

# طبقه‌بندی چندکلاسه سرطان پوست با استفاده از شبکه‌های Vision Transformer و مدل‌های از پیش آموزش‌دیده مبتنی بر شبکه عصبی کانولوشنی

Muhammad Asad Arshed <sup>1,2,\*</sup>, Shahzad Mumtaz <sup>3</sup> , Muhammad Ibrahim <sup>2</sup>, Saeed Ahmed <sup>1</sup> ,  
Muhammad Tahir <sup>4,5</sup> and Muhammad Shafi <sup>6</sup>



**Citation:** Arshed, M.A.; Mumtaz, S.;

Ibrahim, M.; Ahmed, S.; Tahir, M.;

Shafi, M. Multi-Class Skin Cancer

Classification Using Vision

Transformer Networks and

Convolutional Neural NetworkBased Pre-

Trained Models. *Information* **2023**, *14*,

415. [https://](https://doi.org/10.3390/info14070415)

[doi.org/10.3390/info14070415](https://doi.org/10.3390/info14070415)

Received: 6 July 2023

Revised: 13 July 2023

Accepted: 14 July 2023

Published: 18 July 2023



**Copyright:** © 2023 by the authors.

Licensee MDPI, Basel, Switzerland. This

article is an open access article

distributed under the terms and

conditions of the Creative Commons

Attribution (CC BY) license ([https://creativecommons.org/licenses/by/](https://creativecommons.org/licenses/by/4.0/)

[https://creativecommons.org/licenses/by/](https://creativecommons.org/licenses/by/4.0/)

4.0/).

<sup>1</sup> School of Systems and Technology, University of Management and Technology, Lahore 54770, Pakistan; saeed.ahmed@umt.edu.pk

<sup>2</sup> Department of Computer Science, The Islamia University of Bahawalpur, Bahawalpur 63100, Pakistan; muhammad.ibrahim@iub.edu.pk

<sup>3</sup> Department of Data Science, The Islamia University of Bahawalpur, Bahawalpur 63100, Pakistan; shahzadmumtaz22@gmail.com

<sup>4</sup> Department of Electrical and Computer Engineering, University of Manitoba, Winnipeg, MB R3T 5V6, Canada; m.tahir@umanitoba.ca

<sup>5</sup> Department of Computer Science, Abdul Wali Khan University, Mardan 23200, Pakistan

<sup>6</sup> Faculty of Computing and Information Technology, Sohar University, Sohar 311, Oman; mshafi@su.edu.om

\* Correspondence: muhammadasadarshed@gmail.com

**چکیده:** سرطان پوست، به ویژه ملانوما، به عنوان یکی از کشنده‌ترین اشکال سرطان شناخته شده است. تشخیص و شناسایی دقیق ضایعات پوستی

می‌تواند به دلیل شباهت‌های چشمگیر بین انواع مختلف ضایعات پوستی، مانند ملانوما و خال‌های پوستی، به ویژه هنگام بررسی تصاویر رنگی پوست،

چالش‌برانگیز باشد. با این حال، تشخیص زودهنگام نقش مهمی در نجات جان‌ها و کاهش بار بر منابع پزشکی ایفا می‌کند. در نتیجه، توسعه یک

سیستم خودکار قوی برای طبقه‌بندی سرطان پوست ضروری می‌شود. شبکه‌های عصبی کانولوشنی (CNNs) طی دهه گذشته به طور گسترده برای

خودکارسازی تشخیص سرطان به کار گرفته شده‌اند. با این وجود، ظهور Vision Transformer (ViT) اخیراً محبوبیت قابل توجهی در این

زمینه به دست آورده و به عنوان جایگزینی رقابتی برای CNN ها مطرح شده است. با توجه به این موضوع، مطالعه حاضر روشی جایگزین مبتنی بر

ViT آماده‌کار را برای شناسایی بیماری‌های مختلف سرطان پوست پیشنهاد کرد. برای ارزیابی عملکرد آن، روش پیشنهادی با ۱۱ روش یادگیری

انتقالی مبتنی بر CNN که شناخته شده‌اند از سایر تکنیک‌های یادگیری عمیق در حال استفاده، بهتر عمل می‌کنند، مقایسه شد. علاوه بر این، این

مطالعه به مسئله عدم تعادل کلاس درون مجموعه داده، که چالشی رایج در طبقه‌بندی سرطان پوست است، می‌پردازد. در پرداختن به این نگرانی،

مطالعه پیشنهادی از Vision Transformer و مدل‌های یادگیری انتقالی مبتنی بر CNN برای طبقه‌بندی هفت نوع مختلف سرطان پوست

بهره می‌برد. از طریق تحقیق خود، دریافتیم که استفاده از Vision Transformer های از پیش آموزش‌دیده به دقت چشمگیر ۹۲.۱۴٪ دست

یافت و از مدل‌های یادگیری انتقالی مبتنی بر CNN در چندین معیار ارزیابی برای تشخیص سرطان پوست پیشی گرفت.

**کلمات کلیدی:** تشخیص سرطان پوست؛ چندکلاسه؛ Vision Transformer؛ مدل‌های از پیش آموزش‌دیده؛ تنظیم دقیق؛ یادگیری انتقالی؛

افزایش داده

## ۱. اطلاعات

این عضو فیزیکی واقعی بدن ما است که بخش‌های مختلف بدن مانند ماهیچه‌ها و استخوان‌ها را می‌پوشاند. اگر پوست تحت تأثیر هر چیزی قرار گیرد، تأثیر شدیدی بر کل سیستم دارد. ویروس‌ها، آلرژی‌ها، مصرف الکل، عفونت‌ها، فعالیت بدنی، قرار گرفتن در معرض نور فرابنفش (UV)، تغییرات محیطی و تورم‌های غیرمعمول بدن انسان چندین نمونه از دلایل زمین‌های سرطان پوست هستند. ناحیه بیمار روی پوست، ناحیه ضایعه نامیده می‌شود و بسیاری از ضایعات بر اساس منشأ خود تقسیم‌بندی می‌شوند.

سرطان پوست به کارسینوم سلول بازال و کارسینوم سلول سنگ‌فرشی تقسیم می‌شود. این‌ها به عنوان سرطان‌های کراتینوسیت اشاره می‌شوند و علت اصلی آن‌ها قرار گرفتن در معرض نور خورشید است. آن‌ها اغلب بر قسمت‌هایی از بدن که مستقیماً در معرض آفتاب قرار می‌گیرند، مانند صورت، بازوها و دست‌ها تأثیر می‌گذارند. بخش دیگری از بدن به ندرت تحت تأثیر BCC (کارسینوم سلول بازال) قرار می‌گیرد، اما اندام‌ها یا غدد لنفاوی مجاور به راحتی تحت تأثیر SCC (کارسینوم سلول سنگ‌فرشی) قرار می‌گیرند. شکل دیگر کشنده و نادر سرطان پوست، ملانوما [۳] است که در ملانوسیت‌ها ایجاد می‌شود. اگر در مرحله اولیه تشخیص داده شود، با جراحی قابل درمان است؛ در غیر این صورت، زنده ماندن دشوار یا حتی غیرممکن خواهد بود. ملانوما نوعی سرطان است که عمدتاً افراد سفیدپوست را تحت تأثیر قرار می‌دهد و اغلب تنه در مردان و اندام تحتانی در زنان را درگیر می‌کند؛ با این حال، می‌تواند در نواحی دیگر بدن نیز ظاهر شود. در ایالات متحده، گزارش شده است که ۷۵ درصد از مرگ‌ومیرهای ناشی از سرطان پوست توسط ملانوما ایجاد شده است، که ۵ درصد از کل سرطان‌های پوست را تشکیل می‌دهد.

سازمان جهانی بهداشت پیش‌بینی می‌کند که سرطان پوست در هر سومین بیمار سرطانی تشخیص داده خواهد شد، بنابراین اولین و مهم‌ترین هدف تحقیقات پزشکی امروز، درمان سرطان پوست است. این موضوع به ویژه در این زمینه حائز اهمیت است زیرا بر اساس آمار، گزارش شده است که بیش از دوازده میلیون نفر از سرطان رنج می‌برند. بنیاد سرطان پوست ایالات متحده گزارش داد که هر ساله آمریکایی‌های بیشتری در مقایسه با تمام سرطان‌های دیگر تحت تأثیر سرطان پوست قرار می‌گیرند؛ انتظار می‌رود حدود ۵.۴ میلیون مورد سرطان پوست سالانه در ایالات متحده تشخیص داده شود. بنابراین، نیاز به غربالگری بالینی به سرعت در حال افزایش است. علاوه بر این، آسیب‌شناسان بافتی تشخیص سرطان پوست از اپیلومینوسکوپی ضایعات پوستی را بسیار دشوار می‌یابند. پزشکان به طور کلی از روش بیوپسی برای تشخیص بیماری‌های سرطان استفاده می‌کنند. این روش یک نمونه پوستی متفاوت را در آزمایشگاه ارزیابی می‌کند، به همین دلیل این یک روش دردناک و زمان‌بر است.

