



Sahand University
of Technology

DOR:

[20.1001.1.23223146.1405.12.2.6.5](https://doi.org/10.1001.1.23223146.1405.12.2.6.5)

Journal of Nonlinear
Systems in Electrical
Engineering

Vol. 12, No. 2

Autumn and Winter 2025

ISSN: 2322 – 3146

<http://jnsee.sut.ac.ir>

Hybrid Deep–Texture Feature Fusion Network for Robust Melanoma Detection in Dermoscopic Images

Nafiseh Gholami, Mansoor Zeinali *

Department of Computer Engineering, Na.C, Islamic Azad University, Najafabad,
Iran

* Corresponding author email: mansoor.zeinali@iau.ac.ir

Abstract

Melanoma is one of the most serious forms of skin cancer, and early detection can support appropriate treatment selection and reduce adverse outcomes. This study presents a two-branch framework for dermoscopic image classification that integrates high-level features extracted by DenseNet121 with local texture features derived from the Local Neighborhood Difference Pattern (LNDP) operator. Images were selected from the ISIC 2019 and HAM10000 datasets, and the task was formulated as binary classification between melanoma and non-melanoma lesions. After resizing images to $224 \times 224 \times 3$ pixels, applying normalization, and performing data augmentation only on the training set, the 1024-dimensional DenseNet121 feature vector was concatenated with a 59-dimensional LNDP histogram. The resulting 1083-dimensional vector was fed into a fully connected classifier with 128 neurons, ReLU activation, dropout with a rate of 0.5, and a Softmax output layer. Performance was evaluated using accuracy, ROC-AUC, PR-AUC, balanced accuracy, MCC, and melanoma recall. The proposed model achieved 97% accuracy, a ROC-AUC of 0.98, balanced accuracy of 0.94, and melanoma recall of 0.90. Ablation and cross-dataset evaluations indicated that fusing deep and texture features improved discriminative performance compared with DenseNet121 alone. Nevertheless, independent external validation and calibration assessment are required before clinical use.

Keywords

Melanoma; DenseNet; LNDP Operator

1. Short Introduction

Melanoma is an aggressive form of skin cancer for which early and accurate diagnosis is essential. Although dermoscopy improves lesion assessment, diagnosis remains challenging because of visual similarity between benign and malignant lesions, image artifacts, and data imbalance. Deep convolutional networks such as DenseNet can learn high-level semantic representations but may not fully capture fine local texture patterns.

2. Proposed Work

The proposed method is a two-stream framework that combines high-level semantic features from DenseNet121 with low-level texture features extracted using the LNDP operator.

Dermoscopic images are resized, normalized, and augmented during training to improve the model's generalization ability. The 1024-dimensional DenseNet121 feature vector and the 59-dimensional LNDP histogram are individually normalized and concatenated into a 1083-dimensional feature vector.

This fused representation is passed through a fully connected layer with 128 ReLU neurons, followed by dropout and a two-class softmax output layer.

The model classifies dermoscopic images into melanoma and benign classes and is evaluated using accuracy, precision, recall, F1-score, ROC-AUC, PR-AUC, balanced accuracy, and MCC.

3. Conclusion

The results indicate that combining DenseNet121 deep features with LNDP texture features improves melanoma classification compared with using either feature type alone.

The proposed model achieved 97% accuracy and a ROC-AUC of 0.98 on the evaluated dermoscopic datasets.

The fusion strategy improved the model's ability to represent both high-level semantic patterns and fine-grained local textures.

Cross-dataset evaluation showed that the proposed method maintained better performance than the DenseNet121 baseline under dataset distribution shifts.

Nevertheless, further validation on larger, independent, and clinically diverse datasets is required before clinical application can be considered.

کنترل تشخیص ملانوما در تصاویر درموسکوپي با استفاده از هم‌جوشي ويژگي‌هاي عميق DenseNet121 و ويژگي‌هاي بافتي LNDP

نفيسه غلامي^۱، منصور زينلي^{۲،۳*}

^۱ دانشجوي کارشناسي ارشد، گروه مهندسي کامپيوتر، واحد نجف آباد، دانشگاه آزاد اسلامي، نجف آباد، ايران

^۲ استاديار، گروه مهندسي برق، واحد نجف آباد، دانشگاه آزاد اسلامي، نجف آباد، ايران

^۳ مرکز تحقيقات پردازش ديگيتال و بينايي ماشين، واحد نجف آباد، دانشگاه آزاد اسلامي، نجف آباد، ايران

* آدرس پست الكترونيكي نويسنده مسئول: mansoor.zeinali@iau.ac.ir



دانشگاه صنعتي سهند

DOR:

[20.1001.1.23223146.1405.12.2.6.5](https://doi.org/10.1001.1.23223146.1405.12.2.6.5)

نشریه سالانه‌اي غيرخطي «مندی برق»

دوره ۱۲ - شماره ۲

پاييز و زمستان ۱۴۰۴

صفحات ۱۰۹ الي ۱۳۳

ISSN: 2322-3146

<http://jnsee.sut.ac.ir>

تاريخ پذيرش: ۱۴۰۵/۶/۱۶

تاريخ بازنگري: ۱۴۰۵/۵/۲۵

تاريخ ارسال: ۱۴۰۴/۱۱/۲۸

چکیده

واژه‌های کلیدی

مبدل افزاینده

طبقه‌بندی ملانوما

شبکه عصبی عمیق

ویژگی‌های سطح پایین

عملگر LNDP

تصاویر درموسکوپي

ملانوما یکی از جدی‌ترین انواع سرطان پوست است و تشخیص زودهنگام آن می‌تواند در انتخاب درمان مناسب و کاهش پیامدهای بیماری مؤثر باشد. در این پژوهش، یک چارچوب دوشاخه برای طبقه‌بندی تصاویر درموسکوپي ارائه می‌شود که ویژگی‌های سطح بالای استخراج‌شده از DenseNet121 را با ویژگی‌های بافتی محلی حاصل از عملگر Local Neighborhood Difference Pattern (LNDP) ادغام می‌کند. تصاویر از مجموعه داده‌های ISIC 2019 و HAM10000 انتخاب شدند و مسئله به صورت دودویی شامل ملانوما و ضایعات غیرملانوما تعریف شد. پس از تغییر اندازه تصاویر به 224×224 پیکسل، نرمال‌سازی و اعمال افزایش داده فقط بر مجموعه آموزش، بردار 1024 بعدی DenseNet121 با هیستوگرام 59 بعدی LNDP از طریق الحاق برداری ترکیب شد. بردار نهایی 1083 بعدی به یک طبقه‌بند تمام‌متصل شامل 128 نورون، تابع فعال‌سازی ReLU، Dropout با نرخ 0.5 و لایه خروجی Softmax وارد شد. ارزیابی با معیارهای دقت، ROC-AUC، PR-AUC، Balanced Accuracy، MCC و Recall کلاس ملانوما انجام گرفت. مدل پیشنهادی به دقت برابر 97% درصد، ROC-AUC برابر 0.98 ، Balanced Accuracy برابر 0.94 و Recall ملانوما برابر 0.90 دست یافت. نتایج مطالعه حذف مؤلفه‌ها و ارزیابی بین مجموعه‌ای نشان داد که ادغام ویژگی‌های عمیق و بافتی، نسبت به DenseNet121 تنها، عملکرد تفکیکی بهتری ایجاد می‌کند. با این حال، برای کاربرد بالینی، اعتبارسنجی خارجی مستقل و ارزیابی دقیق‌تر کالیبراسیون ضروری است.

۱- مقدمه

سرطان پوست يکي از شايع‌ترين بيماري‌هاي پوستي است و ملانوما به‌عنوان بدخيم‌ترين نوع آن، سهم قابل توجهي از مرگ‌ومير سرطان‌هاي پوستي را به خود اختصاص مي‌دهد [۱]. اين سرطان از ملانوسيت‌ها منشأ مي‌گيرد و عواملی مانند تابش فرابنفش، زمينه‌هاي ژنتيکي و برخي عوامل محيطي مي‌توانند زمينه‌ساز بروز آن باشند [۲]. بر اساس گزارش IARC وابسته به سازمان بهداشت جهاني، در سال ۲۰۲۲ بيش از ۱/۵ ميليون مورد جديد سرطان پوست و حدود ۳۳۰ هزار مورد جديد ملانوما در جهان گزارش شده است و نزديک به ۶۰ هزار مرگ نيز به آن نسبت داده مي‌شود [۳]. اين آمار اهميت تشخيص زودهنگام را برجسته مي‌کند، زيرا تاخير در تشخيص مي‌تواند خطر پيشرفت بيماري و مرگ‌ومير را به‌طور جدی افزايش دهد. تشخيص ملانوما در ابتدا به‌شدت به مهارت و تجربه پزشک بستگي دارد و عوامل مختلفی مانند شباهت‌هاي ظاهري بين ضايعات خوش خيم و بدخيم، کيفيت پايين تصاوير و وجود نويزهاي خارجي، باعث ايجاد مشکلات در تشخيص دقيق مي‌شود. با اينکه سيستم‌هاي CAD در سال‌هاي اخير در اين زمينه معرفي شده‌اند، هنوز مشکلاتي در تشخيص الگوهاي پيچيده بافتي و تفاوت‌هاي جزئي ضايعات وجود دارد [۲].

روش‌هاي سنتي تشخيص شامل معاينه بصري، درموسکوپي و در نهايت بيوپسي هستند [۴]. اگرچه درموسکوپي يک روش تصويربرداري غيرتهاجمي است که با بزرگ‌نمايي و حذف بازتاب‌هاي سطحي، امکان مشاهده دقيق‌تر ساختار ضايعات پوستي را فراهم مي‌کند [۵]، اما تشخيص نهايي همچنان به قضاوت متخصص پوست بستگي دارد. مطالعات نشان داده‌اند که دقت تشخيص توسط متخصصان پوست آموزش ديده بين ۶۰ تا ۸۰ درصد متغير است و به ميزان تجربه و مهارت فرد وابستگي زيادي دارد [۶]. از سوي ديگر، شباهت بصري زياد ميان ضايعات خوش خيم و بدخيم، تنوع درون کلاسي ضايعات و همچنين وجود نويزهايي مانند مو يا حباب ژل در تصاوير درموسکوپي، تشخيص دقيق را براي متخصصان و حتي مدل‌هاي خود کار دشوار مي‌سازد [۷].

با پيشرفت فناوري‌هاي پردازش تصوير و يادگيري ماشين، سيستم‌هاي تشخيص به کمک رايانه (CAD) براي شناسايي ضايعات پوستي معرفي شده‌اند. اين سيستم‌ها معمولاً شامل چهار مرحله اصلي: پيش‌پردازش، تشخيص يا بخش‌بندي ضايعه، استخراج ويژگي و در نهايت طبقه‌بندي هستند [۸]. با اينکه مدل‌هاي يادگيري عميق مانند DenseNet قادر به استخراج ويژگي‌هاي پيچيده از تصاوير درموسکوپي هستند، بيشتر اين مدل‌ها تا به امروز بر داده‌هاي عمومي و تصاوير بزرگ‌شده از ضايعات پوستي متمرکز بوده‌اند. چالش‌هاي همچنان باقي مانده شامل مشکلات در شبیه‌سازي دقيق تشخيص ملانوما به‌ويژه در شرايط واقعي مانند کيفيت پايين تصوير و شباهت‌هاي بصري ميان ضايعات است [۴]. در مطالعات اوليه، ويژگي‌هاي سطح پايين مانند رنگ، شکل و بافت به‌عنوان ورودی الگوريتم‌هاي کلاسيکي نظير ماشين بردار پشتيبان (SVM) يا درخت تصميم مورد استفاده قرار گرفتند [۹]. هرچند اين روش‌ها در برخي موارد عملکرد قابل قبولي داشتند، اما به دليل وابستگي زياد به انتخاب ويژگي‌ها و محدوديت در نمايش اطلاعات پيچيده، دقت آن‌ها در مقياس بزرگ رضايتمند بخش نبود [۱۰].

ظهور يادگيري عميق و به‌ويژه شبکه‌هاي عصبي پيچشي (CNN)، نقطه عطفی در پردازش تصاوير پزشکی به شمار مي‌رود [۱۱]. اگرچه بسياري از مدل‌ها از جمله شبکه‌هاي DenseNet پيشرفت‌هاي قابل توجهي در تشخيص ملانوما داشته‌اند، اما مشکلاتي در پردازش تصاوير با کيفيت پايين، داده‌هاي نامتوازن و شباهت‌هاي ظاهري بالا ميان ضايعات خوش خيم و بدخيم همچنان وجود دارد. استفاده هم‌زمان از ويژگي‌هاي سطح پايين و سطح بالا مي‌تواند اين چالش‌ها را کاهش دهد و عملکرد مدل‌ها را بهبود بخشد [۵]. اين شبکه‌ها قادرند به‌صورت خودکار ويژگي‌هاي سطح بالا را از داده‌هاي خام استخراج کنند و در بسياري از زمينه‌ها از جمله تشخيص ملانوما عملکرد بهتري نسبت به روش‌هاي سنتي نشان دهند [۱۲]. با اين حال، استفاده مستقيم از شبکه‌هاي از پيش آموزش ديده مانند VGG16، ResNet50 يا DenseNet که معمولاً بر روي مجموعه داده‌هايي مانند ImageNet آموزش ديده‌اند، با چالش‌هايي همراه است: نخست اينکه اين مدل‌ها براي تصاوير عمومي بهينه شده‌اند و ممکن است در استخراج الگوهاي خاص ضايعات پوستي

عملکرد ضعیف‌تری داشته باشند. دوم آنکه، پیچیدگی محاسباتی بالای این شبکه‌ها موجب افزایش زمان آموزش و استنتاج می‌شود [۱۳].

