision 93.73%, Recall 92.53% and F-score 92.74%
Adegun and viriri [27]	Deep convolutional encoder-decoder architecture	International Symposium on Biomedical Imaging (ISBI)2017 and Hospital Pedro Hispano (PH2)	Accuracy and dice coefficient of 95% and 92% on ISIC 2017 dataset and accuracy and dice coefficient of 95% and 93% on PH2 datasets
Sarkar et al. [28]	Depthwise separable residual convolutional algorithm	International skin imaging collaboration (ISIC), PH2, DermIS, and MED-NODE datasets	Acc. of 99.50% on ISIC, 96.77% on PH2, 94.44% on DermIS and 95.23% on MED-NODE
Filali et al. [30]	Fusion of handcrafted features (shape, skeleton, color, and texture) and features extracted from the most powerful DL architectures	ISIC Challenge dataset, PH2 dataset	94.69%, 96.63%, and 98% for the PH2 database and 62.73%, 55.68%, and 87.8% for the ISIC challenge using the F-measure, Kappa index, and accuracy rate, respectively
Almaraz-Damian et al. [31]	A fusion of handcraft features related to the medical algorithm ABCD rule and deep learning features employing Mutual Information (MI) measurements	ISIC 2018 public dataset (binary classification)	Accuracy, sensitivity, specificity, precision, and F-score of 92.40%, 86.41%, 90.0%, 92.08% and 89.16%, respectively
Ali A.A., Al-Marzouqi [41]	Regular convolutional neural network	Dataset of the ISBI 2016 challenge (binary classification)	Accuracy, sensitivity, specificity of 81.6%, 14.9%, 98.0%,
Hossain et al. [33]	A multi-layered CNN with multiple regularization techniques	ISIC (International Skin Imaging Collaboration) 1497 malign, and 1800 benign images	Accuracy, sensitivity, precision, and F-score of 93.58%, 91.70%, 88.84%, 92.08% and 88.76%, respectively
Proposed Method	Explainable stacked ensemble of CNNs	ISIC (International Skin Imaging Collaboration) 1497 malign, and 1800 benign images	

جدول 6: مقایسه عملکرد رویکرد پیشنهادی با روش های موجود از ادبیات مقاله