تصاویر ماکروسکوپی که عموماً با دوربین دیجیتال یا ویدیو به دست می‌آیند، معمولاً به صورت محاسباتی تجزیه و تحلیل می‌شوند. بررسی ضایعات پوستی در صورتی چالش‌برانگیز است که تصاویر بالینی وجود اشیایی از جمله خطوط پوستی، سایه‌ها و مو را در تصاویر نشان دهند و وضوح تصویرضعیفی داشته باشند. با کمک هوش مصنوعی، میزان بار مرگ‌ومیر و بیماری مرتبط با این بیماری می‌تواند به دلیل تشخیص زودهنگام بیماری‌ها به سرعت کاهش یابد. این از تکنیک‌های مختلفی مانند یادگیری ماشین (ML) استفاده می‌کند که شامل الگوریتم‌ها و مدل‌های مختلفی است که از داده‌های آموزشی یاد می‌گیرند. این سعی می‌کند پیش‌بینی کند که خروجی بر روی داده‌های آزمایشی/نمونه‌های جدید برای انجام وظایف مورد نظر برای مغز انسان دشوار است. چندین نوع سیستم مبتنی بر کامپیوتر برای تشخیص سرطان پوست معرفی شده است. الگوریتم‌های بنیایی کامپیوتر سنتی عمدتاً به عنوان یک طبقه‌بندی برای تشخیص سرطان و استخراج ویژگی‌های مختلف از تصاویر استفاده می‌شوند. شبکه‌های عصبی عمیق (DNN)، شبکه‌های عصبی کانولوشنی (CNN)، حافظه کوتاه‌مدت بلندمدت (LSTM) و شبکه‌های عصبی بازگشتی (RNN) پرکاربردترین مدل‌های یادگیری عمیق در صنعت مراقبت‌های بهداشتی هستند.

مدل‌های ML سنتی نیاز به استخراج ویژگی‌های کارآمد از تصاویر پوست دارند و تصاویر پوست بر اساس این ویژگی‌های ارزشمند طبقه‌بندی می‌شوند [۴]. به دلیل محدودیت‌های طول ویژگی، مدل‌های ML برای سرطان‌های پوست به جای یک مدل تعمیم‌یافته برای انواع مختلف بیماری‌ها استفاده می‌شوند [۵]. اخیراً از یادگیری عمیق (DL) برای طبقه‌بندی سرطان پوست بدون دانش عمیق از DL و استخراج ویژگی استفاده شده است. در مقایسه با مدل‌های ML، مدل‌های DL برای مجموعه داده‌های بزرگ مقیاس عملکرد خوبی دارند [۶]. در مقایسه با یک متخصص پوست برای شناسایی سرطان، سیستم‌های مبتنی بر ML در حال طراحی هستند [۷،۸]، اما هنوز به تکنیک‌های بهبود یافته‌ای برای سیستم‌های مراقبت‌های بهداشتی مؤثر نیاز است. در طول طراحی مدل DL، تعادل مجموعه داده و تصاویر زیادی باید برای آموزش و ارزیابی عملی مدل در نظر گرفته شود. علاوه بر این، DL هنگامی که مجموعه داده شامل تصاویر با وضوح بالا باشد، به زمان و هزینه آموزش اضافی نیاز دارد [۹]. کومار و همکاران [۱۰] روشی را برای ارزیابی عفونت‌های پوستی با ترکیبی از بنیایی کامپیوتر و رویکرد ML پیشنهاد کردند، که از آن بنیایی کامپیوتر برای استخراج ویژگی و ML به ترتیب برای ارزیابی بیماری استفاده شد. دقت مدل پیشنهادی ۹۵ درصد بود CNN، های مبتنی بر ناحیه در مطالعه منتشر شده توسط نویسندگان [۱۱] برای تشخیص عفونت با سه فناوری (پیشنهاد ناحیه، تبدیل بردار و طبقه‌بندی) استفاده شده‌اند. برای طبقه‌بندی سرطان پوست و سایر بیماری‌های مرتبط، محققان از GoogleNet V3 CNN استفاده کرده‌اند. آن‌ها مجموعه داده‌های مشکل از تصاویر درماتوسکوپی و تصاویر بالینی سرطان پوست را در نظر گرفته‌اند و توانسته‌اند به دقت  $0.9 \pm 0.0721$  دست یابند. طبقه‌بندی‌های ضایعه پوستی و تکنیک‌های چک‌لیست ۷ نقطه‌ای نیز برای تشخیص بیماری پوستی استفاده شده‌اند. [12]

روش استخراج ویژگی به طور مؤثر زمان یادگیری را کاهش می‌دهد و عملکرد مدل‌های ML و DL را بهبود می‌بخشد [۱۳]. برای طبقه‌بندی دودویی تصاویر بالینی، یک رویکرد مبتنی بر CNN در سال ۲۰۱۸ به کار گرفته شد [۱۴]. رویکرد یادگیری انتقال (TL) نیز برای پیش‌بینی سرطان پوست و دستیابی به بالاترین دقت استفاده شده است [۱۵]. یک مدل مبتنی بر CNN جدید به نام SkNet شامل ۱۹ لایه برای طبقه‌بندی

چهار نوع سرطان پوست پیشنهاد شد [۱۶]. برای تشخیص سرطان پوست، مدل مبتنی بر CNN دیگر VGG-16 نیز به دلیل عملکرد تشخیص مؤثر آن استفاده شده است. [17]

به منظور تشخیص و طبقه‌بندی سرطان پوست، سرطان پوست اولیه با استخراج ویژگی‌های تصویر، مانند شکل، بافت، هندسه و سایر روش‌های ایجاد شده دستی طبقه‌بندی می‌شود. CNN ها به روش محبوبي برای شناسایی تصاویر پزشکی تبدیل شده‌اند. CNN ها به دلیل دقت قابل توجه خود با موفقیت در طبقه‌بندی سرطان پوست استفاده شده‌اند و ارزش خود را در این زمینه نشان داده‌اند. اگرچه هنگام استفاده از یک CNN با معماری عمیق، امکان استخراج ویژگی‌ها برای بسیاری از اشیاء کوچک در یک تصویر وجود دارد، اما ممکن است شناسایی دقیق مناطق واقعاً حیاتی دشوار باشد. برای کاهش این مشکل، ما در این مطالعه از مدل Vision Transformer (ViT) استفاده کرده‌ایم. تصویر ورودی در طول فرآیند آموزش عمومی برای این مدل به بلوک‌هایی تقسیم می‌شود و هر بلوک به عنوان یک موجودیت جداگانه در نظر گرفته می‌شود. مدل ViT از ماژول‌های Self-Attention برای درک چگونگی ارتباط این پیک‌های تعبیه‌شده با یکدیگر استفاده می‌کند. مدل ViT اخیراً عملکردی برجسته در وظایف طبقه‌بندی معمولی نشان داده است. مکانیسم Self-Attention ترنسفورمر اهمیت ویژگی‌های مهم را افزایش می‌دهد در حالی که تأثیر ویژگی‌هایی که باعث نویز می‌شوند را کاهش می‌دهد. با انگیزه این دیدگاه، مطالعه حاضر یک شبکه طبقه‌بندی سرطان پوست مبتنی بر ViT را پیشنهاد کرد. یافته‌ها نشان داد که شبکه پیشنهادی نتایج رضایت‌بخشی در طبقه‌بندی سرطان پوست ارائه می‌دهد. علاوه بر این، این مطالعه استفاده از رویکردهای مبتنی بر CNN و تنظیم دقیق آن‌ها را برای نشان دادن استحکام مدل ViT بررسی کرد. این تحقیق به صورت زیر در این زمینه مشارکت می‌کند:

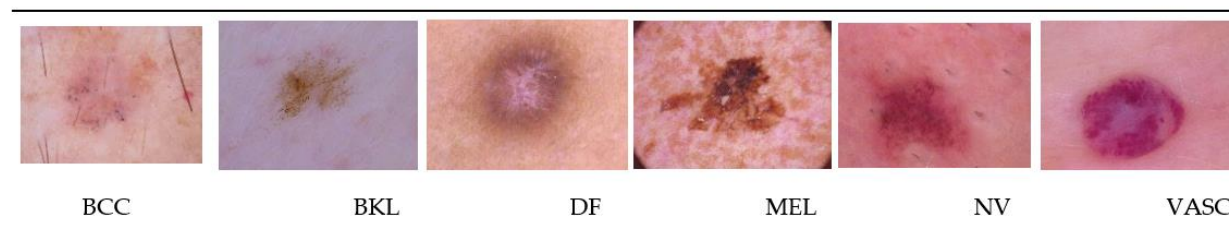
- برای حل مسئله عدم تعادل کلاس، یک تکنیک مؤثر افزایش داده برای افزایش مصنوعی نمونه‌های مجموعه داده پیاده‌سازی شد.
- مدل ViT تنظیم‌شده پیشنهادی از مدل‌های پیشرفته موجود برای طبقه‌بندی چندکلاسه سرطان پوست پیشی گرفت.
- در این مطالعه، ما همچنین مدل‌های از پیش آموزش‌دیده مبتنی بر CNN از جمله ResNet50، ResNet101، ResNet152، ResNet50V2، ResNet101V2، ResNet152V2، DenseNet121، DenseNet169، DenseNet201، VGG16 و VGG19 را به ترتیب تنظیم دقیق کردیم.
- آزمایش‌های گسترده با استفاده از تکنیک افزایش داده برای پیشنهاد یک مدل مؤثر انجام شد.
- سیستم طبقه‌بندی ضایعات پوستی چندکلاسه تکامل یافته است و اطلاعات تشخیصی دقیقی را در اختیار متخصصان و بیماران قرار می‌دهد.