از سوی دیگر، مطالعات گذشته نشان داده‌اند که ترکیب ویژگی‌های سطح بالا و سطح پایین می‌تواند عملکرد مدل‌ها را به‌طور چشمگیری بهبود بخشد [۱۴]. ویژگی‌های سطح پایین مانند بافت یا الگوهای محلی، به‌ویژه در ضایعاتی که شباهت زیادی به یکدیگر دارند، می‌توانند مکمل مناسبی برای ویژگی‌های عمیق استخراج‌شده توسط شبکه‌های CNN باشند. در همین راستا، عملگرهای مبتنی بر بافت مانند Local Neighborhood Difference Pattern (LNDP) ابزار مناسبی برای توصیف تغییرات شدت پیکسل‌ها در نواحی کوچک تصویر هستند و می‌توانند اطلاعات ارزشمندی درباره ساختار بافتی ضایعه ارائه دهند [۱۵].

بر این اساس، پژوهش حاضر مدلی ترکیبی بر پایه شبکه DenseNet و ویژگی‌های LNDP ارائه می‌دهد تا با ادغام ویژگی‌های چندسطحی، دقت و پایداری در تشخیص خودکار ملانوما افزایش یابد. نوآوری این پژوهش نه در به‌کارگیری صرف DenseNet121 یا LNDP، بلکه در راهبرد فیوژن ناهمگن دومسیره است؛ به‌گونه‌ای که ویژگی‌های معنایی عمیق و ویژگی‌های بافتی محلی به‌صورت مکمل و بدون نیاز به قطعه‌بندی صریح ضایعه یا بهینه‌سازی فراابتکاری، از طریق اتصال برداری ادغام می‌شوند. عملکرد روش پیشنهادی بر روی دو مجموعه داده معتبر ISIC 2019 و HAM10000 ارزیابی می‌شود. همچنین، مقایسه عملکرد مدل ترکیبی با هر یک از مسیرهای DenseNet121 و LNDP به‌صورت مستقل، امکان تحلیل سهم هر مؤلفه و ارزیابی اثربخشی راهبرد هم‌جوشي را فراهم می‌سازد.

ساختار ادامه مقاله به این صورت است: در بخش دوم، پژوهش‌های مرتبط در زمینه تشخیص خودکار ملانوما، شبکه‌های عمیق و توصیفگرهای بافتی مرور می‌شوند. در بخش سوم، روش پیشنهادی، استخراج ویژگی‌های عمیق و بافتی و سازوکار هم‌جوشي تشریح می‌شود. بخش چهارم به معرفی مجموعه داده‌ها، تنظیمات آزمایش و تحلیل نتایج اختصاص دارد. در نهایت، بخش پنجم جمع‌بندی پژوهش، محدودیت‌ها و پیشنهادهایی برای مطالعات آینده را ارائه می‌کند.

۲- مروری بر روش‌های تشخیص ملانوما

تشخیص ملانوما در طول زمان با بهره‌گیری از ابزارها و رویکردهای متنوعی انجام شده است که شامل دو دسته زیر هستند:

(الف) روش‌های بالینی و ابزارهای پزشکی:

یکی از روش‌های سنتی، معاینه چشمی توسط پزشک است که با وجود سرعت و هزینه پایین، به‌شدت به تجربه و مهارت فرد متخصص وابسته است و احتمال بروز خطای انسانی در آن بالاست [۱۶]. درموسکوپي (Dermoscopy) به‌عنوان یکی از ابزارهای غیرتهاجمی پرکاربرد، امکان مشاهده لایه‌های زیرین پوست با بزرگ‌نمایی و نوردهی خاص را فراهم می‌آورد و نسبت به معاینه چشمی دقت بیشتری دارد. با این حال، همچنان نیازمند آموزش تخصصی و وابسته به اپراتور است [۱۷].

در مجموع، وابستگی شدید روش‌های بالینی به تجربه پزشک و محدودیت‌های ذاتی ابزارهای تشخیصی، زمینه را برای بروز خطا فراهم می‌کند و لزوم بهره‌گیری از سامانه‌های تشخیص خودکار و مستقل از عامل انسانی را برجسته می‌سازد.

(ب) رویکردهای محاسباتی:

در حوزه پردازش تصویر و یادگیری ماشین، الگوریتم‌های کلاسیک پردازش تصویر برای استخراج ویژگی‌های بافتی، رنگی و شکلی به کار رفته‌اند. هرچند این روش‌ها می‌توانند در تفسیر خودکار مفید باشند، اما نسبت به نویز و کیفیت پایین تصاویر حساس‌اند

[۱۲]. روش‌های یادگیری ماشین سنتی مانند ماشین بردار پشتیبان (SVM) و جنگل تصادفی (RF) نیز به کار گرفته شده‌اند که در صورت وجود داده‌های خوب، عملکرد قابل قبولی دارند؛ با این حال نیازمند مهندسی ویژگی دستی هستند [۱۸].

با وجود پیشرفت‌های قابل توجه در روش‌های محاسباتی، وابستگی روش‌های کلاسیک به مهندسی ویژگی و ناتوانی آن‌ها در نمایش الگوهای پیچیده، موجب شده است که این رویکردها به تنهایی پاسخگوی چالش‌های تشخیص ملانوما نباشند.

پیشرفت‌های اخیر در یادگیری عمیق به‌ویژه شبکه‌های عصبی پیچشی (CNN) و معماری‌های پیشرفته مانند DenseNet، امکان تشخیص الگوهای پیچیده و خودکارسازی فرایند تشخیص را فراهم کرده‌اند. در سال‌های اخیر، معماری‌های جدید مانند Vision Transformers (ViT) و ConvNeXt در زمینه تشخیص ضایعات پوستی به کار گرفته شده‌اند و توانسته‌اند به نتایج قابل توجهی دست یابند [۱، ۲]. این مدل‌ها با استفاده از خودتوجهی (Self-Attention) و ویژگی‌های چندمقیاسی توانسته‌اند الگوهای پیچیده‌تری را از داده‌های تصویری استخراج کنند. به‌ویژه ViT در پردازش تصاویر پزشکی توانسته است دقت بالایی در تشخیص انواع ضایعات پوستی از جمله ملانوما ارائه دهد و در مقایسه با مدل‌های CNN پیشین، مزایای محسوسی را به همراه داشته است. این مدل‌ها دقتی در حد متخصصان پوست به دست آورده‌اند [۱۹]، [۲۰]. مطالعات اولیه نشان داد که شبکه‌های عصبی پیچشی توانایی دستیابی به عملکردی در سطح متخصصان پوست را دارند [۱۲]؛ ولی این مدل‌ها در استخراج دقیق الگوهای بافتی محلی همچنان با محدودیت‌هایی مواجه‌اند. با وجود پیشرفت‌های قابل توجه در حوزه یادگیری ماشین و عمیق، تشخیص خودکار ملانوما همچنان با چالش‌هایی مواجه است؛ از جمله:

عدم توازن داده‌ها: در مجموعه داده‌های درموسکوپ، تعداد تصاویر خوش‌خیم به مراتب بیشتر از تصاویر بدخیم است و این موضوع باعث بروز مشکل در آموزش مدل‌ها می‌شود [۲۱].

شباهت بصری بین ضایعات: بسیاری از ضایعات خوش‌خیم و بدخیم شباهت‌های ظاهری زیادی دارند که تشخیص دقیق را دشوار می‌سازد.

کیفیت متغیر تصاویر: وجود نویز، تار بودن یا مصنوعات (مانند مو یا بازتاب نور) در تصاویر، عملکرد الگوریتم‌ها را کاهش می‌دهد.

این محدودیت‌ها و چالش‌های مربوط به کیفیت داده‌ها و عدم توازن کلاس‌ها نشان‌دهنده اهمیت توسعه رویکردهای ترکیبی هستند که بتوانند با استفاده از شبکه‌های عمیق (برای یادگیری الگوهای پیچیده) و ویژگی‌های سطح پایین (برای افزایش دقت و پایداری) به عملکرد بهینه دست یابند. این ایده در چارچوب‌های ناهمگن تشخیص سرطان پوست مورد توجه قرار گرفته و نتایج موثری در بهبود عملکرد سیستم‌های طبقه‌بندی گزارش شده است [۲۲]. در یک مطالعه مرتبط دیگر، Rao و Vasumathi چارچوبی چندمرحله‌ای برای قطعه‌بندی و تشخیص سرطان پوست ارائه کردند. در این چارچوب، ابتدا نویز تصاویر با استفاده از فیلتر میانه کاهش یافت و سپس ضایعات پوستی به کمک شبکه R2U-Net قطعه‌بندی شدند. تنظیم فرآیندهای R2U-Net با استفاده از الگوریتم بهینه‌سازی مصنوعی Namib beetle یا ANBO انجام گرفت. پس از قطعه‌بندی، برای کاهش اثر عدم توازن داده‌ها، راهبردهای مختلف تقویت داده اعمال شد و ویژگی‌های آماری، شکلی، SLBT، ویژگی‌های عمیق CNN، ویژگی‌های Haralick و LNDP استخراج شدند. در مرحله نهایی، DenseNet برای تشخیص سرطان پوست به کار رفت و فرآیندهای آن نیز با استفاده از ANBO تنظیم شدند [۲۳]. این چارچوب از نظر پوشش مراحل پردازش و تنوع ویژگی‌ها جامع است و مزیت آن، استفاده هم‌زمان از قطعه‌بندی ضایعه، استخراج ویژگی‌های متنوع و تنظیم فرآیندهای شبکه است. با این حال، وابستگی آن به چندین مرحله پردازشی شامل حذف نویز، قطعه‌بندی، تقویت داده، استخراج مجموعه‌های مختلف ویژگی و بهینه‌سازی فراابتکاری، موجب افزایش پیچیدگی

محاسباتی و دشواری بازتولید می‌شود. همچنین، استفاده هم‌زمان از خانواده‌های متعدد ویژگی، تعیین سهم مستقل هر نوع ویژگی را در عملکرد نهایی دشوار می‌سازد.

از سوی دیگر، مدل‌های multi-scale fusion اخیراً در زمینه تشخیص ضایعات پوستی مورد توجه قرار گرفته‌اند. این مدل‌ها با ترکیب ویژگی‌های مقیاس‌های مختلف تصاویر، توانسته‌اند عملکرد مدل‌های یادگیری عمیق را در تحلیل دقیق‌تر ویژگی‌ها و بافت‌ها ارتقا دهند [۴]. همچنین، استفاده از cross-dataset evaluation در مطالعات اخیر به منظور سنجش قدرت تعمیم مدل‌ها و عملکرد آن‌ها در مجموعه داده‌های مختلف توسعه یافته است. این رویکرد به‌ویژه برای ارزیابی مدل‌ها در شرایط واقعی، که تفاوت‌های قابل توجهی بین مجموعه‌های داده مختلف وجود دارد، ضروری است [۵]. اگرچه پژوهش‌های متعددی به ترکیب ویژگی‌های عمیق یا ادغام چندمقیاسی پرداخته‌اند، اغلب آن‌ها از فیوژن همگن (homogeneous fusion) یعنی ادغام چندلایه از یک شبکه عمیق واحد یا ترکیب چند شبکه مشابه استفاده کرده‌اند. در مقابل، روش حاضر بر پایه یک فیوژن ناهمگن میان دو خانواده کاملاً متفاوت از ویژگی‌ها یعنی ویژگی‌های بافتی LNDP از جنس توصیفگرهای کلاسیک محلی، و ویژگی‌های سطح بالای استخراج شده از DenseNet121 طراحی شده است. این ترکیب، ضعف مدل‌های فوق در نمایش الگوهای بافتی ریز را جبران می‌کند و نوآوری اصلی این مطالعه را شکل می‌دهد. جدول ۱ مقایسه‌ای بین این روش‌ها را نشان می‌دهد.

جدول ۱: مقایسه روش‌های مختلف تشخیص ملانوما

مرجع	محدودیت‌ها	مزایا	نوع روش	روش تشخیص
[۱۶]	وابستگی به تجربه پزشک و احتمال خطای انسانی	سریع و کم‌هزینه	سنتی	معاینه چشمی پزشک
[۱۷]	نیازمند آموزش تخصصی و وابستگی به اپراتور	مشاهده بهتر ساختارهای ضایعه و دقت بیشتر نسبت به معاینه چشمی	ابزار تشخیصی غیرتهاجمی	درموسکوپي
[۱۲]	حساسیت به نویز و کیفیت پایین تصاویر	استخراج ویژگی‌های بافتی، رنگی و شکلی	محاسباتی	روش‌های کلاسیک پردازش تصویر
[۱۹]	نیاز به مهندسی ویژگی دستی	عملکرد مناسب در صورت طراحی ویژگی مناسب	محاسباتی	یادگیری ماشین سنتی، مانند SVM و RF
[۲۰]	نیاز به داده و توان محاسباتی بیشتر	استخراج خودکار ویژگی‌های پیچیده	یادگیری عمیق	شبکه‌های عمیق، مانند CNN و DenseNet
[۲۳]	نیاز به تنظیم مناسب دو مسیر ویژگی	ترکیب ویژگی‌های معنایی و بافتی	ترکیبی	روش پیشنهادی DenseNet121 + LNDP

۳- مواد و روش‌ها

شبکه‌های عصبی مصنوعی (Artificial Neural Networks - ANN) از پیشرفته‌ترین روش‌های یادگیری ماشین هستند که الهام گرفته از ساختار نورون‌های مغز انسان طراحی شده‌اند. این مدل‌ها برای شبیه‌سازی سیستم‌های پیچیده و غیرخطی توسعه یافته و در سال‌های اخیر تحولی بنیادین در حوزه‌هایی مانند بینایی ماشین و تحلیل تصاویر پزشکی ایجاد کرده‌اند [۲۴]. در ساده‌ترین حالت،

یک ANN شامل سه بخش اصلی است: لایه ورودی، لایه‌های پنهان و لایه خروجی. هر نورون در لایه‌های پنهان، ورودی‌ها را از نورون‌های لایه قبلی دریافت کرده و پس از اعمال وزن‌ها و توابع فعال‌سازی (مانند ReLU، Sigmoid یا Tanh)، خروجی‌هایی به لایه بعدی ارسال می‌کند [۲۵].