## ۲. مواد و روش‌ها

در این بخش، ما روش‌هایی که برای دستیابی به یک طبقه‌بندی مؤثر سرطان پوست اتخاذ و پیشنهاد شدند را معرفی کرده‌ایم.

### ۲.۱. مجموعه داده

برای آزمایش خود، ما از یک مجموعه داده که از [18] Kaggle به دست آمده بود استفاده کردیم، همانطور که در شکل ۱ نشان داده شده است. با این حال، توجه به این نکته مهم است که مجموعه داده مورد استفاده ما نامتعادل بود، که می‌تواند منجر به ارائه چالش‌هایی مانند Overfitting یا Underfitting شود. برای اطمینان از عملکرد بهینه مدل، بنابراین لازم است مسئله عدم تعادل داده، به ویژه در مرحله آموزش، مورد توجه قرار گیرد.

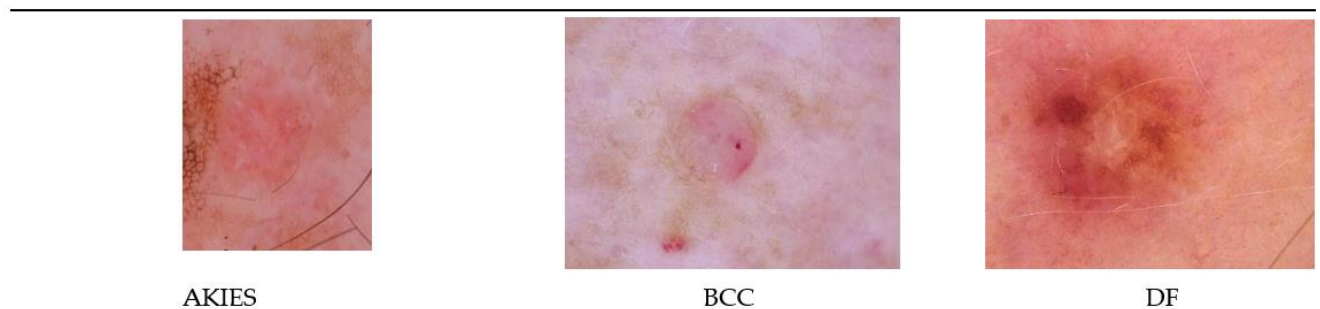


شکل ۱. تصاویر نمونه سرطان پوست به دست آمده از مجموعه داده HAM10000

### ۲.۲. افزایش داده‌ها

افزایش داده‌ها، تکنیکی است که با اعمال تغییرات بر روی داده‌های اصلی، اندازه مجموعه‌داده را افزایش می‌دهد. این روش به ویژه در مدل‌های مبتنی بر یادگیری عمیق ارزشمند است، چرا که داشتن مجموعه‌داده آموزشی گسترده برای دستیابی به عملکرد مؤثر ضروری است. اگرچه افزایش داده می‌تواند در حوزه‌های مختلف به کار گرفته شود، اما به طور گسترده برای مسائل بینایی کامپیوتر استفاده می‌شود. در این مطالعه، ما بر چرخش به عنوان یک تکنیک افزایش داده تمرکز کردیم و به طور خاص، حداکثر چرخش‌های چپ و راست تا ۸ درجه را اعمال نمودیم. با گنجاندن افزایش داده، هدف ما تنوع بخشیدن به مجموعه‌داده و جلوگیری از ایجاد سوگیری در مدل ناشی از عدم تعادل کلاس‌ها بود.

فرآیند افزایش داده شامل دو دسته اصلی بود: افزایش موقعیت و افزایش رنگ. برای افزایش موقعیت، از زیرتکنیک‌هایی مانند مقیاس گذاری، برش، تبدیل آفین، حاشیه‌گذاری، چرخش، انتقال و برگرداندن (افقی/عمودی) استفاده کردیم. در مورد افزایش رنگ، تکنیک‌هایی مانند تنظیمات فام (hue)، روشنایی، اشباع و کنتراست را در نظر گرفتیم. نمونه‌های افزایش‌یافته در شکل ۲ نمایش داده شده‌اند.



شکل ۲. تصاویر نمونه افزایش‌یافته از کلاس‌های اقلیت.

برای ایجاد یک مجموعه‌داده آموزشی متعادل، تصاویری را به‌طور تصادفی از مجموعه‌داده اصلی انتخاب و افزایش دادیم که در نهایت منجر به دقیقاً ۹۸۰ نمونه برای هر کلاس تنها برای آموزش شد. این تصاویر افزایش‌یافته سپس با داده‌های اصلی ترکیب شدند و در مجموع ۹۸۰ تصویر برای هر کلاس جهت آموزش به دست آمد (مجموع نمونه‌های آموزشی:  $980 \times 7$ ). برای اعتبارسنجی، از ۱۴۰ تصویر جداگانه استفاده شد و برای آزمایش، مجموعه جداگانه‌ای متشکل از ۱۴۰ تصویر برای این مطالعه در نظر گرفته شد.

به‌طور کلی، این تکنیک‌ها، یعنی افزایش داده و نمونه‌برداری تصادفی (کاهش نمونه‌های اکثریت)، برای حل مسئله عدم تعادل داده و اطمینان از نتایج مؤثرتر در مدل ما مورد استفاده قرار گرفتند.

مدل‌های از پیش آموزش‌دیده (مانند ResNet50، VGG16 و VGG19) بر روی این مجموعه‌داده افزایش‌یافته آموزش و ارزیابی شدند که به ما امکان داد تا به‌طور مؤثری عدم تعادل کلاس را برطرف کرده و عملکرد کلی وظیفه طبقه‌بندی سرطان پوست را بهبود بخشیم.

### ۲.۳. مدل مؤثر از پیش آموزش‌دیده مبتنی بر CNN (تنظیم دقیق ResNet50)

برای اطمینان از تعمیم‌پذیری مؤثر، مراحل پیش‌پردازش حداقلی را برای مدل پیشنهادی به کار گرفتیم. تصاویر مجموعه‌داده به اندازه ثابت  $224 \times 224$  پیکسل استاندارد شدند.

معماری‌های مختلف شبکه عصبی اهداف مختلفی را دنبال می‌کنند. برای مثال، شبکه‌های عصبی بازگشتی (RNN) در پردازش زبان طبیعی و تشخیص گفتار عالی عمل می‌کنند، در حالی که شبکه‌های عصبی کانولوشنی (CNN) در تحلیل ورودی‌های بصری و پردازش تصاویر بسیار مؤثر هستند.

با این حال، افزایش عمق این شبکه‌ها چالش‌هایی را به همراه می‌آورد. آموزش شبکه‌های عمیق به دلیل مشکل گرادیان محو شونده می‌تواند دشوار باشد، جایی که گرادیان‌ها در حین انتشار به عقب در لایه‌ها کاهش می‌یابند. برای حل این مشکل، یک شبکه عصبی باقی‌مانده (ResNet) معرفی شد که از اتصالات پرشی استفاده می‌کند. با پرش از چندین لایه، مقادیر به پایین‌ترین نقطه سقوط نمی‌کنند و در نتیجه مشکل گرادیان محو شونده کاهش می‌یابد. اتصالات پرشی شامل افزودن ورودی به خروجی یک لایه خاص است [۱۹].

**ResNet50**، یک شبکه کانولوشنی عمیق متشکل از ۵۰ لایه، به‌طور خاص برای وظایف طبقه‌بندی تصویر طراحی شده است. این مدل توسط کیمینگ هه و همکاران معرفی شد [۲۰]. **ResNet50** از دو نوع بلوک تشکیل شده است: بلوک‌های هویتی و بلوک‌های کانولوشنی. بلوک‌های هویتی زمانی استفاده می‌شوند که ابعاد ورودی و خروجی یکسان باقی بمانند، در حالی که بلوک‌های کانولوشنی

زمانی استفاده می‌شوند که ابعاد ورودی و خروجی متفاوت باشند. در مواردی که این ابعاد مطابقت ندارند، یک بلوک کانولوشنی می‌تواند به مسیر میان‌بر اضافه شود تا برابری بین ابعاد ورودی و خروجی تضمین شود.

با گنجاندن **ResNet50** در تحقیق خود، هدف ما بهره‌گیری از معماری و اتصالات پرشی آن برای بهبود عملکرد مدل پیشنهادی‌مان برای طبقه‌بندی سرطان پوست بود. در این مطالعه، حذف تصادفی (**Dropout**) عمدتاً روی لایه‌های پنهان و ورودی شبکه عصبی اعمال شد، نه لایه خروجی. این به این دلیل است که در طول آزمایش، تمام نورون‌ها و اتصالات در شبکه باید برای پیش‌بینی‌ها و استنتاج دقیق در دسترس باشند. بنابراین، حذف تصادفی روی لایه خروجی اعمال نشد تا در دسترس بودن ساختار کامل شبکه در طول آزمایش تضمین شود [۲۱]. علاوه بر این، با استفاده از ادغام میانگین (**Average Pooling**) در تحقیق خود، هدف ما کوچک‌سازی مؤثر نقشه‌های ویژگی در حالی بود که اطلاعات معنادار را برای لایه‌های بعدی در شبکه عصبی حفظ کنیم. این به ما امکان می‌دهد تا ویژگی‌های مرتبط را استخراج کرده و طبقه‌بندی سرطان پوست را با دقت و کارایی بهبود یافته تسهیل کنیم.

بیش‌برازش (**Overfitting**) یک چالش رایج در یادگیری ماشین است، جایی که یک مدل روی داده‌های آموزشی عملکرد خوبی دارد اما پیش‌بینی‌های ناپایداری روی داده‌های آزمایشی نشان می‌دهد. برای حل این مسئله، از تکنیک‌های تنظیم (**Regularization**) استفاده شد. یکی از این تکنیک‌ها، نرمال‌سازی دسته‌ای (**Batch Normalization**) است که نقش حیاتی در تسريع فرآیند آموزش شبکه‌های عصبی کانولوشنی و افزایش پایداری آن‌ها ایفا می‌کند [۲۲].

معمولاً، یک دسته از داده‌های ورودی جمع‌آوری و برای آموزش یک شبکه عصبی استفاده می‌شود [۲۲]. با گنجاندن نرمال‌سازی دسته‌ای در فرآیند آموزش، می‌توانیم فعال‌سازی‌ها را در هر دسته نرمال‌سازی کنیم و در نتیجه تغییر کوواریانس داخلی (**Internal Covariate Shift**) را کاهش دهیم. این نرمال‌سازی به تثبیت فرآیند یادگیری کمک می‌کند و به مدل اجازه می‌دهد تا بهتر تعمیم یابد و پیش‌بینی‌های سازگارتری در مجموعه داده‌های مختلف انجام دهد.

تابع فعال‌سازی یک مؤلفه حیاتی در شبکه‌های عصبی است که سطح فعال‌شدن هر نورون را تعیین می‌کند. این شامل محاسبه مجموع وزندار ورودی هر نورون، افزودن عبارت بایاس و عبور دادن مقدار حاصل از طریق تابع فعال‌سازی است. با معرفی غیرخطی بودن، تابع فعال‌سازی به شبکه‌های عصبی اجازه می‌دهد تا مسائل پیچیده را حل کنند، زیرا بدون آن، شبکه تنها رگرسیون خطی انجام می‌داد [۲۲].

تابع فعال‌سازی در محدوده‌های خروجی خاصی عمل می‌کند، که معمولاً مقادیر را بین ۱ و ۱- یا ۰ و ۱ محدود می‌کند. دو دسته اصلی از توابع فعال‌سازی وجود دارد: تابع فعال‌سازی خطی/هویتی و توابع فعال‌سازی غیرخطی. توابع فعال‌سازی غیرخطی خود شامل انواع مختلفی مانند سیگموئید/لجستیک، سافت‌مکس، تانژانت هذلولوی (**Tanh**)، **ReLU**، **ReLU** نشت‌دار (**Leaky ReLU**) و واحد خطی اصلاح‌شده (**ReLU**) می‌شوند [۲۳، ۲۴].

تابع فعال‌سازی غیرخطی بودن را به شبکه‌های عصبی اضافه می‌کند و به آن‌ها اجازه می‌دهد تا به طور مؤثری با وظایف پیچیده برخورد کنند. این تابع مشخص می‌کند کدام نورون‌ها فعال هستند و توابع فعال‌سازی را به انواع خطی و غیرخطی دسته‌بندی می‌کند که هر کدام ویژگی‌ها و موارد استفاده خاص خود را دارند.

**Adam** [۲۵] یک گسترش از نزول گرادیان تصادفی (**SGD**) است (که هدف آن یافتن حداقل یک تابع ضرر داده‌شده با تنظیم تکراری پارامترهای مدل است) و برای به‌روزرسانی مکرر وزن‌های شبکه بر اساس داده‌های آموزشی استفاده می‌شود. این بهینه‌ساز دو روش را ترکیب می‌کند - الگوریتم گرادیان تطبیقی و انتشار جذر میانگین مربعات - تا به گرادیان‌های پراکنده در مسائل پُر‌نویز کمک کند. ما **Adam** را در این مطالعه در نظر گرفتیم، زیرا بهتر، سریع‌تر، نیاز حافظه کمتری در مقایسه با سایر الگوریتم‌های بهینه‌سازی دارد و تنها پارامترهای تنظیم کمی دارد.

### تنظیم ابرپارامترها (Hyperparameter Tuning)

ابرپارامترها پارامترهای ضروری هستند که باید قبل از آموزش یک مدل تعریف شوند. آن‌ها نقش حیاتی در تعیین رفتار و عملکرد مدل دارند. نمونه‌هایی از ابرپارامترها شامل نرخ حذف تصادفی، تابع فعال‌سازی، نورون‌های لایه پنهان، تعداد دوره‌های آموزش و اندازه دسته هستند [۲۲].

در مطالعه ما، مقادیر این ابرپارامترها به صورت تجربی تعیین شدند و اطمینان حاصل شد که برای وظیفه خاص در دست بهینه هستند. مقادیر خاص انتخاب شده برای هر ابرپارامتر را می‌توان در جدول ۱ یافت که شفافیت و قابلیت بازتولید را در تنظیمات آزمایشی ما فراهم می‌کند. با انتخاب دقیق و تنظیم دقیق این ابرپارامترها، هدف ما حداکثر کردن عملکرد و اثربخشی مدل ما در طبقه‌بندی سرطان پوست بود.

جدول ۱. تنظیم ابرپارامترهای شبکه.

مقدار	پارامتر
1024, 512, 256, 128, 64, 7	Hidden neurons
20	Epochs
0.3	Dropout
ReLU, SoftMax	Activation function
Categorical	Loss-Function
Adam	Optimizer
0.001	Learning rate
32	Batch size
بله	Early stopping
3	Patience

### ۲.۴. یادگیری انتقالی و تغییرات معماری شبکه در مدل‌های از پیش آموزش‌دیده مبتنی بر CNN

یادگیری انتقالی یک مفهوم مؤثر است که از دانش کسب‌شده از وظایف برای حل وظایف مرتبط از طریق تنظیم دقیق استفاده می‌کند. چندین رویکرد برای انجام تنظیم دقیق وجود دارد. یک رویکرد، تنظیم دقیق برخی یا تمام پارامترهای آخرین لایه مدل از پیش آموزش‌دیده است [۲۶].

در این کار، یک استخراج‌کننده ویژگی یکپارچه برای انجام استخراج مؤثر ویژگی با استفاده از مدل از پیش آموزش‌دیده به کار گرفته شد. این رویکرد از مفهوم یادگیری انتقالی بهره می‌برد و به مدلی که برای یک مسئله خاص آموزش دیده است اجازه می‌دهد تا با تنظیم دقیق مدل، برای مسئله‌ای متفاوت مورد استفاده قرار گیرد. به عنوان مثال، آخرین لایه کانولوشنی ResNet50 با لایه‌های چگال (۱۰۲۴، ۵۱۲، ۲۵۶، ۱۲۸، ۶۴ و ۷) نگاشت شد. نرمال‌سازی دسته‌ای و نرخ حذف تصادفی ۰.۳ پس از هر لایه چگال اعمال شد. لایه‌های پنهان از تابع فعال‌سازی ReLU استفاده کردند، در حالی که آخرین لایه شامل ۷ نورون با تابع فعال‌سازی سافت‌مکس بود.

لایه‌های ResNet50 در این مطالعه ثابت (فریز) شدند و وزن‌های از مجموعه داده ImageNet مورد استفاده قرار گرفتند. پیکربندی یکسان برای تمام معماری‌های مدل از پیش آموزش‌دیده برای حفظ یکنواختی اعمال شد.

به طور کلی، هر دو تکنیک تنظیم دقیق و استخراج ویژگی در این مطالعه پیاده‌سازی شدند و یک استخراج‌کننده ویژگی یکپارچه برای استخراج مؤثر ویژگی مورد استفاده قرار گرفت. پیکربندی‌ها و رویکردهای خاص توصیف‌شده در بالا به طور یکنواخت در تمام معماری‌های مدل از پیش آموزش‌دیده مورد استفاده در تحقیق ما اعمال شدند.