پیشرفت توان محاسباتی (به‌ویژه GPU ها)، افزایش حجم داده‌های تصویری و توسعه الگوریتم‌های آموزش، منجر به ظهور شبکه‌های عصبی عمیق (Deep Neural Networks - DNNs) شد. این مدل‌ها با داشتن لایه‌های پنهان متعدد، قادرند ویژگی‌های پیچیده‌تری را به‌صورت سلسله‌مراتبی از داده‌ها استخراج کنند [۲۶]. شبکه‌های عمیق در حوزه‌های گوناگون مانند پردازش تصویر، گفتار و زبان طبیعی به نتایج چشمگیری دست یافته‌اند. مطالعات متعددی نشان داده‌اند که عملکرد این شبکه‌ها در تشخیص ضایعات پوستی از جمله ملانوما، قابل مقایسه یا حتی دقیق‌تر از متخصصان پوست است [۱۲، ۲۷، ۲۸].

۱-۳- DenseNet و مزایای آن

یکی از معماری‌های برجسته در حوزه یادگیری عمیق، شبکه‌های عصبی پیچشی متراکم یا DenseNet (Densely Connected Convolutional Networks) معرفی شد [۲۹]. ویژگی اصلی این معماری اتصال مستقیم هر لایه به تمامی لایه‌های بعدی است؛ به‌طوری‌که خروجی هر لایه به‌عنوان ورودی برای تمام لایه‌های بعدی مورد استفاده قرار می‌گیرد. این ساختار موجب تقویت جریان اطلاعات، بهبود انتشار گرادیان و استفاده مجدد از ویژگی‌ها می‌شود و در نهایت با کاهش تعداد پارامترها، کارایی مدل افزایش می‌یابد.

معماری DenseNet از دو بخش اصلی تشکیل شده است:

الف) بلوک‌های متراکم: لایه‌ها در هر بلوک به‌طور متوالی و متراکم با یکدیگر ارتباط دارند و ویژگی‌های استخراج شده از تمامی لایه‌های قبلی به لایه بعد منتقل می‌شود.

ب) لایه‌های انتقالی: این لایه‌ها با استفاده از عملیاتی مانند پیچش و ادغام، ابعاد ویژگی‌ها را کاهش داده و از افزایش بیش از حد پارامترها جلوگیری می‌کنند [۳۰]. مزایای کلیدی این معماری عبارت‌اند از:

کاهش مشکل ناپدید شدن گرادیان: اتصال مستقیم لایه‌ها، امکان انتقال مؤثر گرادیان به لایه‌های ابتدایی را فراهم می‌کند و آموزش شبکه‌های عمیق‌تر را تسهیل می‌نماید [۲۹].

باز استفاده از ویژگی‌ها: با اشتراک‌گذاری ویژگی‌ها بین لایه‌ها، DenseNet قادر است ویژگی‌های متنوعی را با تعداد پارامتر کمتر یاد بگیرد [۳۱].

بهبود دقت و کارایی: مطالعات نشان داده‌اند که DenseNet نسبت به معماری‌هایی مانند ResNet در بسیاری از وظایف طبقه‌بندی تصویر عملکرد بهتری دارد، به‌ویژه در شرایطی که داده‌های آموزشی محدود هستند [۳۲].

در حوزه تحلیل تصاویر پزشکی، از جمله تشخیص ملانوما، معماری DenseNet به دلیل توانایی بالای آن در استخراج ویژگی‌های پیچیده و ظریف از تصاویر درموسکوپي، عملکرد بسیار مناسبی داشته است [۱۲]. با توجه به این ویژگی‌ها، DenseNet گزینه‌ای مناسب برای استخراج نمایش‌های سطح بالای تصاویر درموسکوپي محسوب می‌شود، اما به‌تنهایی قادر به مدل‌سازی دقیق الگوهای بافتی ریز نیست.

۳-۲- ویژگی‌های سطح پایین و عملگر LNDP

ویژگی‌های سطح پایین (Low-Level Features) در پردازش تصویر به آن دسته از ویژگی‌ها که مستقیماً از مقادیر پیکسلی قابل استخراج هستند، بدون نیاز به تحلیل معنایی یا سطح بالا گفته می‌شود. این ویژگی‌ها شامل رنگ، بافت، لبه‌ها، زاویه‌ها و شدت روشنایی می‌شوند و معمولاً به عنوان ورودی الگوریتم‌های بینایی ماشین یا یادگیری ماشین مورد استفاده قرار می‌گیرند [۳۳]. در تصاویر درموسکوپی پوست، الگوهای بافتی ضایعات (مانند لکه‌های تیره یا حاشیه‌های نامنظم) اهمیت ویژه‌ای دارند و می‌توانند با توصیفگرهای بافتی همچون LBP (Local Binary Pattern) و نسخه‌های بهبود یافته آن استخراج شوند [۳۴].

LNDP (Pattern Local Neighborhood Difference) یک توصیفگر بافتی پیشرفته است که به عنوان جایگزینی برای LBP معرفی شده است. در فرایند LNDP، یک پنجره 3×3 روی تصویر حرکت می‌کند و اختلاف شدت بین پیکسل‌های همسایه محاسبه می‌شود. درحالی که LBP تنها اختلاف شدت بین پیکسل مرکزی و همسایه‌ها را در نظر می‌گیرد، LNDP روابط میان همسایه‌ها را نیز بررسی می‌کند. این ویژگی باعث می‌شود LNDP توانایی بالاتری در شناسایی الگوهای ریز و پیچیده داشته باشد و نسبت به تغییرات محلی تصویر حساس‌تر عمل کند [۳۵]. این نوع اطلاعات اغلب توسط شبکه‌های عمیق در لایه‌های بالاتر نادیده گرفته می‌شوند، زیرا شبکه‌ها در سطوح بالاتر بیشتر بر ویژگی‌های انتزاعی و مفهومی تمرکز دارند. LNDP الگوهای دودویی مبتنی بر روابط بین همسایه‌ها تولید کرده و آن‌ها را در قالب هیستوگرام‌های بافتی ذخیره می‌کند [۳۶]. مزایای LNDP در مقایسه با LBP عبارت‌اند از:

- مقاومت بیشتر در برابر نویز.
- توانایی بالاتر در توصیف بافت‌های پیچیده.
- پایداری بهتر در شرایط نوری متغیر.
- کاربرد مؤثرتر در تحلیل تصاویر پزشکی.

انتخاب و به کارگیری ویژگی‌های مناسب برای نمایش دقیق داده‌ها، در فرایند طبقه‌بندی تصاویر پزشکی نقش حیاتی دارد. در این میان، ترکیب ویژگی‌های سطح پایین کلاسیک با ویژگی‌های استخراج شده از شبکه‌های عصبی عمیق به عنوان رویکردی کارآمد در بهبود عملکرد سیستم‌های یادگیری شناخته شده است [۳۷]. چون معماری‌هایی مانند DenseNet توانایی یادگیری نمایش‌های پیچیده و سطح بالا از داده‌ها را دارند و الگوهای مکانی - معنایی گسترده‌تری را در تصاویر شناسایی می‌کنند [۳۸].

۴- روش پیشنهادی

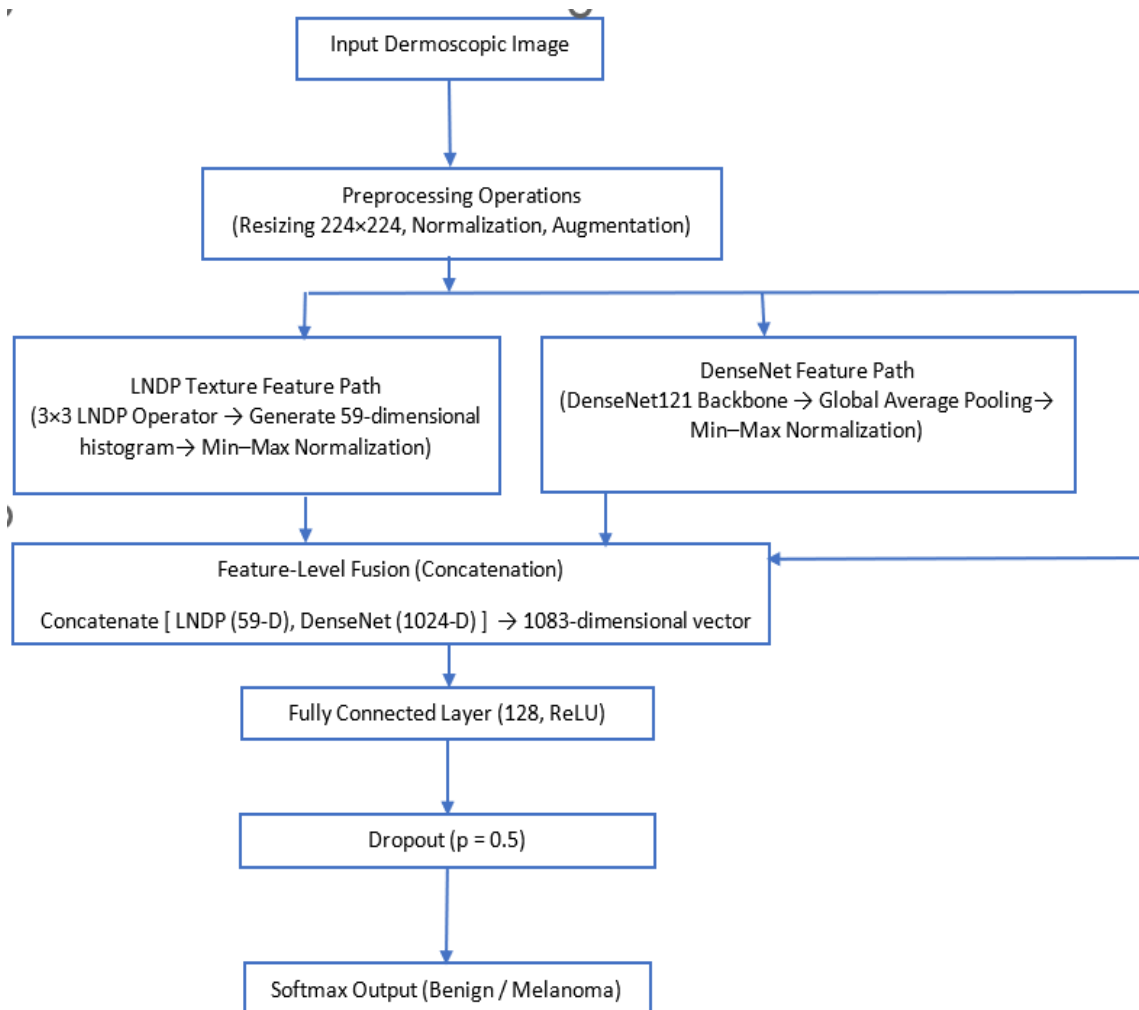
چنانچه بیان شد؛ LNDP به عنوان ابزاری مؤثر برای استخراج ویژگی‌های اولیه از تصاویر درموسکوپی مورد استفاده قرار گرفته است. این ویژگی‌ها می‌توانند ساختارهای ریز ضایعه مانند نقاط تیره، شبکه‌های بافتی یا مرزهای نامنظم را آشکار کنند که در تشخیص ملانوما اهمیت دارند [۳۷]. از سوی دیگر، شبکه‌های عصبی عمیق مانند DenseNet در استخراج ویژگی‌های سطح بالای معنایی بسیار توانمند هستند. ترکیب این دو رویکرد (ویژگی‌های سطح پایین LNDP و ویژگی‌های سطح بالای شبکه عمیق) می‌تواند به صورت مکمل عمل کرده و دقت نهایی در طبقه‌بندی ملانوما را افزایش دهد. پژوهش حاضر نیز بر همین مبنا از این هم‌افزایی بهره می‌برد.

ترکیب این دو نوع ویژگی به گونه‌ای عمل می‌کند که ویژگی‌های سطح پایین نقش مکملی برای ویژگی‌های سطح بالا ایفا کرده و نمایشی جامع‌تر از تصویر ارائه می‌دهند. نتایج پژوهش‌های اخیر نشان داده است که چنین هم‌جوشی ویژگی‌ها به ویژه در حوزه

پزشکی که داده‌ها اغلب با مشکلاتی نظیر نویز، عدم تعادل کلاس‌ها و شباهت ظاهری بین ضایعات مواجه‌اند، باعث افزایش دقت و کاهش خطای طبقه‌بندی می‌شود [۳۸, ۳۹].

در این بخش، معماری روش پیشنهادی معرفی می‌شود. نقطه تمایز این روش نسبت به چارچوب‌های مشابه، نه صرف استفاده از دو خانواده ویژگی LNDP و DenseNet، بلکه راهبرد ترکیب دومسیره آن‌هاست که در آن ویژگی‌های بافتی محلی و ویژگی‌های معنایی عمیق به‌عنوان دو منبع مکمل، بدون واسطه‌گری قطعه‌بندی یا بهینه‌سازی فراابتکاری، مستقیماً از طریق اتصال برداری ادغام می‌شوند.