#### ۲.۵. مدل از پیش آموزش‌دیده (ViT) Vision Transformer

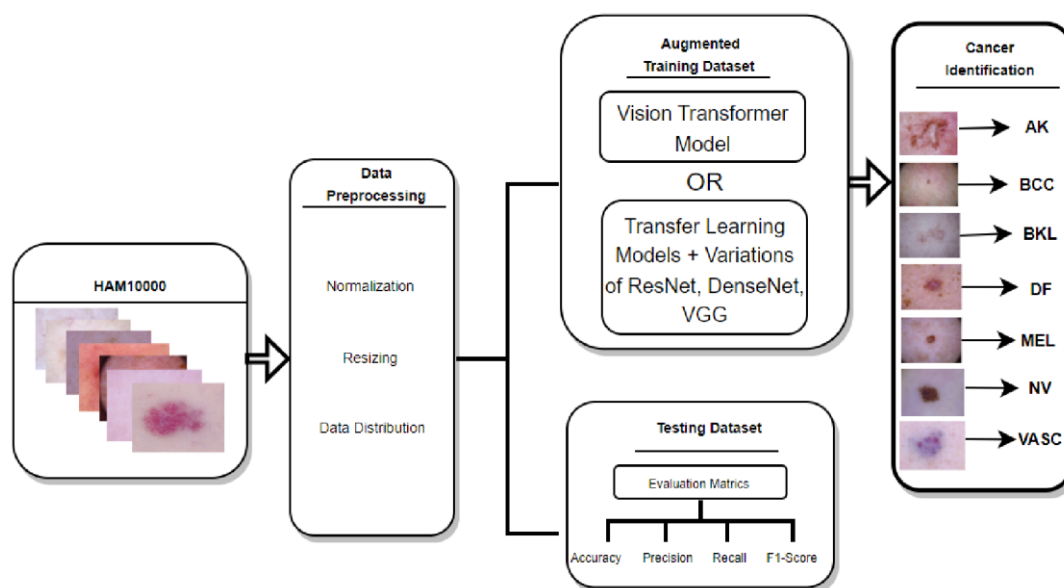
ViT، یک شبکه عصبی عمیق [۲۷]، برای وظایف تشخیص تصویر با پردازش تصاویر ورودی از طریق یک سری تبدیل‌های یادگرفته‌شده طراحی شده است. در تضاد با CNN‌های سنتی، ViT از یک مکانیسم توجه خودی (Self-Attention) برای تمرکز بر بخش‌های مرتبط تصویر ورودی استفاده می‌کند که منجر به دقت بالا در طیف متنوعی از وظایف تشخیص تصویر می‌شود.

در این مطالعه، تصاویر ورودی پس از تغییر اندازه به  $224 \times 224$  پیکسل، به قطعات  $16 \times 16$  تقسیم شدند. قطعات  $16 \times 16$  به فرآیند تقسیم یک تصویر ورودی به قطعات کوچک‌تر با اندازه ثابت اشاره دارد، جایی که هر قطعه ۱۶ پیکسل عرض و ۱۶ پیکسل ارتفاع داشت. این مدل بر روی ImageNet-21k آموزش دیده است، یک مجموعه داده طبقه‌بندی تصویر در مقیاس بزرگ با بیش از ۱۴ میلیون تصویر که به ۲۱,۸۴۱ دسته تقسیم شده‌اند. این مدل از ۱۲ لایه ترنسفورمر تشکیل شده است که هر کدام ۷۶۸ واحد پنهان و ۸۵.۸ میلیون پارامتر قابل آموزش دارند. برای مقادیر پارامترها و پیکربندی‌های مدل ViT، جدول ۲ را ببینید.

جدول ۲. پیکربندی‌های Vision Transformer (ViT)

مقدار	پارامتر
۷۶۸	ابعاد لایه‌های کُدگذار و ادغام (پولینگ)
۱۲	تعداد لایه‌های پنهان کُدگذار ترنسفورمر
۳۰۷۲	ابعاد لایه پیش‌خور (فید-فوروارد)
Gelu	تابع فعال‌سازی لایه پنهان
۰.۱	نرخ حذف لایه پنهان (دراپ‌اوت)
$224 \times 224$	اندازه تصویر
۳	تعداد کانال‌ها
$16 \times 16$	اندازه قطعات (پچ‌ها)
True	وضعیت متعادل‌سازی

شکل ۳ نمودار سطح انتزاعی روش‌شناسی پیشنهادی را نشان می‌دهد.



شکل ۳. نمودار سطح انتزاعی روش‌شناسی پیشنهادی

### ۳. نتایج و بحث

این بخش بحث جامعی در مورد معیارهای ارزیابی، جزئیات آزمایش‌ها و نتایج به دست آمده از طریق روش‌شناسی پیشنهادی ارائه می‌دهد.

#### ۳.۱. معیارهای ارزیابی

معیارهای ارزیابی نقش حیاتی در ارزیابی عملکرد مدل‌های یادگیری ماشین و یادگیری عمیق ایفا می‌کنند. این معیارها در قلمرو یادگیری ماشین، یادگیری عمیق و تحقیق آماری از اهمیت بالایی برخوردارند. در این مطالعه، ما بر چندین معیار ارزیابی کلیدی برای سنجش اثربخشی مدل پیشنهادی خود تمرکز کرده‌ایم:

- **دقت (Accuracy):** صحت کلی پیش‌بینی‌های مدل را اندازه‌گیری می‌کند. این معیار نسبت نمونه‌های به درستی طبقه‌بندی شده به کل نمونه‌ها را محاسبه می‌کند. دقت به تنهایی همیشه برای ارزیابی کافی نیست، به ویژه هنگامی که با مجموعه داده‌های نامتعادل سر و کار داریم یا هنگامی که انواع مختلف خطاها پیامدهای متفاوتی دارند.

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

- **صحت (Precision):** توانایی مدل را در شناسایی صحیح نمونه‌های مثبت در بین نمونه‌های پیش‌بینی شده مثبت کمی‌سازی می‌کند. این معیار نسبت مثبت‌های درست (TP) به مجموع مثبت‌های درست و مثبت‌های کاذب (FP) را محاسبه می‌کند. صحت بر قابلیت اطمینان پیش‌بینی‌های مثبت تمرکز دارد.

$$(2) \quad P = \frac{TP}{TP + FP}$$

- **احساس (Recall):** همچنین به عنوان حساسیت یا نرخ مثبت درست شناخته می‌شود، احساس، توانایی مدل را در شناسایی صحیح نمونه‌های مثبت در بین تمام نمونه‌های مثبت واقعی اندازه‌گیری می‌کند. این معیار نسبت مثبت‌های درست (TP) به مجموع مثبت‌های درست و منفی‌های کاذب (FN) را محاسبه می‌کند. احساس بر کامل بودن پیش‌بینی‌های مثبت تمرکز دارد.

$$(3) \quad R = \frac{TP}{TP + FN}$$

- **نمره F1:** میانگین هارمونیک صحت و احساس است. این معیار، متریکی واحد ارائه می‌دهد که هم صحت و هم احساس را متعادل می‌کند و آن را برای مواقعی که توزیع کلاس ناهمگن است یا تأکید یکسانی بر هر دو نوع خطا وجود دارد، مفید می‌سازد. نمره F1 بین ۰ تا ۱ متغیر است که ۱ بهترین عملکرد را نشان می‌دهد.

$$(4) \quad F1 = (2 \times P \times R) / (P + R)$$

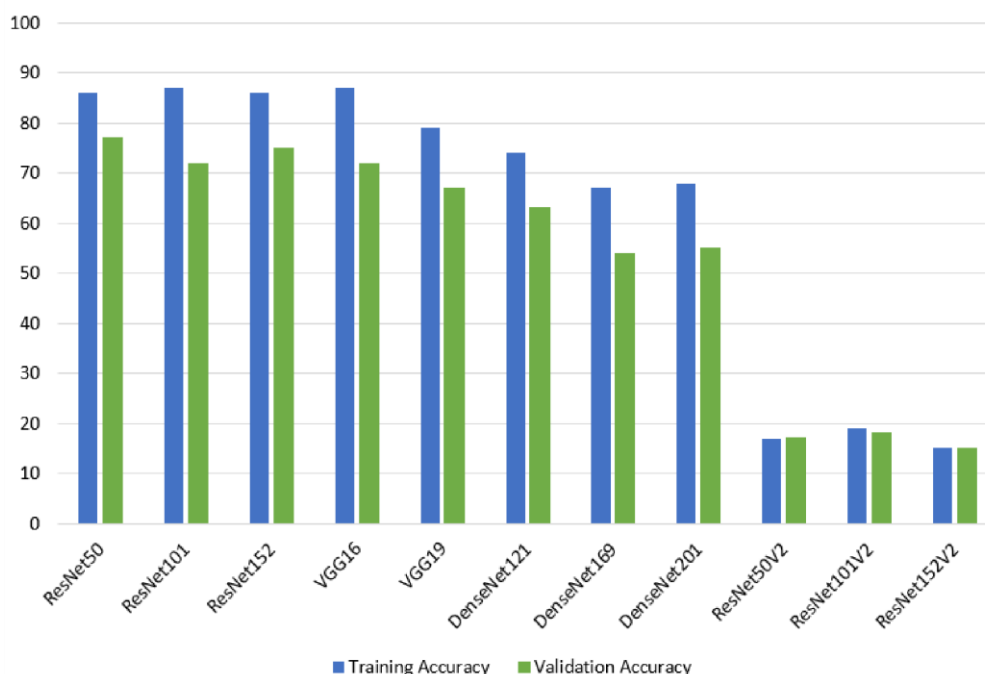


در طبقه‌بندی چندکلاسه، دقت به عنوان نسبت پیش‌بینی‌های صحیح (مثبت‌های درست و منفی‌های درست) به کل پیش‌بینی‌ها، صرف نظر از کلاس، محاسبه می‌شود. در مقایسه، صحت، احساس و F1 به صورت میانگین‌های وزنی برای طبقه‌بندی چندکلاسه در نظر گرفته می‌شوند. میانگین‌گیری وزنی به هر کلاس وزنی بر اساس نسبت آن در مجموعه داده اختصاص می‌دهد. برای به دست آوردن معیارهای وزنی، صحت، احساس و نمره F1 هر کلاس در وزن متناظر آن ضرب شده و سپس نتایج جمع و بر وزن کل تقسیم می‌شوند. این روش، عدم تعادل کلاس در مجموعه داده را در نظر می‌گیرد.