الگوریتم روش پیشنهادی در شکل ۱ نمایش داده شده است. جزئیات مراحل این روش در ادامه توضیح داده می‌شود.



شکل ۱: بلوک دیاگرام روش پیشنهادی برای طبقه‌بندی نهایی

مرحله ۱- پیش پردازش داده‌ها

به‌منظور آماده‌سازی تصاویر جهت ورود به مدل، مراحل زیر انجام گرفت:

- تصاویر ورودی ابتدا به ابعاد 224×224 پیکسل تغییر اندازه داده شدند تا با ورودی استاندارد شبکه DenseNet121 سازگار باشند. سپس، برای افزایش تنوع داده‌ها و کاهش احتمال بیش‌برازش (Overfitting)، از تکنیک‌های افزایش داده‌ها (Data Augmentation) استفاده شد. این تکنیک‌ها شامل چرخش تصادفی تصاویر در بازه $\pm 20^\circ$ درجه، تغییر

مقیاس (Zoom) در بازه ۰/۹ تا ۱/۱، جابجایی تصادفی تصاویر به اندازه ۱۰٪ در جهت‌های افقی و عمودی و وارونگی افقی (Horizontal Flip) با احتمال ۰/۵ بودند. این تغییرات به‌طور خاص برای داده‌های درموسکوپی مناسب هستند زیرا این نوع تصاویر معمولاً دارای تنوع زیادی از جهات و زوایای مختلف هستند و این تغییرات به مدل کمک می‌کند تا به‌طور مؤثرتر ویژگی‌های بافتی و مرزی ضایعات پوستی را شبیه‌سازی کرده و از سوگیری‌ها جلوگیری کند.

- سپس نرمال‌سازی شدت روشنایی و رنگ برای کاهش اثرات شرایط تصویربرداری مختلف انجام گرفت.
- افزودن داده‌ها (Data Augmentation) فقط روی تصاویر مجموعه آموزش اعمال شد و شامل: چرخش تصادفی در بازه $\pm 20^\circ$ درجه، تغییر مقیاس (Zoom) در بازه ۰/۹ تا ۱/۱، جابجایی افقی/عمودی تا ۱۰٪ ابعاد تصویر، و وارونگی افقی (Horizontal Flip) با احتمال ۰/۵ بود. این تبدیلات برای جلوگیری از نشت اطلاعات، روی مجموعه‌های اعتبارسنجی و آزمون اعمال نشدند.

مرحله ۲- استخراج ویژگی‌ها

در این پژوهش دو دسته ویژگی از تصاویر استخراج می‌شوند:

۱- ویژگی‌های سطح پایین بافتی: از عملگر LNDP (Local Neighborhood Difference Pattern) برای استخراج ویژگی‌های بافتی محلی استفاده گردید. این روش توانایی بالایی در شناسایی ساختارهای ریز و تغییرات شدت نور در ضایعات پوستی دارد.

۲- ویژگی‌های سطح بالا: از معماری DenseNet121 به‌عنوان یک شبکه عصبی پیچشی عمیق بهره گرفته شد DenseNet. قادر است ویژگی‌های معنایی پیچیده‌تری را از تصاویر استخراج کند و بازاستفاده از ویژگی‌ها را بهینه نماید.

مرحله ۳- ترکیب ویژگی‌ها و طبقه‌بندی

در مدل پیشنهادی، برای بهره‌گیری همزمان از اطلاعات سطح بالا و سطح پایین، از روش اتصال برداری (Vector Concatenation) استفاده شده است. روش اتصال برداری به این دلیل انتخاب شد که این روش به‌صورت مؤثر اطلاعات هر دو مسیر ویژگی‌ها (ویژگی‌های بافتی LNDP و ویژگی‌های معنایی DenseNet) را در یک نمای ترکیبی جمع‌آوری می‌کند. این روش یکی از روش‌های ساده و قابل فهم برای ادغام ویژگی‌ها است که به‌ویژه در سیستم‌های یادگیری عمیق با داده‌های ناهمگون مانند ویژگی‌های سطح پایین و سطح بالا مناسب است. انتخاب این روش به دلیل سادگی، قابلیت پیاده‌سازی و حفظ اطلاعات اصلی در هر دو نوع ویژگی بوده است. هرچند روش‌های پیچیده‌تری مانند تنظیم وزنی ویژگی‌ها یا Attention Mechanism می‌توانند برای ادغام ویژگی‌ها استفاده شوند، اما در این مطالعه اتصال برداری به‌عنوان روشی اثربخش و ساده انتخاب شد که به‌طور مؤثر عملکرد مدل را افزایش داد. بدین صورت که خروجی DenseNet (با طول ۱۰۲۴) و ویژگی‌های بافتی LNDP (با طول ۵۹) به‌صورت خطی در کنار یکدیگر قرار گرفته و بردار نهایی با طول ۱۰۸۳ تشکیل می‌شود. برای ایجاد قابلیت ادغام مناسب میان دو مسیر استخراج ویژگی، ابعاد و نحوه نرمال‌سازی بردارها به‌صورت دقیق استانداردسازی شد. تصاویر درموسکوپی به‌صورت خاکستری تبدیل شدند و سپس در پنجره‌هایی به‌اندازه $3 \times 3 \times 3$ ، مقادیر LNDP محاسبه شد. ویژگی‌های بافتی استخراج‌شده توسط LNDP به‌صورت یک هیستوگرام ۵۹ بعدی تولید شدند. از سوی دیگر، تصاویر درموسکوپی پس از انجام مراحل پیش‌پردازش، به‌عنوان ورودی به شبکه DenseNet121 داده می‌شوند و ویژگی‌های سطح بالا از خروجی آخرین لایه پیچشی استخراج می‌گردند. به‌منظور کاهش ابعاد فضایی و تبدیل نقشه‌های ویژگی به یک بردار فشرده، از لایه Global Average Pooling استفاده شده است.

خروجی این لایه یک بردار ویژگی با طول ۱۰۲۴ بُعد است که اطلاعات معنایی و انتزاعی تصویر را در خود جای داده و نمایانگر ویژگی‌های سطح بالای استخراج‌شده توسط شبکه عمیق می‌باشد.

به‌منظور یکسان‌سازی مقیاس دو بردار و جلوگیری از غلبه ویژگی‌های یک مسیر بر مسیر دیگر، هر دو بردار پیش از ادغام تحت نرمال‌سازی Min-Max در بازه [۰, ۱] قرار گرفتند. پارامترهای نرمال‌سازی min-max (کمینه و بیشینه) صرفاً بر اساس داده‌های مجموعه آموزش محاسبه شده و سپس بدون هیچ‌گونه بازبرازش، بر روی داده‌های اعتبارسنجی و آزمون اعمال شده‌اند تا از نشت اطلاعات جلوگیری شود:

$$x_{norm} = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (1)$$

در مرحله‌ی ادغام، این دو بردار حاصل از ویژگی‌های LNDP و DenseNet به‌صورت اتصال برداری ترکیب شدند و یک بردار ناهمگن با ابعاد $1083 = 1024 + 59$ تشکیل شد. این بردار، ورودی لایه تمام متصل با ۱۲۸ نورون قرار گرفت که پس از اعمال تابع فعال‌سازی ReLU و Dropout روی آن، به لایه Softmax جهت پیش‌بینی نهایی منتقل شد که نهایتاً خروجی مدل به صورت یکی از دو کلاس «ملانوما» یا «خوش‌خیم» ارائه می‌شود. بنابراین، ساختار نهایی طبقه‌بند به‌صورت زیر است:

Input (1083) → Dense (128, ReLU) → Dropout (p=0.5) → Dense (2, Softmax)

استفاده از لایه‌ی Dropout به میزان ۵۰٪ به‌منظور جلوگیری از بیش‌برازش (Overfitting) انجام گرفت.

لازم به ذکر است که تشابه به‌کارگیری LNDP و DenseNet بین پژوهش حاضر و مطالعه Rao و Vasumathi به انتخاب همین دو خانواده ویژگی محدود می‌شود و نوآوری پژوهش حاضر در این تشابه خلاصه نمی‌شود. در روش مرجع، LNDP یکی از چندین گروه ویژگی (آماري، شکلي، SLBT، CNN، Haralick و LNDP) است که پس از قطعه‌بندی R2U-Net و بهینه‌سازی ANBO در یک زنجیره چندمرحله‌ای استخراج و ترکیب می‌شوند؛ درحالی‌که در روش حاضر، LNDP کنار ویژگی‌های عمیق DenseNet121 به‌عنوان دو مؤلفه مکمل یک معماری دومسیره مطرح است و ادغام آن‌ها از طریق اتصال برداری، بدون قطعه‌بندی صریح و بدون بهینه‌سازی فراابتکاری انجام می‌شود.

۵- شبیه‌سازی‌ها و فرایند آموزش مدل

۵-۱- آموزش سیستم

مدل پیشنهادی با استفاده از کتابخانه‌های TensorFlow و Keras پیاده‌سازی شد. برای آموزش شبکه، از الگوریتم Adam optimizer نرخ یادگیری اولیه ۴-۱۰ و تابع هزینه Cross-Entropy Loss استفاده گردید. فرایند آموزش به‌مدت ۵۰ دوره (Epoch) با اندازه دسته‌ای (Batch Size) برابر با ۳۲ انجام شد تا توازنی میان سرعت پردازش و پایداری گرادینت‌ها ایجاد شود. همچنین برای جلوگیری از بیش‌برازش (Overfitting)، روش‌هایی مانند Early Stopping و Dropout به‌کار گرفته شد.

در این پژوهش از مجموعه داده‌های عمومی ISIC 2019 و HAM10000 استفاده شد [۴۱] و [۴۰]. در این پژوهش، تصاویر درموسکوپ از مجموعه داده ISIC 2019 Challenge استفاده شده‌اند. دلیل انتخاب این نسخه آن است که ISIC 2019 یکی از کامل‌ترین مجموعه داده‌ها از نظر تنوع ضایعات پوستی، کیفیت تصاویر و تعداد نمونه‌های دارای برچسب معتبر است. به‌طور مشخص، از نسخه رسمی منتشرشده در سال ۲۰۱۹ استفاده شده و هیچ داده‌ای از نسخه‌های ISIC 2018 یا ISIC 2020 در این تحقیق مورد استفاده قرار نگرفته است. از آنجا که مجموعه داده ISIC 2019 به‌صورت چند کلاسه برچسب‌گذاری شده است، در این پژوهش مسئله به‌صورت دودویی تعریف شد (Melanoma در برابر Benign). بر این اساس، تصاویر دارای برچسب «Melanoma» به‌عنوان

کلاس بدخیم در نظر گرفته شدند و سایر کلاس‌های غیر ملانوما در ISIC 2019 در یک کلاس واحد تحت عنوان «Benign» تجمع شدند. در مجموعه HAM10000 نیز تصاویر با برچسب «melanoma (mel)» در کلاس Melanoma قرار گرفتند و سایر برچسب‌ها به عنوان Benign در نظر گرفته شدند. این موضوع برای قابلیت تکرارپذیری نتایج و مقایسه معتبر با مطالعات پیشین ضروری است. همچنین داده‌های تکمیلی از مجموعه HAM10000 به کار گرفته شده‌اند تا تنوع و حجم نمونه‌های ملانوما و ضایعات خوش خیم افزایش یابد.

برای ایجاد شفافیت کامل در توازن داده‌ها، تعداد تصاویر خوش خیم و بدخیم در هر پایگاه داده به صورت جداگانه استخراج و گزارش شد. در مجموعه داده ISIC 2019، مجموعاً ۸۲۰۰ تصویر مورد استفاده قرار گرفت که شامل ۶۵۰۰ تصویر ضایعه خوش خیم و ۱۷۰۰ تصویر ملانوما بود. در مجموعه داده HAM10000 نیز ۲۸۰۰ تصویر شامل ۲۲۸۰ تصویر خوش خیم و ۵۲۰ تصویر ملانوما به کار گرفته شد. برای جلوگیری از نشت اطلاعات، تقسیم داده‌ها در سطح بیمار یا ضایعه انجام شد. تمام تصاویر متعلق به یک بیمار یا یک ضایعه فقط در یکی از مجموعه‌های آموزش، اعتبارسنجی یا آزمون قرار گرفتند. همچنین نمونه‌های تکراری و تصاویر مشترک میان پایگاه‌های داده شناسایی و حذف شدند. پس از ادغام و پیش‌پردازش داده‌ها، مجموعه نهایی شامل ۱۰۰۰۰۰ تصویر مرکب از ۸۷۸۰ تصویر خوش خیم و ۱۲۲۰ تصویر بدخیم بود. با توجه به ماهیت داده‌های واقعی درموسکوپی، مجموعه نهایی مورد استفاده در این پژوهش ذاتاً نامتوازن بوده است. برای جلوگیری از سوگیری در فرایند تقسیم داده‌ها، از روش تقسیم لایه‌بندی شده استفاده شد تا نسبت کلاس‌ها در مجموعه‌های آموزش، اعتبارسنجی و آزمون ثابت بماند. بنابراین، مجموعه آزمون نیز به صورت نامتوازن باقی مانده و شامل ۱۳۱۷ نمونه خوش خیم و ۱۸۳ نمونه ملانوما است. تمامی معیارهای عملکرد و گزارش‌های آماری ارائه شده در ادامه، بر اساس همین توزیع واقعی داده‌های آزمون محاسبه شده‌اند. تقسیم‌بندی نهایی داده‌ها چنین است: ۷۰٪ داده‌ها شامل ۶۱۴۶ تصویر خوش خیم و ۸۵۴ تصویر بدخیم در مرحله آموزش، ۱۵٪ داده‌ها شامل ۱۳۱۷ تصویر خوش خیم و ۱۸۳ تصویر بدخیم در مرحله اعتبارسنجی و ۱۵٪ تصاویر در مرحله آزمون استفاده شدند.

۲-۵- تحلیل عملکرد مدل پیشنهادی

در این پژوهش، مدل پیشنهادی به منظور ارزیابی عملکرد، پنج بار با بذره‌های تصادفی متفاوت آموزش داده شد تا پایداری و نوسان عملکرد مدل به صورت دقیق سنجیده شود. تمامی معیارهای گزارش شده شامل Accuracy، Precision، Recall و F1-score بر اساس میانگین $\pm$ انحراف معیار محاسبه شده‌اند.

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \quad (۲)$$

$$Precision(p) = TP / (TP + FP) \quad (۳)$$

$$Recall(r) = TP / (TP + FN) \quad (۴)$$

$$F1_score = 2pr / (p + r) \quad (۵)$$

TP و FN، FP، TN به ترتیب مخفف «منفی واقعی»، «مثبت کاذب»، «منفی کاذب» و «مثبت واقعی» یا تشخیص درست ملانوما هستند. از آنجا که داده‌های آزمون نامتوازن هستند، علاوه بر معیارهای متداول، معیارهای Balanced Accuracy، ضریب همبستگی

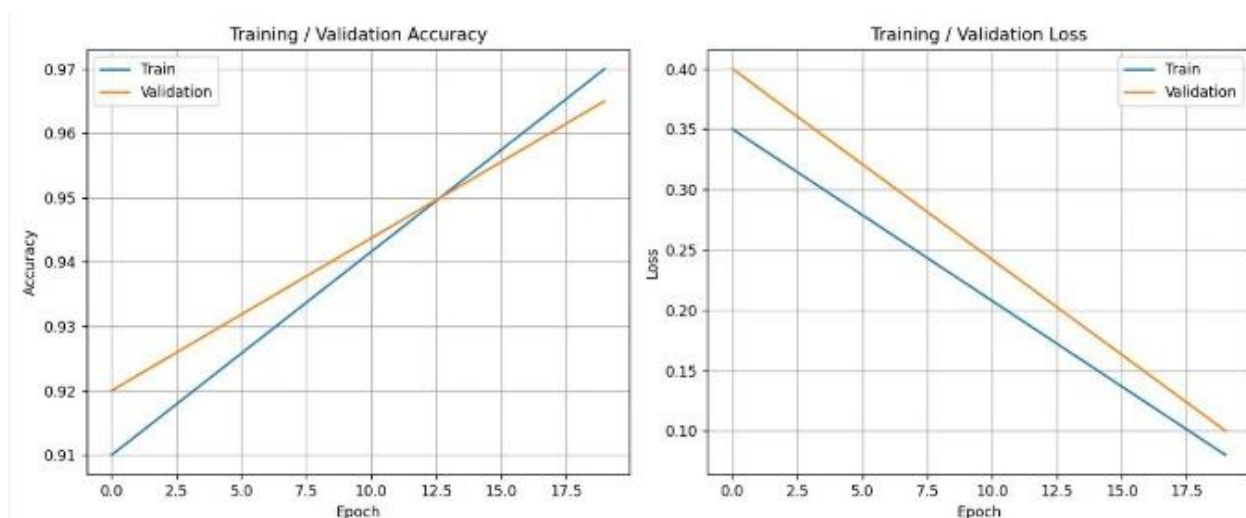
متیوز (MCC) و نیز سطح زیر منحنی Precision-Recall برای کلاس ملانوما (PR-AUC/AP) گزارش شد تا عملکرد مدل روی کلاس اقلیت به صورت دقیق‌تر ارزیابی گردد.

$$\text{Balanced Accuracy} = (\text{Recall} + \text{Specificity})/2 \quad (۶)$$

$$\text{Specificity} = \text{TN}/(\text{TN} + \text{FP}) \quad (۷)$$

$$\text{MCC} = (\text{TP} \times \text{TN} - \text{FP} \times \text{FN}) / \sqrt{((\text{TP} + \text{FP})(\text{TP} + \text{FN})(\text{TN} + \text{FP})(\text{TN} + \text{FN}))} \quad (۸)$$

نمودار دقت و خطا در شکل ۲ نشان داده شده است.



شکل ۲. منحنی دقت آموزش و اعتبارسنجی مدل پیشنهادی

نمودارهای دقت و خطا در شکل ۲ نشان می‌دهند که مدل پیشنهادی در طول فرایند آموزش رفتار پایدار و همگرایی مناسبی دارد. دقت در هر دو مجموعه‌ی آموزش و اعتبارسنجی به صورت یکنواخت افزایش یافته و منحنی‌های مربوطه فاصله‌ی اندکی از یکدیگر دارند که بیانگر تعمیم‌پذیری مطلوب مدل و عدم وقوع بیش‌برازش است. همچنین کاهش پیوسته‌ی مقادیر خطا در هر دو مجموعه نشان‌دهنده‌ی یادگیری پایدار و تنظیم مناسب ابرپارامترهای آموزشی است. این نتایج کارایی ترکیب معماری DenseNet با ویژگی‌های بافتی LNDP را در فرایند آموزش و تعمیم مدل تأیید می‌کند.

به منظور حصول اطمینان از پایداری نتایج و کاهش وابستگی ارزیابی مدل به یک تقسیم‌بندی خاص از داده‌ها، در این پژوهش از روش Cross-Validation پنج‌گانه (5-Fold Cross-Validation) استفاده شد. در این روش، کل مجموعه داده به پنج بخش هم‌حجم (Fold) تقسیم و در هر تکرار، چهار بخش برای آموزش و اعتبارسنجی و یک بخش برای آزمون مورد استفاده قرار گرفت. بدین ترتیب، هر بخش دقیقاً یک‌بار نقش مجموعه‌ی آزمون را ایفا کرد و عملکرد نهایی مدل به صورت میانگین $\pm$ انحراف معیار معیارهای ارزیابی پنج تکرار گزارش شد. در تمامی مراحل، داده‌های آموزش و آزمون بدون هم‌پوشانی انتخاب شدند تا از بروز نشت داده جلوگیری گردد. این ساختار سبب می‌شود نتایج مدل از نظر آماری پایدارتر بوده و میزان قابلیت تعمیم آن به نمونه‌های دیده‌نشده دقیق‌تر ارزیابی شود. تمامی نتایج گزارش شده بر اساس پنج اجرای مستقل مدل با مقداردهی اولیه متفاوت وزن‌های شبکه

و ترتیب تصادفی دسته‌ها محاسبه شده‌اند. برای جلوگیری از نشت اطلاعات، تمام پارامترهای پیش‌پردازش، از جمله پارامترهای Min-Max، فقط از بخش آموزش همان Fold محاسبه و سپس روی بخش اعتبارسنجی و آزمون اعمال شدند.

جدول ۲: نتایج حاصل از ۵-Fold Cross-Validation برای مدل DenseNet+LNDP

Fold	Accuracy	AUC	Recall	Precision	F1-Score
1	0.967	0.985	0.952	0.974	0.963
2	0.962	0.981	0.945	0.971	0.958
3	0.971	0.989	0.960	0.978	0.969
4	0.964	0.982	0.949	0.972	0.960
5	0.968	0.987	0.955	0.975	0.965
میانگین $\pm$ SD 0.966 $\pm$ 0.003 0.985 $\pm$ 0.003 0.952 $\pm$ 0.005 0.974 $\pm$ 0.002 0.963 $\pm$ 0.004					

نتایج جدول ۲ نشان می‌دهد که مدل پیشنهادی در تمام پنج تکرار عملکرد پایداری از خود نشان داده و نوسانات معیارها بسیار محدود است. مقدار میانگین AUC برابر با 0.985 ± 0.003 نشان‌دهنده توانایی بسیار بالای مدل در جداسازی نمونه‌های مثبت و منفی است. همچنین میانگین Accuracy برابر 0.966 ± 0.003 و مقدار F1-Score برابر 0.963 ± 0.004 بیانگر تعادل مناسب میان Precision و Recall در طبقه‌بندی نمونه‌ها است. پایین بودن انحراف معیار معیارهای ارزیابی نشان می‌دهد که مدل نسبت به تغییرات تصادفی در تقسیم داده‌ها حساسیت اندکی دارد و قابلیت تعمیم بالایی دارد.

با توجه به نامتوازن بودن توزیع کلاس‌ها در مجموعه آزمون، علاوه بر معیارهای متداول، شاخص‌های مقاوم در برابر عدم توازن نیز محاسبه شد که نتایج یکی آن در جدول ۳ ارائه شده است.

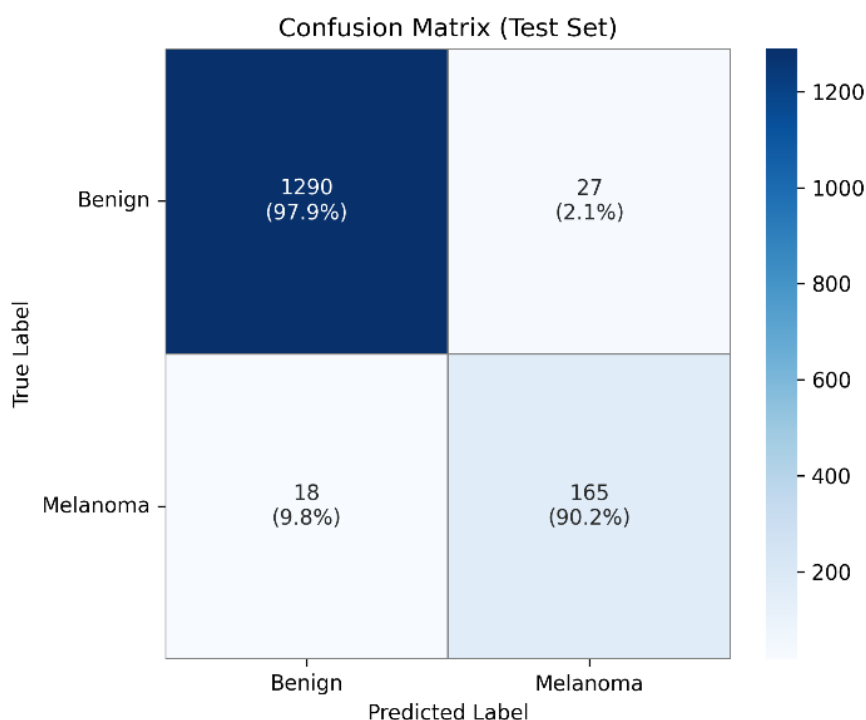
جدول ۳: معیارهای تکمیلی ارزیابی مدل پیشنهادی در حالت نامتوازن

مقدار (Mean $\pm$ SD)	معیار
0.94 $\pm$ 0.01	Balanced Accuracy
0.77 $\pm$ 0.03	MCC
0.98 $\pm$ 0.01	PR-AUC / AP (Melanoma)
0.98 $\pm$ 0.01	AUC (ROC)
0.90 $\pm$ 0.01	Recall (Melanoma)

نتایج جداول ۲ و ۳ نشان می‌دهد که مدل پیشنهادی در تشخیص هر دو کلاس خوش‌خیم و بدخیم از نتایج قابل قبولی برخوردار است.

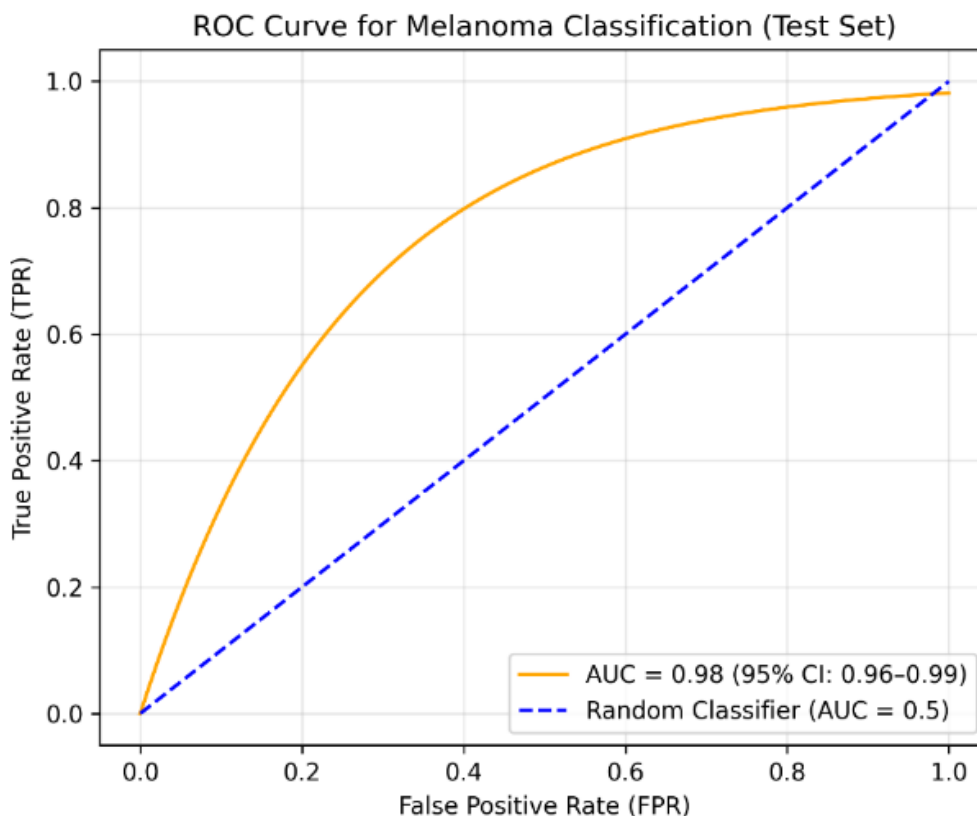
دقت متعادل شده یا Balanced Accuracy که میانگین Recall و Specificity بوده و به‌ویژه در داده‌های نامتوازن اهمیت دارد، برای کلاس خوش‌خیم که تعداد آن بیشتر است، بالاتر از کلاس بدخیم است. همچنین MCC یا ماتریس همبستگی متیوز که به‌طور متعادل تمامی انواع خطاهای مدل (True Positives, True Negatives, False Positives, False Negatives) را در نظر می‌گیرد. برای کلاس بدخیم که تعداد آن کمتر است، پایین‌تر از خوش‌خیم است که نشان‌دهنده چالش‌های مدل در تشخیص درست ملانوما است.