### ۳.۲. نتایج و بحث

هدف اصلی این مطالعه، طبقه‌بندی بیماری سرطان پوست در هفت کلاس بود. در این مطالعه، مدل با تصاویر عادی و افزایش‌یافته آموزش داده شد. از افزایش داده برای افزایش نمونه‌های آموزشی استفاده شد. مدل با ۱۴۰ تصویر (۲۰ تصویر از هر کلاس) اعتبارسنجی و با ۱۴۰ تصویر دیگر آزمایش شد. از تکنیک انتقال وزن با استفاده از وزن‌های ResNet50 که از مجموعه داده ImageNet بازیابی شده بود، بهره گرفته شد. لایه‌های شبکه ثابت (فریز) شدند و با ۱۰۲۴، ۵۱۲، ۲۵۶، ۱۲۸، ۶۴ و ۷ نرون از لایه‌های چگال به ترتیب تنظیم دقیق شدند. برای تعمیم‌پذیری مدل، نرمال‌سازی دسته‌ای و یک لایه حذف تصادفی با نرخ ۰.۳ نیز در این معماری استفاده شد. همانطور که از شکل ۴ مشاهده می‌شود، حداکثر دقت اعتبارسنجی ۷۷٪ و دقت آموزش ۸۶٪ به ترتیب با مدل ResNet50 تنظیم‌شده به دست آمد.

11-CNN Based Fine Tuned Pretrained Architectures Performance



شکل ۴. عملکردهای معماری مدل‌های از پیش آموزش‌دیده تنظیم‌شده مبتنی بر یادگیری انتقالی برای سرطان پوست.

ما در این مطالعه همه انواع ResNet را پیاده‌سازی کرده‌ایم. مدل همچنین با داده‌های آزمایشی متشکل از ۱۴۰ تصویر ارزیابی شد و به دقت آزمایشی ۷۵٪ دست یافتیم. مسئله طبقه‌بندی سرطان پوست به دلیل ماهیت مجموعه داده، چالشی بسیار دشوار است. تصاویر دسته‌های مختلف شبیه به هم به نظر می‌رسند که چالش‌های متعددی را برای مدل طبقه‌بندی ایجاد می‌کند. اگرچه مطالعات متعددی در زمینه طبقه‌بندی سرطان پوست منتشر شده است، اما استحکام مدل پیشنهادی ما همچنان دقیق‌تر ارزیابی می‌شود. تاکنون، بیشتر کارها بر روی تشخیص سرطان پوست (طبقه‌بندی دودویی) انجام شده است.

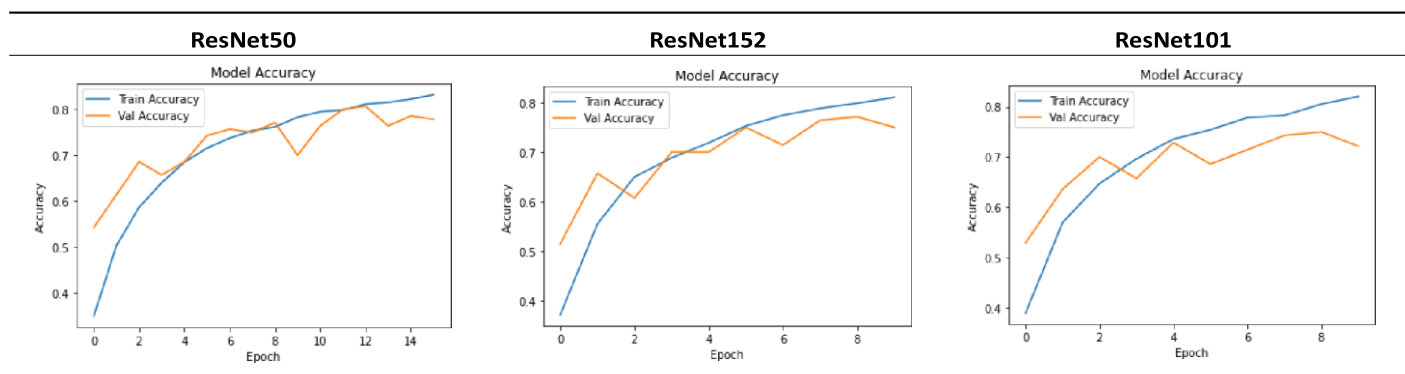
سوفی الف [۲۸] مطالعه‌ای برای طبقه‌بندی تصاویر ملانوما و غیرملانوما پیشنهاد کرد و به دقت ۸۳٪ دست یافت. در مطالعه هوسنی ک [۲۹]، او نیز رویکرد یادگیری انتقالی مدل AlexNet را برای طبقه‌بندی سرطان پوست در سه کلاس پیشنهاد داد. مدل AlexNet پیشنهادی، تصاویر سرطان پوست را به کلاس‌های ملانوما، خال معمولی و خال آتیپیک با دقت حدود ۹۷٪ طبقه‌بندی می‌کند. دورج یو [۳۰] نیز مطالعه‌ای را پیشنهاد کرد که سرطان پوست را در چهار کلاس با نمره دقت ۹۵٪ طبقه‌بندی می‌کند. نویسندگان [۳۰] از مجموعه داده ISIC برای طبقه‌بندی سرطان پوست با تنها دو کلاس ملانوما و غیرملانوما استفاده کردند.

از تحلیل مطالعات مختلف موجود در ادبیات این حوزه می‌توان دریافت که تحلیل مقایسه‌ای مطالعات طبقه‌بندی سرطان پوست نیز کاری چالش‌برانگیز است. مجموعه داده از مطالعه‌ای به مطالعه دیگر متفاوت است، که مقایسه منصفانه‌ای ارائه نمی‌دهد. علاوه بر این، تعداد کلاس‌ها در کارهای منتشر شده بسیار متغیر است. مقایسه مطالعات با مجموعه داده‌ها و تعداد کلاس‌های مختلف، تحلیل مقایسه‌ای دقیقی ارائه نخواهد داد.

در مطالعه بودیمن الف [۳۱]، طبقه‌بندی دودویی روی مجموعه داده ISIC انجام شد و نمره دقت ۰.۸۳ به همراه نمره مثبت بسیار پایین (۰.۴۶) به دست آمد. در همان حال، مدل تنظیم‌شده مبتنی بر یادگیری انتقالی ResNet50 در مطالعه ما دقت ۷۵٪ را بر روی هفت نوع کلاس سرطان پوست نشان داد. با افزایش مقادیر در متغیر هدف، دقت مدل آموزش‌دیده کاهش یافت. اما تعداد کلاس‌ها نیز در مطالعه پیشنهادی ما چندین برابر افزایش یافت و با این حال نمره دقت معقولی به دست آمد. بنابراین، مدل تنظیم‌شده ResNet50 پیشنهادی، افزوده‌ای مناسب برای طبقه‌بندی هفت نوع سرطان پوست در مقایسه با مدل از پیش آموزش‌دیده مبتنی بر CNN است. عملکرد تمام ۱۱ مدل از پیش آموزش‌دیده مبتنی بر CNN با پیکربندی یکسان را می‌توان در شکل ۴ مشاهده کرد.

نمودارهای یادگیری که فرآیندهای اعتبارسنجی و آموزش را به تصویر می‌کشند، برای کسب بینش نسبت به عملکرد مدل‌های یادگیری عمیق در حین آموزش ضروری هستند. این نمودارها به ما امکان می‌دهند رفتار مدل را نظارت و تحلیل کنیم، مسائل احتمالی بیش‌برازش را تشخیص دهیم، تنظیمات لازم را در معماری و ابرپارامترهای مدل ایجاد کنیم و در نهایت قابلیت اطمینان و دقت مدل‌های یادگیری عمیق را افزایش دهیم. برای نمایش بصری نمودارهای یادگیری، لطفاً به جدول ۳ مراجعه کنید که نمودارهای یادگیری برای سه مدل از پیش آموزش‌دیده برتر (ResNet50، ResNet101 و ResNet152) را نشان می‌دهد.