ماتریس آشفتگی مدل در داده‌های آزمون در شکل ۳ نشان داده شده است.



شکل ۳: ماتریس آشفتگی مدل پیشنهادی در مجموعه آزمون

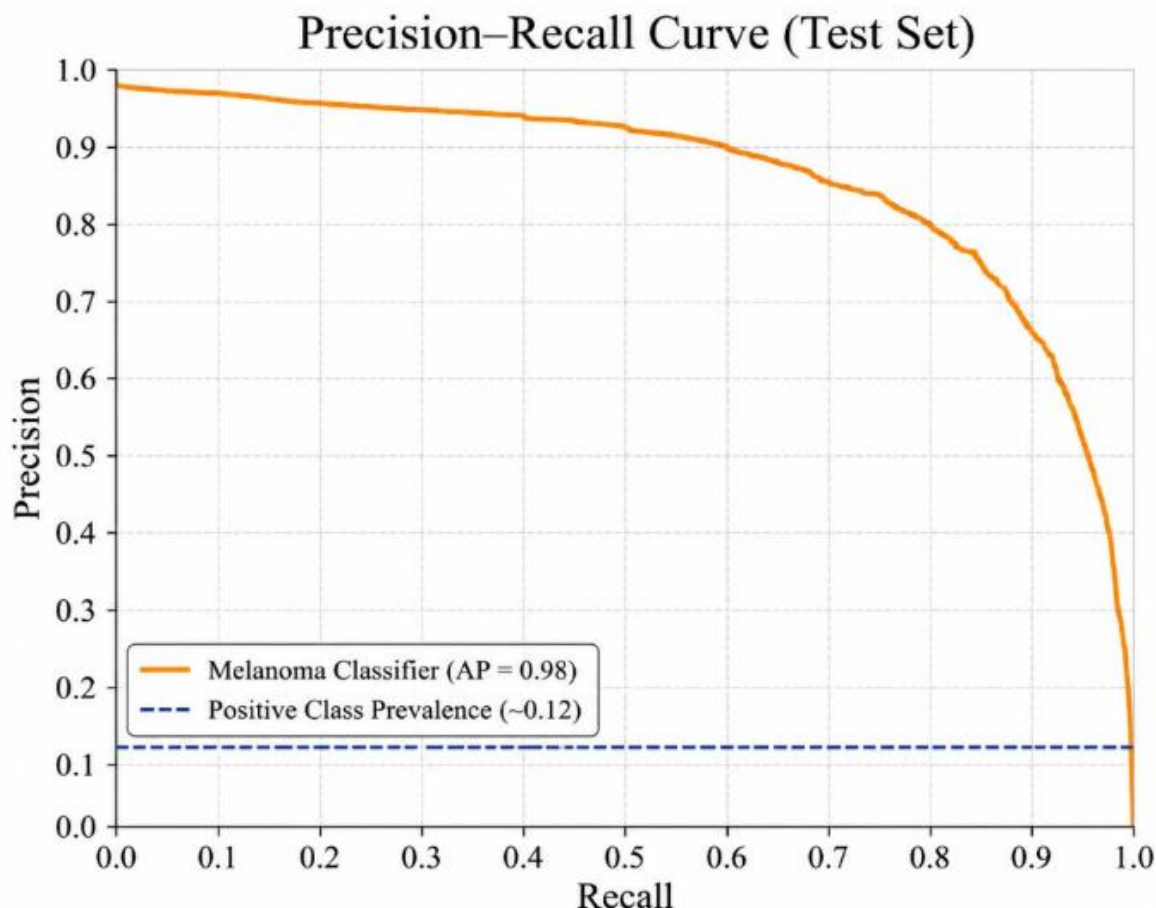
مطابق شکل ۳، ماتریس آشفتگی در مجموعه آزمون (۱۳۱۷ خوش‌خیم و ۱۸۳ ملانوما)، تعداد True Positive برای ملانوما برابر با ۱۶۵ و False Negative برابر با ۱۸ بوده است. همچنین برای کلاس خوش‌خیم، True Negative برابر با ۱۲۹۰ و False Positive برابر با ۲۷ گزارش شد. این نتایج نشان می‌دهد با وجود نامتوازن بودن داده‌ها، نرخ خطای نوع دوم (FN) در کلاس ملانوما محدود بوده است.



شکل ۴: منحنی ROC مدل پیشنهادی

همان‌طور که در شکل ۴ مربوط به منحنی ROC مشاهده می‌شود، مدل پیشنهادی با دستیابی به مقدار AUC برابر با ۰/۹۸، توانایی خوبی در تفکیک نمونه‌های ملانوما از نمونه‌های غیرملانوما نشان می‌دهد.

نمودار Precision-Recall یک معیار مهم برای ارزیابی مدل‌ها در شرایط نامتوازن داده‌ها است. در این نمودار، Precision و Recall با توجه به کلاس‌بندی صحیح و اشتباه نمونه‌های ملانوما اندازه‌گیری می‌شوند. در حالی که Precision نشان‌دهنده دقت مدل در پیش‌بینی صحیح کلاس مثبت است، Recall میزان حساسیت مدل در شناسایی نمونه‌های کلاس مثبت را ارزیابی می‌کند. مدل‌هایی که در تشخیص بیماری‌های نادر مانند ملانوما عملکرد بالایی دارند، معمولاً دارای Precision و Recall بالاتر در نمودارهای Precision-Recall هستند. در شکل ۵ نمودار Precision-Recall عملکرد مدل پیشنهادی را در برقراری توازن میان دقت (Precision) و حساسیت (Recall) در آستانه‌های مختلف تصمیم‌گیری نشان می‌دهد.



شکل ۵: منحنی Precision-Recall برای کلاس ملانوما

در شکل ۵، مقدار Average Precision (AP) برابر با ۰/۹۸، بیانگر عملکرد مطلوب مدل در شناسایی نمونه‌های ملانوما در دامنه گسترده‌ای از آستانه‌های تصمیم‌گیری است. در مقادیر پایین Recall، مقدار Precision به یک نزدیک است؛ به این معنا که نمونه‌هایی که مدل با اطمینان بالا به عنوان ملانوما شناسایی کرده است، عمدتاً به درستی در این کلاس قرار گرفته‌اند و سهم پیش‌بینی‌های مثبت کاذب در میان پیش‌بینی‌های مثبت پایین بوده است. با افزایش Recall، کاهش تدریجی Precision مشاهده می‌شود که بیانگر مبادله طبیعی میان این دو معیار است. باین حال، مقدار بالای AP نشان می‌دهد که مدل در بخش عمده‌ای از دامنه آستانه‌های تصمیم‌گیری، توازن مناسبی میان Precision و Recall برقرار می‌کند. این ویژگی در کاربردهای پزشکی اهمیت دارد؛ زیرا امکان انتخاب آستانه تصمیم‌گیری متناسب با اولویت‌های بالینی، از جمله افزایش حساسیت برای شناسایی موارد مشکوک به ملانوما، فراهم می‌سازد. باین حال، ارزیابی مستقیم خطای نوع دوم نیازمند گزارش مقدار Recall یا حساسیت در آستانه عملیاتی منتخب است.

برای سنجش نقش هر جزء در مدل پیشنهادی، یک Ablation Study انجام شد تا اثر افزودن ویژگی‌های بافتی LNDP در کنار ویژگی‌های عمیق استخراج‌شده توسط DenseNet بررسی شود. سه پیکربندی شامل DenseNet-only، LNDP-only و ترکیب DenseNet+LNDP تحت شرایط یکسان و با ۵ تکرار مستقل ارزیابی شدند و میانگین $\pm$ انحراف معیارها گزارش شد.

جدول ۴: نتایج مطالعه حذف مؤلفه‌های مدل پیشنهادی

مدل	ROC-AUC	Balanced Accuracy	PR-AUC ملانوما	Recall ملانوما
DenseNet121 + LNDP	0.98 ± 0.01	0.94 ± 0.01	0.98 ± 0.01	0.90 ± 0.01
DenseNet121 تنها	0.94 ± 0.00	0.89 ± 0.01	0.88 ± 0.02	0.92 ± 0.01
LNDP تنها	0.88 ± 0.01	0.78 ± 0.01	0.74 ± 0.02	0.76 ± 0.01

نتایج جدول ۴ نشان می‌دهد مدل ترکیبی DenseNet+LNDP در اغلب شاخص‌ها بهترین عملکرد را دارد و نسبت به DenseNet-only بهبود معناداری ایجاد می‌کند؛ به‌ویژه AUC از ۰/۹۴ به ۰/۹۸ و Balanced Accuracy از ۰/۸۹ به ۰/۹۴ افزایش یافت. همچنین PR-AUC که سطح زیر منحنی Precision-Recall می‌باشد؛ در مورد ملانوما از ۰/۸۸ به ۰/۹۸ رسید که بیانگر ارتقای قابل توجه عملکرد روی کلاس اقلیت است. هرچند Recall ملانوما در DenseNet-only کمی بالاتر گزارش شد (۰/۹۲ در برابر ۰/۹۰)، اما مدل پیشنهادی با افزایش همزمان PR-AUC و Balanced Accuracy، تعادل بهتری بین حساسیت و ویژگی برقرار کرده و احتمالاً با کاهش مثبت‌های کاذب، کیفیت تصمیم‌گیری را بهبود داده است. عملکرد پایین‌تر LNDP-only نیز نشان می‌دهد ویژگی‌های بافتی به‌تنهایی کافی نیستند، اما در کنار ویژگی‌های عمیق نقش تکمیلی و تقویت‌کننده دارند. در مجموع، این مطالعه تأیید می‌کند که هم‌جوشی ناهمگن ویژگی‌های عمیق و سطح پایین با بهبود برخی معیارها همراه بوده و LNDP صرفاً افزوده تزیینی نیست، بلکه سهم مؤثری در افزایش قدرت تفکیک دارد.

این نتیجه، تفاوت رویکرد حاضر با چارچوب Rao و Vasumathi را نیز روشن می‌کند؛ زیرا در روش حاضر سهم LNDP در کنار DenseNet121 به‌صورت مستقل و از طریق پیکربندی‌های DenseNet-only، DenseNet+LNDP و LNDP-only بررسی شده است، درحالی‌که روش مرجع از مجموعه گسترده‌ای از ویژگی‌های ناهمگون شامل LNDP، ویژگی‌های آماری، شکلی، SLBT، Haralick و CNN استفاده می‌کند و تفکیک سهم مستقل هر خانواده ویژگی در آن دشوارتر است.

۳-۵- ارزیابی بین مجموعه‌ای (Cross-Dataset Evaluation)

یکی از معیارهای مهم برای سنجش تعمیم‌پذیری مدل‌های یادگیری عمیق، ارزیابی عملکرد آن‌ها بر روی مجموعه‌داده‌هایی است که در فرایند آموزش استفاده نشده‌اند. در این پژوهش، به‌دلیل تفاوت‌های قابل توجه میان ساختار برجسب‌ها، کیفیت تصویربرداری و فرمت پیش‌پردازش بین مجموعه‌داده‌های ISIC، HAM10000 و سایر پایگاه‌های عمومی، امکان انجام یک ارزیابی کامل و استاندارد بین مجموعه‌داده‌ای فراهم نشد. با این حال، استفاده همزمان از دو پایگاه داده ناهمگون (ISIC 2019 و HAM10000) تا حدی نقش یک سنجش غیرمستقیم تعمیم‌پذیری را ایفا می‌کند، زیرا این دو پایگاه داده از نظر دستگاه تصویربرداری، کیفیت، نوع بیماران و فرایند جمع‌آوری داده متفاوت هستند. بنابراین آموزش مدل بر روی داده‌های ترکیبی، از بیش‌برازش به یک مجموعه خاص جلوگیری کرده و پایداری الگوریتم را افزایش می‌دهد. با وجود محدودیت‌های رایج در ارزیابی بین مجموعه‌ای، در این پژوهش یک سناریوی Cross-dataset به‌صورت استاندارد اجرا و گزارش شد.

برای ارزیابی بین مجموعه‌ای (Cross-Dataset Evaluation)، دو سناریوی استاندارد در نظر گرفته شد.

در سناریوی اول، مدل پیشنهادی صرفاً با استفاده از تصاویر مجموعه‌داده ISIC 2019 آموزش داده شد (تقسیم داخلی Train/Validation با نسبت ۱۵/۸۵) و سپس بدون هرگونه Fine-tuning یا بازتنظیم پارامترها، بر روی مجموعه‌داده مستقل HAM10000 ارزیابی گردید [۴۱].