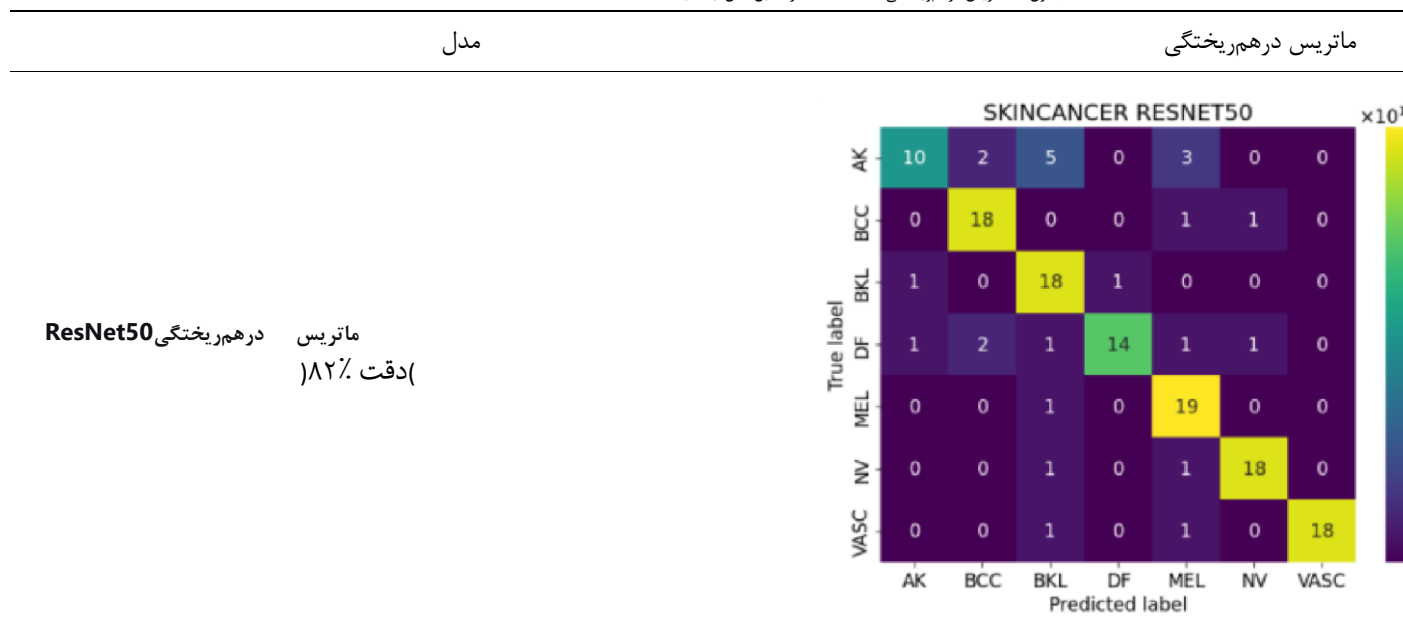
جدول ۳. نمودارهای یادگیری سه مدل برتر از پیش آموزش‌دیده مبتنی بر CNN



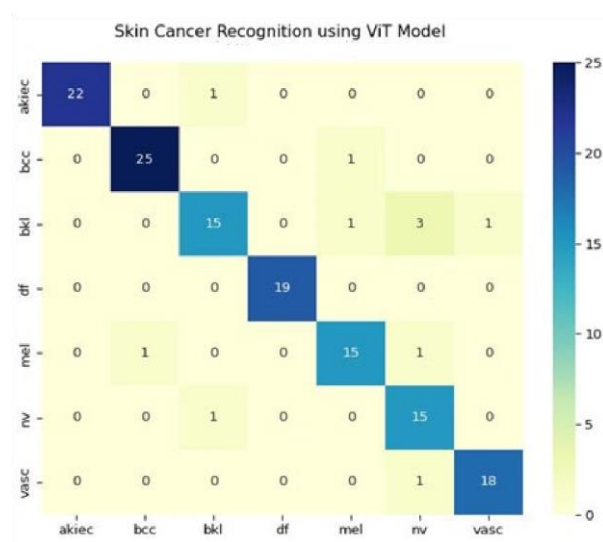
عملکرد ResNet50 از نظر دقت اعتبارسنجی در مقایسه با دیگر مدل‌های از پیش آموزش‌دیده متعارف برتر است ResNet50. از یک سری لایه‌های کانولوشنی برای استخراج ویژگی‌ها از تصاویر قبل از طبقه‌بندی آن‌ها با استفاده از لایه‌های کاملاً متصل بهره می‌برد، در حالی که ViT تصاویر را با استفاده از یک مکانیسم توجه خودی پردازش می‌کند. تصویر به قطعات (پیچ) تقسیم می‌شود و به هر قطعه مجموعه‌ای از بلوک‌های ترنسفورمر برای اعمال اختصاص داده می‌شود. به دلیل بلوک‌های ترنسفورمر و توجه خودی، مدل ViT در طبقه‌بندی تصاویر قدرتمندتر است. ما مجموعه داده متعادل را با نسبت آزمایش ۰.۲ برای مدل ViT در نظر گرفتیم و به دقت ۹۲.۱۴٪ و همچنین نمرات ۹۲.۶۱، ۹۲.۱۴ و ۹۲.۱۷ به ترتیب برای صحت، احساس و F1 دست یافتیم.

هنگامی که کلاس‌ها نامتعادل هستند یا زمانی که هزینه طبقه‌بندی اشتباه یک کلاس به طور قابل توجهی بیشتر از هزینه طبقه‌بندی اشتباه کلاس دیگر است، یک ماتریس درهم‌ریختگی (Confusion Matrix) می‌تواند به ارزیابی عملکرد یک مدل طبقه‌بندی کمک کند. یک ماتریس درهم‌ریختگی می‌تواند معیارهای دقت، صحت، احساس و نمره F1 را استخراج کند. ماتریس درهم‌ریختگی مقایسه‌ای ResNet50 و مدل ViT در جدول ۴ نمایش داده شده است.

جدول ۴. ماتریس درهم‌ریختگی ResNet50 در مقابل مدل Vision Transformer (ViT).



در اینجا معمولاً ماتریس درهم‌ریختگی ViT نمایش داده می‌شود و به دقت ۹۲.۱۴٪ و همچنین نمرات ۹۲.۶۱، ۹۲.۱۴ و ۹۲.۱۷ به ترتیب برای صحت، احساس و F1 دست یافتیم.



معیارهای صحت، احساس و نمره F1 به تفکیک کلاس نیز برای اعتبارسنجی مدل پیشنهادی استفاده شدند که در جدول ۵ نشان داده شده است. ستون پشتیبانی (Support) تعداد نمونه‌های هر کلاس را نشان می‌دهد. به عنوان مثال، کلاس AKIEC دارای ۲۳ نمونه برای آزمایش است در حالی که کلاس BCC دارای ۲۶ نمونه برای اهداف آزمایشی می‌باشد. مجموع ستون پشتیبانی برابر با ۱۴۰ است زیرا مدل را برای ۱۴۰ نمونه آزمایش کرده‌ایم.

جدول ۵- عملکرد ViT به تفکیک کلاس.

نام کلاس	Precision	Recall	F1	Support
AKIEC	1.0000	0.9565	0.9778	23
BCC	0.9615	0.9615	0.9615	26
BKL	0.8824	0.7500	0.8108	20
DF	1.0000	1.0000	1.0000	19
MEL	0.8824	0.8824	0.8824	17
NV	0.7500	0.9375	0.8333	16
VASC	0.9474	0.9474	0.9474	19

با توجه به مقایسه عملکرد مدل پیشنهادی، که در جدول ۶ نشان داده شده است، روش پیشنهادی از روش‌های پیشرفته موجود عملکرد بهتری دارد.

جدول ۶. مقایسه مطالعه پیشنهادی با مطالعات پیشرفته موجود (state-of-the-art).

Authors and Year	Classes	Method	Evaluation Metric	Results
(Ali et al., 2021) [32]	2	Custom CNN-Based Model named DCNN Proposed	Accuracy	Train 93.16% Test 91.93%
(Bassel et al., 2022) [14]	2	Stacking-CV (Proposed) + Xception Features	Accuracy	Test 90.9%
(Jain et al., 2021) [33]	7	Xception Net Transfer Learning-Based Model	Accuracy	Test 90.48%
(Ali et al., 2021) [34]	7	Efficient-Nets B0-B7 Transfer Learning-Based Models (Top Accuracy achieved with Efficient-Net B4)	Accuracy Precision Recall F1	87.91% 88% 88% 87%
(Huang et al., 2023) [35]	3	YoloV5 (RGB Images, HSI Images)	Accuracy Precision Recall F1 Specificity	<u>RGB</u> 79.2% 88.8% 75.8% 81.8% 79.8% <u>HSI</u> 78.7% 80% 72.6% 76.1% 78.6%
Proposed	7	Vision Transformers (RGB Images)	Accuracy Precision Recall F1	92.14% 92.61% 92.14% 92.17%

## ۴. نتیجه‌گیری

طبقه‌بندی سرطان پوست به دلیل ظواهر متنوع دسته‌های مختلف، وظیفه‌ای چالش‌برانگیز است. در این مطالعه، ما اثربخشی رویکرد Vision Transformer (ViT) و مدل‌های از پیش آموزش‌دیده CNN را برای طبقه‌بندی چندکلاسه سرطان پوست بررسی کردیم. با به کارگیری تکنیک‌های تنظیم دقیق، یادگیری انتقالی و افزایش داده، به نتایج چشمگیری دست یافتیم.

مدل ViT از مدل‌های یادگیری انتقالی مبتنی بر CNN پیشی گرفت و به ترتیب به دقت ۹۲.۱۴٪، صحت ۹۲.۶۱٪، احساس ۹۲.۱۴٪ و نمره F1 معادل ۹۲.۱۷٪ دست یافت. با این حال، بهبودهای بیشتری در تکنیک‌های پیش‌پردازش برای مدل‌های یادگیری عمیق مورد نیاز است.

به طور کلی، مطالعه ما پتانسیل رویکرد ViT و مدل‌های از پیش آموزش‌دیده CNN را برای طبقه‌بندی قابل اطمینان سرطان پوست برجسته می‌کند.