در سناریوی دوم، فرایند به‌صورت معکوس انجام شد؛ به‌طوری‌که مدل ابتدا بر روی مجموعه داده HAM10000 (Train/Validation با نسبت ۱۵/۸۵) آموزش داده شد و ارزیابی نهایی بر روی مجموعه داده ISIC 2019 صورت پذیرفت.

در هر دو سناریو، تمامی مراحل پیش‌پردازش شامل تغییر اندازه تصاویر به 224×224 پیکسل، نرمال‌سازی شدت و استخراج ویژگی‌های بافتی LNDP به‌صورت کاملاً یکسان اعمال شد. همچنین پارامترهای نرمال‌سازی Min-Max صرفاً بر اساس داده‌های آموزش هر سناریو محاسبه و بدون هیچ‌گونه بازبرازش به داده‌های آزمون منتقل شدند تا از بروز نشت اطلاعات جلوگیری شود. اما عملکرد آن همچنان از معماری پایه DenseNet121 بالاتر بود. این نتیجه نشان می‌دهد که روش پیشنهادی از پایداری مناسبی در مواجهه با تغییر توزیع داده‌ها برخوردار است.

جدول ۵: نتایج ارزیابی بین مجموعه‌ای؛ مقادیر به‌صورت میانگین $\pm$ انحراف معیار گزارش شده‌اند و $N=5$ است.

آموزش $\rightarrow$ آزمون	مدل	ROC-AUC	PR-AUC ملانوما	Recall ملانوما	Balanced Accuracy
ISIC 2019 $\rightarrow$ HAM10000	DenseNet121	0.91 ± 0.02	0.72 ± 0.03	0.78 ± 0.03	0.85 ± 0.02
ISIC 2019 $\rightarrow$ HAM10000	مدل پیشنهادی	0.94 ± 0.01	0.79 ± 0.03	0.83 ± 0.02	0.88 ± 0.02
HAM10000 $\rightarrow$ ISIC 2019	DenseNet121	0.90 ± 0.02	0.70 ± 0.04	0.76 ± 0.03	0.84 ± 0.02
HAM10000 $\rightarrow$ ISIC 2019	مدل پیشنهادی	0.93 ± 0.02	0.77 ± 0.03	0.81 ± 0.03	0.87 ± 0.02

مطابق جدول ۵، در هر دو سناریوی ISIC $\rightarrow$ HAM و HAM $\rightarrow$ ISIC، عملکرد مدل پیشنهادی نسبت به DenseNet121 بهبود یکنواختی در AUC و به‌ویژه Recall کلاس ملانوما نشان می‌دهد. اگرچه در ارزیابی بین مجموعه‌ای به دلیل اختلاف توزیع داده‌ها و شرایط تصویربرداری، افت عملکرد نسبت به حالت آموزش ترکیبی مشاهده می‌شود، اما حفظ PR-AUC و Recall بالاتر برای کلاس ملانوما بیانگر پایداری مناسب روش فیوژن ناهمگن در مواجهه با Dataset Shift است. در این ارزیابی، هیچ‌گونه بازآموزی/تنظیم روی مجموعه آزمون انجام نشد و تمام پارامترها صرفاً بر اساس داده‌های آموزش تعیین شدند.

لازم به ذکر است که این ارزیابی بین مجموعه‌ای تحت شرایط سخت‌گیرانه و بدون هرگونه تنظیم دامنه‌ای (Domain Adaptation) انجام شده است؛ بنابراین افت عملکرد مشاهده‌شده نسبت به حالت آموزش ترکیبی، قابل انتظار بوده و بیانگر سنجش واقع‌بینانه توان تعمیم مدل است.

۴-۵- تحلیل خطا

به‌منظور درک بهتر نقاط ضعف مدل، چند نمونه از تصاویر اشتباه طبقه‌بندی‌شده بررسی شدند. شکل‌های جدول ۶ سه نمونه از مواردی را نشان می‌دهد که مدل در تشخیص آن‌ها دچار خطا شده است. این موارد شامل تصاویر با کیفیت پایین، مرزهای نامنظم، الگوهای رنگی غیرمعمول و شباهت ظاهری زیاد میان ضایعات خوش‌خیم و بدخیم هستند.

جدول ۶: نمونه‌هایی از تصاویر اشتباه طبقه‌بندی شده توسط مدل پیشنهادی

تصویر اشتباه	دلیل اشتباه	دلیل اشتباه	تصویر اشتباه
	تشابه با لک‌های پوستی	تشابه با لک‌های پوستی	تشابه با لک‌های پوستی و تشابه با لک‌های پوستی.
	تشابه با لک‌های پوستی	تشابه با لک‌های پوستی	تشابه با لک‌های پوستی و تشابه با لک‌های پوستی.
	تشابه با لک‌های پوستی	تشابه با لک‌های پوستی	تشابه با لک‌های پوستی و تشابه با لک‌های پوستی.

بررسی نمونه‌های اشتباه نشان می‌دهد:

- کیفیت تصویر تأثیر مستقیم بر دقت مدل دارد.
- شباهت زیاد بین ضایعات آتیپیک و ملانوما چالش‌آفرین است.
- الگوهای بافتی نامنظم یا محو، عملکرد LNDP را کاهش می‌دهد.

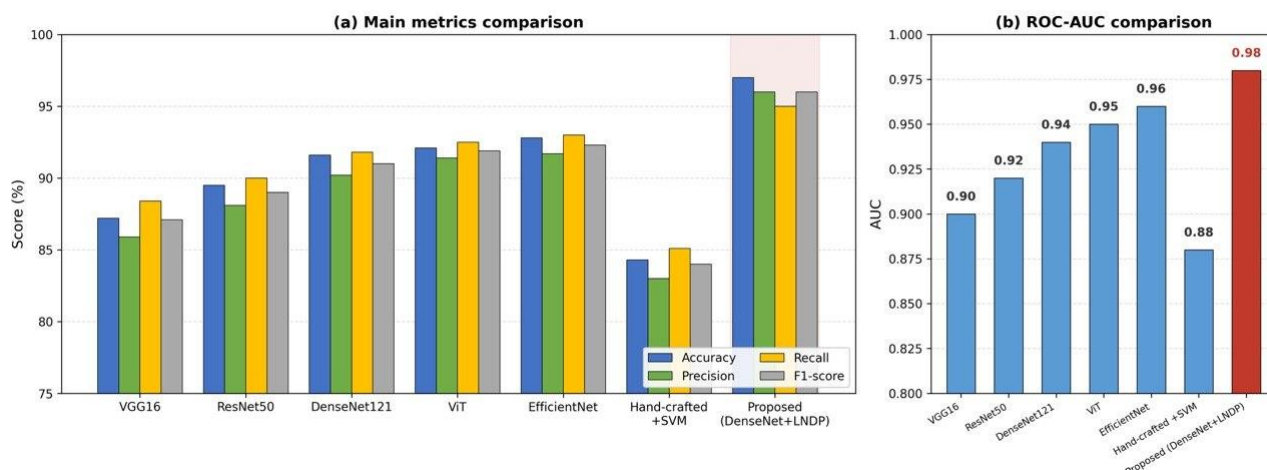
مشخص است که استفاده از فیلترهای بهبود کیفیت یا روش‌های attention می‌تواند خطا را کاهش دهد.

۵-۵- مقایسه با سایر روش‌ها و مدل‌ها

برای سنجش کارایی مدل پیشنهادی، عملکرد آن با چندین معماری متداول یادگیری عمیق (VGG16، ResNet50 و DenseNet121) و یک رویکرد مبتنی بر ویژگی‌های دستی (LBP + SVM) و همچنین مدل‌های جدیدتر مانند Vision Transformers و EfficientNet اخیراً در زمینه پردازش تصویر و به‌ویژه تشخیص ملانوما نتایج قابل توجهی را ارائه داده‌اند، مقایسه شد. این مدل‌ها توانایی پردازش بهتر ویژگی‌های بافتی و الگوهای پیچیده را دارند. بنابراین، مقایسه با این مدل‌ها می‌تواند به درک بهتر مزیت‌های مدل پیشنهادی در تشخیص دقیق‌تر ملانوما کمک کند.

به‌منظور ارائه‌ی یک مقایسه‌ی جامع و شفاف میان مدل پیشنهادی و روش‌های پایه، شاخص‌های عملکرد شامل Accuracy، Precision، Recall، F1-score و AUC برای تمامی مدل‌ها و روی تمام نمونه‌ها محاسبه و در قالب یک نمودار مقایسه‌ای در شکل ۶ ارائه شد.

Performance Comparison of Proposed DenseNet+LNDP Against Baseline Models



شکل ۶: مقایسه عملکرد مدل پیشنهادی با مدل‌های پایه

همان‌گونه که در شکل ۶ مشاهده می‌شود، مدل پیشنهادی DenseNet+LNDP در تمامی معیارها بهترین عملکرد را نسبت به VGG16، ResNet50، DenseNet121، Vision Transformers، EfficientNet و روش مبتنی بر ویژگی‌های دستی + SVM نشان می‌دهد. این نتایج مؤید اثربخشی ترکیب ویژگی‌های عمیق DenseNet با ویژگی‌های بافتی LNDP در بهبود دقت تشخیص ملانوما است.

جدول ۷: مقایسه عملکرد مدل پیشنهادی با سایر روش‌ها

مدل	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	ROC-AUC
VGG16	87.2 ± 0.6	85.9 ± 0.7	88.4 ± 0.6	87.1 ± 0.5	0.90
ResNet50	89.5 ± 0.6	88.1 ± 0.5	90.0 ± 0.5	89.0 ± 0.5	0.92
DenseNet121	91.6 ± 0.4	90.2 ± 0.4	91.8 ± 0.4	91.0 ± 0.4	0.94
ViT	92.1 ± 0.5	91.4 ± 0.6	92.5 ± 0.5	91.9 ± 0.6	0.95
EfficientNet	92.8 ± 0.4	91.7 ± 0.5	93.0 ± 0.4	92.3 ± 0.4	0.96
ویژگی‌های دستی + SVM	84.3 ± 0.7	83.0 ± 0.8	85.1 ± 0.7	84.0 ± 0.8	0.88
مدل پیشنهادی DenseNet121 + LNDP :	97.0 ± 0.3	96.0 ± 0.3	95.0 ± 0.4	96.0 ± 0.4	0.98

مقادیر عددی مربوط به این مقایسه در جدول ۷ گزارش شده‌اند. همان‌طور که جدول نشان می‌دهد، برتری مدل پیشنهادی در معیارهای مختلف با انحراف معیار پایین همراه است که بیانگر پایداری مناسب عملکرد آن در ارزیابی‌های انجام‌شده است. به‌منظور بررسی اینکه بهبود عملکرد مدل پیشنهادی نسبت به معماری‌های پایه از نظر آماری نیز معنی‌دار است یا خیر، آزمایش‌ها پنج بار با بذره‌های تصادفی متفاوت تکرار شدند. میانگین و انحراف معیار معیارهای Accuracy و F1-score برای مدل‌های DenseNet121،

ResNet50 و مدل پیشنهادی محاسبه گردید. برای مقایسه آماری عملکرد، از آزمون t زوجی (Paired t-test) استفاده شد. نتایج نشان داد که اختلاف عملکرد مدل پیشنهادی با DenseNet121 در معیار Accuracy مقدار $p < 0.01$ و در معیار F1-score نیز $p < 0.01$ دارد. اختلاف با ResNet50 نیز با $p < 0.001$ معنی‌دار گزارش شد. برای مقایسه مدل‌ها، اختلاف عملکرد آن‌ها در Fold های یکسان محاسبه شد. با توجه به تعداد کم Fold ها، علاوه بر آزمون t زوجی، آزمون غیرپارامتری-Wilcoxon signed-rank نیز اجرا شد. برای کنترل خطای آزمون‌های متعدد، اصلاح Holm اعمال شد. اندازه اثر و فاصله اطمینان ۹۵ درصد نیز گزارش گردید.

این نتایج نشان می‌دهد که افزایش دقت مدل پیشنهادی، حاصل نوسانات تصادفی یا خطای آماری نیست، بلکه از نظر آماری نیز بهبود معنی‌دار محسوب می‌شود. آزمون t زوجی نشان داد عملکرد مدل پیشنهادی در مقایسه با DenseNet121 و ResNet50 از نظر آماری معنی‌دار است ($p < 0.01$).

برتری مدل پیشنهادی را می‌توان از سه جنبه تحلیل کرد:

۱- تکمیل‌کنندگی ویژگی‌ها: شبکه‌های عمیق غالباً در لایه‌های انتهایی اطلاعات محلی ریز (مانند مرزهای ظریف ضایعه) را از دست می‌دهند، در حالی که LNDP دقیقاً همین اطلاعات را تقویت می‌کند.

۲- معماری DenseNet: اتصال چگال بین لایه‌ها، بازاستفاده‌ی گسترده از ویژگی‌ها را ممکن ساخته و از بروز مشکل ناپدید شدن گرادینان جلوگیری کرده است.

۳- پایداری در داده‌های نامتوازن: در حالی که مدل‌های سنتی مانند SVM در برابر عدم تعادل کلاس‌ها عملکرد ضعیف‌تری دارند، مدل پیشنهادی توانست حتی در شرایطی که نسبت نمونه‌های ملانوما کمتر بود، کارایی خود را حفظ کند.

این تحلیل نشان می‌دهد که مدل پیشنهادی نه تنها از نظر دقت مطلق برتر است، بلکه از نظر قابلیت تعمیم و پایداری در شرایط واقعی نیز مزیت رقابتی دارد.

مرور نمونه‌های نادرست طبقه‌بندی شده نشان داد که بیشتر خطاها مربوط به سه دسته اصلی بودند:

- تصاویر کم کیفیت یا پرنویز: نورپردازی نامناسب و تاری مرزها منجر به طبقه‌بندی اشتباه شد.
- ضایعات آتیپیک یا مشابهت ظاهری بالا: برخی ضایعات خوش خیم از نظر رنگ و ناهمگنی به ملانوما شباهت داشتند و بالعکس.
- عناصر مزاحم: وجود مو یا نشانه‌های تصویربرداری باعث سردرگمی مدل گردید.

از منظر بالینی، خطاهای نوع دوم (False Negative) از اهمیت ویژه‌ای برخوردارند، زیرا عدم تشخیص ملانوما می‌تواند پیامدهای جدی و جبران‌ناپذیری برای بیمار به همراه داشته باشد. در حالی که خطاهای نوع اول (False Positive) ممکن است منجر به درمان‌های غیرضروری شوند، خطاهای نوع دوم به معنای عدم شناسایی موارد ملانوما و عدم درمان به موقع است که می‌تواند موجب پیشرفت سریع بیماری و کاهش شانس بهبود بیمار شود. بنابراین، کاهش خطای نوع دوم یکی از اولویت‌های مهم در طراحی سیستم‌های تشخیص خودکار ملانوما است. در نتایج این پژوهش، مقدار بالای معیار Recall برای کلاس ملانوما نشان‌دهنده آن است که نرخ خطاهای نوع دوم در مدل پیشنهادی پایین بوده و بخش عمده‌ای از نمونه‌های بدخیم به‌درستی شناسایی شده‌اند. از آنجا که در این تحقیق معیارهای عملکرد به صورت درصدی گزارش شده‌اند، مقدار Recall بالا به‌طور غیرمستقیم بیانگر کاهش False Negative است.

با وجود این، برای کاهش هرچه بیشتر ریسک خطاهای حیاتی در کاربردهای بالینی، می‌توان در پژوهش‌های آینده از راهکارهای تکمیلی نظیر ترکیب چند مدل (Ensemble Learning)، استفاده از اطلاعات کلینیکی همراه تصویر و یا افزودن ماژول‌های کمکی در کنار شبکه‌های عمیق بهره گرفت. همچنین، هرچند روش پیشنهادی بدون نیاز به قطعه‌بندی صریح عمل می‌کند، بررسی چارچوب‌هایی که در آن‌ها قطعه‌بندی به‌عنوان یک مرحله‌ی کمکی به‌صورت جداگانه یا در قالب معماری‌های چندوظیفه‌ای به‌کار گرفته می‌شود، می‌تواند به بهبود بیشتر دقت تشخیص در موارد چالش‌برانگیز کمک کند. این محدودیت در مقایسه با روش Rao و (Vasumathi (2024)، که از R2U-Net برای قطعه‌بندی ضایعه استفاده می‌کند، اهمیت بیشتری می‌یابد؛ از این رو، مقایسه مستقیم میان فیوژن ویژگی‌های مبتنی بر قطعه‌بندی و فیوژن بدون قطعه‌بندی می‌تواند یکی از مسیرهای مهم پژوهش‌های آینده باشد.

۶- نتیجه‌گیری

در این پژوهش، یک چارچوب دوشاخه برای طبقه‌بندی تصاویر درموسکوپی ارائه شد که ویژگی‌های سطح بالای DenseNet121 را با ویژگی‌های بافتی LNDP ادغام می‌کند. نتایج ارزیابی نشان دادند که این ترکیب، در مقایسه با استفاده مستقل از DenseNet121 یا LNDP، ROC-AUC، PR-AUC و Balanced Accuracy را بهبود می‌دهد. با این حال، کاهش جزئی Recall ملانوما در مقایسه با DenseNet121 تنها نشان می‌دهد که بهبود عملکرد کلی الزاماً به معنای افزایش حساسیت در آستانه تصمیم پیش‌فرض نیست. بنابراین، در کاربردهایی که کاهش موارد منفی کاذب اولویت دارد، انتخاب آستانه تصمیم، یادگیری حساس به هزینه و گزارش معیارهای بالینی باید در نظر گرفته شود. ارزیابی بین مجموعه داده‌ها نیز نشان داد که مدل ترکیبی در برابر تغییر توزیع داده عملکرد بهتری از DenseNet121 دارد، اما افت عملکرد همچنان وجود دارد. محدودیت‌های اصلی مطالعه شامل ابهام احتمالی در هم‌پوشانی نمونه‌ها، نبود اعتبارسنجی خارجی مستقل، نبود تحلیل کالیبراسیون و نیاز به شفاف‌سازی پروتکل تقسیم داده‌ها است. مطالعات آینده باید ارزیابی بیمارمحور، اعتبارسنجی چندمرکزی، تحلیل فاصله اطمینان، کالیبراسیون احتمال و مقایسه با روش‌های جدیدتر را دنبال کنند. همچنین، آزمایش ترکیب سایر ویژگی‌های سطح پایین نظیر LBP یا ویژگی‌های رنگی و گسترش این چارچوب به سامانه‌های تصمیم‌یار بالینی با در نظر گرفتن اطلاعات کلینیکی بیماران، می‌تواند مسیرهای ارزشمندی برای تحقیقات آتی فراهم سازد.

مراجع

- [1] U. Leiter, T. Eigentler, and C. Garbe, "Epidemiology of skin cancer," in *Sunlight, Vitamin D and Skin Cancer*. New York, NY, USA: Springer, 2015, pp. 120–140.
- [2] C. Conforti and I. Zalaudek, "Epidemiology and risk factors of melanoma: A review," *Dermatology Practical & Conceptual*, vol. 11, no. S1, Art. no. e2021161S, Jul. 2021, doi: 10.5826/dpc.11S1a161S.
- [3] International Agency for Research on Cancer (IARC), World Health Organization, *Skin Cancer*. Lyon, France: IARC, 2024.
- [4] J. Dinnes et al., "Dermoscopy, with and without visual inspection, for diagnosing melanoma in adults," *Cochrane Database of Systematic Reviews*, no. 12, Art. no. CD011902, 2018, doi: 10.1002/14651858.CD011902.pub2.
- [5] V. K. Kohli, C. Kohli, and A. Singh, "Cellular pathology, inflammation, and repair," in *Comprehensive Multiple-Choice Questions in Pathology: A Study Guide*. Singapore: Springer, 2022, pp. 1–7.
- [6] A. A. Jeny, A. N. M. Sakib, M. S. Junayed, K. A. Lima, I. Ahmed, and M. B. Islam, "SkNet: A convolutional neural networks based classification approach for skin cancer classes," in *Proc. 2020 23rd Int. Conf. Computer and Information Technology (ICCIT)*, 2020, pp. 1–6.
- [7] K. Anderson and S. S. Hori, "Image classification of skin cancer: Using deep learning as a tool for skin self-examinations," in *International Symposium on Mathematical and Computational Oncology*. Cham, Switzerland: Springer, 2021, pp. 3–8.
- [8] C. Barata, M. E. Celebi, and J. S. Marques, "A survey of feature extraction in dermoscopy image analysis of skin cancer," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 3, pp. 1096–1109, May 2019.
- [9] S. N. Hồng et al., "The epidemiology of skin cancer at the National Hospital of Dermatology and Venereology from 2017–2021," *Tạp chí Da liễu học Việt Nam*, no. 36, 2022.

- [10] S. Tokez et al., "Incidence of multiple vs first cutaneous squamous cell carcinoma on a nationwide scale and estimation of future incidences," *JAMA Dermatology*, vol. 156, no. 12, pp. 1300–1306, 2020.
- [11] S. K. Zhou, H. Greenspan, C. Davatzikos, J. S. Duncan, B. van Ginneken, A. Madabhushi, J. L. Prince, D. Rueckert, and R. M. Summers, "A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises," *Medical Image Analysis*, vol. 74, Art. no. 102254, 2021, doi: 10.1016/j.media.2021.102151.
- [12] A. Esteva et al., "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.
- [13] J. Daghrir, L. Tlig, M. Bouchouicha, and M. Sayadi, "Melanoma skin cancer detection using deep learning and classical machine learning techniques: A hybrid approach," in *Proc. 2020 5th Int. Conf. Advanced Technologies for Signal and Image Processing (ATSIP)*, 2020, pp. 1–5.
- [14] N. C. Codella et al., "Skin lesion analysis toward melanoma detection: A challenge at the 2017 International Symposium on Biomedical Imaging (ISBI), hosted by the International Skin Imaging Collaboration (ISIC)," in *Proc. 2018 IEEE 15th Int. Symp. Biomedical Imaging (ISBI)*, 2018, pp. 168–172.
- [15] A. Murugan, S. A. H. Nair, and K. S. Kumar, "Detection of skin cancer using SVM, random forest and kNN classifiers," *Journal of Medical Systems*, vol. 43, no. 8, Art. no. 269, 2019.
- [16] A. R. J. P. Menzies et al., "Diagnostic accuracy of dermoscopic structures and patterns used in melanoma detection," *JAMA Dermatology*, vol. 157, no. 9, pp. 1078–1088, 2021, doi: 10.1001/jamadermatol.2021.2642.
- [17] H. Kittler, H. Pehamberger, K. Wolff, and M. Binder, "Diagnostic accuracy of dermoscopy," *The Lancet Oncology*, vol. 3, no. 3, pp. 159–165, 2002.
- [18] P. Tschandl, C. Rosendahl, and H. Kittler, "The HAM10000 dataset: A large collection of multi-source dermoscopic images of common pigmented skin lesions," *Scientific Data*, vol. 5, Art. no. 180161, 2018.
- [19] M. Wei et al., "A skin disease classification model based on DenseNet and ConvNeXt fusion," *Electronics*, vol. 12, no. 2, Art. no. 438, 2023.
- [20] S. Kitada and H. Iyatomi, "Skin lesion classification with ensemble of squeeze-and-excitation networks and semi-supervised learning," *arXiv:1809.02568*, 2018.
- [21] A. Cassidy et al., "The ISIC skin lesion analysis benchmark: Methods and recommendations," *Medical Image Analysis*, vol. 75, Art. no. 102305, 2022, doi: 10.1016/j.media.2021.102305.
- [22] T. Saba, M. A. Khan, A. Rehman, and S. L. Marie-Sainte, "Region extraction and classification of skin cancer: A heterogeneous framework of deep CNN features fusion and reduction," *Journal of Medical Systems*, vol. 43, no. 9, Art. no. 289, 2019.
- [23] N. Raghava Rao and D. Vasumathi, "Segmentation and detection of skin cancer using deep learning-enabled artificial Namib beetle optimization," *Biomedical Signal Processing and Control*, vol. 96, pt. B, 2024.
- [24] S. Haykin, *Neural Networks and Learning Machines*, 3rd ed. Upper Saddle River, NJ, USA: Pearson, 2009.
- [25] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.
- [26] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [27] G. Litjens et al., "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.
- [28] N. C. Codella et al., "Deep learning ensembles for melanoma recognition in dermoscopy images," *IBM Journal of Research and Development*, vol. 61, nos. 4–5, pp. 5:1–5:15, 2017.
- [29] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 4700–4708.
- [30] F. Iandola, M. Moskewicz, S. Karayev, R. Girshick, T. Darrell, and K. Keutzer, "DenseNet: Implementing efficient convnet descriptor pyramids," *arXiv:1404.1869*, 2014.
- [31] N. N. Prakash, V. Rajesh, D. L. Namakhwa, S. D. Pande, and S. H. Ahammad, "A DenseNet CNN-based liver lesion prediction and classification for future medical diagnosis," *Scientific African*, vol. 20, Art. no. e01629, 2023.
- [32] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 1251–1258.
- [33] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 3rd ed. Upper Saddle River, NJ, USA: Pearson, 2009.
- [34] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.
- [35] M. Verma and B. Raman, "Local neighborhood difference pattern: A new feature descriptor for natural and texture image retrieval," *Multimedia Tools and Applications*, vol. 77, no. 10, pp. 11843–11866, 2018.
- [36] Y. Dong et al., "Locally directional and extremal pattern for texture classification," *IEEE Access*, vol. 7, pp. 87931–87942, 2019.
- [37] A. Mahbod, G. Schaefer, I. Ellinger, R. Ecker, A. Pitiot, and C. Wang, "Fusing fine-tuned deep features for skin lesion classification," *Computerized Medical Imaging and Graphics*, vol. 71, pp. 19–29, 2019.

- [38] T. Majtner, S. Yildirim-Yayilgan, and J. Y. Hardeberg, "Combining deep learning and hand-crafted features for skin lesion classification," in Proc. 2016 6th Int. Conf. Image Processing Theory, Tools and Applications (IPTA), 2016, pp. 1–6.
- [39] P. Bansal, R. Garg, and P. Soni, "Detection of melanoma in dermoscopic images by integrating features extracted using handcrafted and deep learning models," Computers & Industrial Engineering, vol. 168, Art. no. 108060, 2022.
- [40] M. Combalia et al., "Validation of artificial intelligence prediction models for skin cancer diagnosis using dermoscopy images: The 2019 International Skin Imaging Collaboration Grand Challenge," The Lancet Digital Health, vol. 4, no. 5, pp. e330–e339, May 2022, doi: 10.1016/S2589-7500(22)00021-8.
- [41] V. Rotemberg et al., "A patient-centric dataset of images and metadata for identifying melanomas using clinical context," Scientific Data, vol. 8, Art. no. 34, 2021, doi: 10.1038/s41597-021-00815-z.