**Author Contributions:** Conceptualization, M.A.A., S.M. and M.I.; methodology, S.M., M.A.A., M.T. and S.A.; validation, S.A., M.T., M.S. and M.A.A.; investigation, S.A., S.M. and M.A.A.; data curation, M.A.A. and M.I.; writing—original draft preparation, M.A.A. and S.A.; writing—review and editing, S.A., M.A.A., M.S., M.T. and M.I. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The dataset HAM10000 used in this study for experiments is openly available to download from Kaggle.

**Conflicts of Interest:** The authors declare no conflict of interest.

#### Abbreviations

AKIEC AK	Actinic Keratoses
BCC	Basal cell carcinoma
BKL	Benign keratosis-like lesions
DF	Dermatofibroma
MEL	Melanoma
NV	Melanocytic nevi
TP	True positive
FP False positive	TN True negative
FN	False negative
ViT	Vision tranformer
MAX	Maximum
MIN	Minumum
AVG	Average
HSI	Hyperspectral imaging
RGB	Red, green, and blue (color images)

#### References

1. Cancer. Available online: <https://www.who.int/news-room/fact-sheets/detail/cancer> (accessed on 4 August 2022).
2. Cancer—NHS. Available online: <https://www.nhs.uk/conditions/cancer/> (accessed on 4 August 2022).
3. Melanoma—The Skin Cancer Foundation. Available online: <https://www.skincancer.org/skin-cancer-information/melanoma/> (accessed on 8 July 2023).
4. Arroyo, J.L.G.; Zapirain, B.G. Automated Detection of Melanoma in Dermoscopic Images. In *Computer Vision Techniques for the Diagnosis of Skin Cancer*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 139–192. [CrossRef]
5. Pomponiu, V.; Nejati, H.; Cheung, N.-M. Deepmole: Deep neural networks for skin mole lesion classification. *Proc. Int. Conf. Image Process.* **2016**, *2016*, 2623–2627. [CrossRef]
6. Guo, Y.; Liu, Y.; Oerlemans, A.; Lao, S.; Wu, S.; Lew, M.S. Deep learning for visual understanding: A review. *Neurocomputing* **2016**, *187*, 27–48. [CrossRef]
7. Li, K.M.; Li, E.C. Skin Lesion Analysis Towards Melanoma Detection via End-to-end Deep Learning of Convolutional Neural Networks. *arXiv* **2018**, arXiv:1807.08332.
8. Li, H.; Pan, Y.; Zhao, J.; Zhang, L. Skin disease diagnosis with deep learning: A review. *Neurocomputing* **2021**, *464*, 364–393. [CrossRef]
9. Goyal, M.; Knackstedt, T.; Yan, S.; Hassanpour, S. Artificial intelligence-based image classification methods for diagnosis of skin cancer: Challenges and opportunities. *Comput. Biol. Med.* **2020**, *127*, 104065. [CrossRef] [PubMed]

10. Kumar, V.B.; Kumar, S.S.; Saboo, V. Dermatological Disease Detection Using Image Processing and Machine Learning. In Proceedings of the 2016 3rd International Conference on Artificial Intelligence and Pattern Recognition (AIPR), Lodz, Poland, 19–21 September 2016; pp. 88–93. [\[CrossRef\]](#)
11. Esteva, A.; Kuprel, B.; Novoa, R.A.; Ko, J.; Swetter, S.M.; Blau, H.M.; Thrun, S. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **2017**, *542*, 115–118. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Kawahara, J.; Daneshvar, S.; Argenziano, G.; Hamarneh, G. Seven-Point Checklist and Skin Lesion Classification Using Multitask Multimodal Neural Nets. *IEEE J. Biomed. Health Inform.* **2019**, *23*, 538–546. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Chandrashekar, G.; Sahin, F. A survey on feature selection methods. *Comput. Electr. Eng.* **2014**, *40*, 16–28. [\[CrossRef\]](#)
14. Haenssle, H.A.; Fink, C.; Schneiderbauer, R.; Toberer, F.; Buhl, T.; Blum, A.; Kalloo, A.; Hassen, A.B.H.; Thomas, L.; Enk, A.; et al. Man against machine: Diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Ann. Oncol.* **2018**, *29*, 1836–1842. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Bassel, A.; Abdulkareem, A.B.; Alyasseri, Z.A.A.; Sani, N.S.; Mohammed, H.J. Automatic Malignant and Benign Skin Cancer Classification Using a Hybrid Deep Learning Approach. *Diagnostics* **2022**, *12*, 2472. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Jeny, A.A.; Sakib, A.N.M.; Junayed, M.S.; Lima, K.A.; Ahmed, I.; Islam, M.B. SkNet: A Convolutional Neural Networks Based Classification Approach for Skin Cancer Classes. In Proceedings of the ICCIT 2020 23rd International Conference on Computer and Information Technology (ICCIT), Dhaka, Bangladesh, 19–21 December 2020. [\[CrossRef\]](#)
17. Tabrizchi, H.; Parvizpour, S.; Razmara, J. An Improved VGG Model for Skin Cancer Detection. *Neural Process. Lett.* **2022**, 1–18. [\[CrossRef\]](#)
18. Skin Cancer MNIST: HAM10000|Kaggle. Available online: <https://www.kaggle.com/datasets/kmader/skin-cancer-mnistham10000> (accessed on 11 September 2022).
19. Residual Neural Network (ResNet). Available online: <https://iq.opengenus.org/residual-neural-networks/> (accessed on 11 August 2022).
20. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. 2016, pp. 770–778. Available online: <http://image-net.org/challenges/LSVRC/2015/> (accessed on 13 July 2023).
21. Dropout Regularization in Neural Networks: How It Works and When to Use It—Programmatically. Available online: <https://programmatically.com/dropout-regularization-in-neural-networks-how-it-works-and-when-to-use-it/> (accessed on 12 August 2022).
22. What Are Hyperparameters? and How to Tune the Hyperparameters in a Deep Neural Network?|by Pranoy Radhakrishnan|Towards Data Science. Available online: <https://towardsdatascience.com/what-are-hyperparameters-and-how-to-tune-the-hyperparameters-in-a-deep-neural-network-d0604917584a> (accessed on 18 August 2022).
23. Activation Functions in Neural Networks—GeeksforGeeks. Available online: <https://www.geeksforgeeks.org/activation-functions-neural-networks/> (accessed on 18 August 2022).
24. What, Why and Which?? Activation Functions|by Snehal Gharat|Medium. Available online: <https://medium.com/@snaily16/what-why-and-which-activation-functions-b2bf748c0441> (accessed on 18 August 2022).
25. Gentle Introduction to the Adam Optimization Algorithm for Deep Learning. Available online: <https://machinelearningmastery.com/adam-optimization-algorithm-for-deep-learning/> (accessed on 18 August 2022).
26. Long, M.; Cao, Y.; Wang, J.; Jordan, M.I.; Edu, J. Learning Transferable Features with Deep Adaptation Networks. *PMLR* **2015**, *37*, 97–105.
27. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image is Worth 16 × 16 Words: Transformers for Image Recognition at Scale. *arXiv* **2020**, arXiv:2010.11929.
28. Sufi, A. Skin Cancer Classification Using Deep Learning. 2022. Available online: <http://dspace.uiu.ac.bd/handle/52243/2483> (accessed on 15 September 2022).
29. Hosny, K.M.; Kassem, M.A.; Foad, M.M. Skin Cancer Classification using Deep Learning and Transfer Learning. In Proceedings of the 2018 9th Cairo International Biomedical Engineering Conference (CIBEC), Cairo, Egypt, 20–22 December 2018; pp. 90–93. [\[CrossRef\]](#)
30. Dorj, U.O.; Lee, K.K.; Choi, J.Y.; Lee, M. The skin cancer classification using deep convolutional neural network. *Multimed. Tools Appl.* **2018**, *77*, 9909–9924. [\[CrossRef\]](#)
31. Budhiman, A.; Suyanto, S.; Arifianto, A. Melanoma Cancer Classification Using ResNet with Data Augmentation. In Proceedings of the 2019 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), Yogyakarta, Indonesia, 5–6 December 2019; pp. 17–20. [\[CrossRef\]](#)
32. Ali, M.S.; Miah, M.S.; Haque, J.; Rahman, M.M.; Islam, M.K. An enhanced technique of skin cancer classification using deep convolutional neural network with transfer learning models. *Mach. Learn. Appl.* **2021**, *5*, 100036. [\[CrossRef\]](#)
33. Jain, S.; Singhan, U.; Tripathy, B.; Nasr, E.A.; Aboudaif, M.K.; Kamrani, A.K. Deep Learning-Based Transfer Learning for Classification of Skin Cancer. *Sensors* **2021**, *21*, 8142. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Ali, K.; Shaikh, Z.A.; Khan, A.A.; Laghari, A.A. Multiclass skin cancer classification using EfficientNets—A first step towards preventing skin cancer. *Neurosci. Inform.* **2022**, *2*, 100034. [\[CrossRef\]](#)
35. Huang, H.Y.; Hsiao, Y.P.; Mukundan, A.; Tsao, Y.M.; Chang, W.Y.; Wang, H.C. Classification of Skin Cancer Using Novel Hyperspectral Imaging Engineering via YOLOv5. *J. Clin. Med.* **2023**, *12*, 1134. [\[CrossRef\]](#) [\[PubMed\]](#)

---

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.