

# Intelligent Skin Cancer Detection Using a Multispectral Metasurface and a Hybrid Deep Learning Architecture Based on CNN–ViT

**Afsane Saeed Arezoomand\***

PhD in Communication Engineering, Young  
Researchers and Elite Club, Islamic Azad  
University, Urmia Branch, Urmia, Iran.

## Abstract

Skin cancer is among the most prevalent malignancies worldwide, and its early detection is essential for improving patient survival and reducing treatment costs. Conventional dermoscopic and visual imaging techniques are primarily limited to the visible spectrum and often fail to capture subtle spectral signatures associated with early-stage malignancies. This study proposes an innovative framework that integrates a multispectral metasurface for imaging with a hybrid deep learning architecture based on Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs). The designed metasurface enables noninvasive acquisition of rich spectral information highly sensitive to tissue alterations, while the hybrid CNN–ViT model simultaneously extracts local and global features to robustly classify skin lesions. Simulation-based evaluations demonstrate that the proposed method achieves approximately 98% accuracy, 95% sensitivity, and 99% specificity—surpassing conventional RGB-based and single-architecture approaches. Qualitative analyses using attention maps reveal that the model focuses on clinically relevant lesion regions, improving interpretability. Overall, the results indicate that combining metasurface-based multispectral imaging with hybrid deep learning can introduce a new generation of diagnostic tools in dermatology and pave the way for portable, fast, and highly accurate clinical systems.

**Keywords:** skin cancer, metasurface, multispectral imaging, deep learning, convolutional neural network (CNN), vision transformer (ViT), skin lesion classification

Received: 25/September/2025

Accepted: 12/July/2025

eISSN: 3060-6144

ISSN: 2980-8936

## تشخیص هوشمند سرطان پوست با استفاده از فراسطح چندطیفی و معماری هیبریدی یادگیری عمیق مبتنی بر CNN-ViT

دکتری تخصصی رشته برق، گرایش مخابرات، باشگاه پژوهشگران جوان و  
نخبگان، دانشگاه آزاد اسلامی، واحد ارومیه، ارومیه، ایران. \* افسانه ساعی آرزومند

### چکیده

سرطان پوست یکی از شایع‌ترین بدخیمی‌ها در سراسر جهان است که تشخیص زودهنگام آن نقش کلیدی در افزایش بقا و کاهش هزینه‌های درمانی دارد. روش‌های رایج تصویربرداری پوستی عمدتاً مبتنی بر تصاویر مرئی بوده و در بسیاری از موارد قادر به شناسایی ویژگی‌های پنهان طیفی در مراحل اولیه بیماری نیستند. لذا در پژوهش حاضر، یک چارچوب نوآورانه برای تشخیص خودکار سرطان پوست ارائه می‌شود که تلفیقی از یک فراسطح چندطیفی (متاسرفیس) جهت تصویربرداری و یک معماری یادگیری عمیق هیبریدی مبتنی بر شبکه‌ی عصبی پیچشی (CNN) و ترنسفورمر بینایی (ViT) است. فراسطح طراحی شده قادر است امضاهای طیفی حساس به تغییرات بافت را به صورت غیرتهاجمی و با دقت بالا ثبت کند، سپس داده‌های طیفی تولیدشده همراه با تصاویر مرئی، طی یک مدل هیبریدی پردازش شده و با استفاده از سازوکارهای توجه و استخراج ویژگی‌های محلی-جهانی، طبقه‌بندی ضایعات انجام می‌گیرد. نتایج ارزیابی شبیه‌سازی شده نشان می‌دهد که روش پیشنهادی توانسته است به دقت حدود ۹۸ درصد، حساسیت ۹۵ درصد و ویژگی ۹۹ درصد دست یابد که نسبت به مدل‌های مرجع مبتنی بر تصاویر RGB یا معماری‌های تک‌لایه برتری قابل ملاحظه‌ای دارد. تحلیل‌های کیفی نیز حاکی از تمرکز مدل بر نواحی کلینیکی مهم و قابلیت تفسیرپذیری بالا است. مجموعه یافته‌ها نشان می‌دهد که ترکیب فراسطح چندطیفی و یادگیری عمیق هیبریدی می‌تواند به عنوان نسل جدید ابزارهای تشخیصی در پوست‌شناسی مطرح شود و مبنایی را برای توسعه‌ی سیستم‌های قابل حمل، سریع و دقیق در کاربردهای بالینی آینده فراهم آورد.

**کلیدواژه‌ها:** سرطان پوست، فراسطح، تصویربرداری چندطیفی، یادگیری عمیق، شبکه عصبی پیچشی، ترنسفورمر بینایی، تشخیص ضایعات پوستی

## ۱- مقدمه

سرطان پوست شایع‌ترین بدخیمی در سطح جهان محسوب می‌شود که قرار گرفتن طولانی مدت در معرض پرتوهای فرابنفش خورشید مهم‌ترین عامل خطر آن است (Smith et al., 2023). سه نوع اصلی این سرطان شامل کارسینوم سلول بازال، کارسینوم سلول سنگفرشی و ملانوم می‌باشد که در میان آن‌ها، ملانوم به دلیل توانایی بالای متاستاز، خطرناک‌ترین نوع شناخته می‌شود (Anderson et al., 2022). تشخیص زودهنگام نقش حیاتی در افزایش بقا دارد، زیرا ضایعات ملانوم در مراحل اولیه اغلب فاقد نشانه‌های واضح بوده و تشخیص آن‌ها تنها از طریق بررسی دقیق ویژگی‌های ظاهری امکان‌پذیر است (Zhang et al., 2024).

روش‌های رایج تشخیص مانند معاینه‌ی چشمی متخصص پوست و درموسکوپ<sup>۱</sup> با وجود مزایایی نظیر تقویت بافت و الگوهای ضایعه، همچنان با محدودیت‌هایی روبه‌رو هستند؛ از جمله این محدودیت‌ها می‌توان به وابستگی شدید به تجربه‌ی متخصص، خطای میان‌مشاهده‌گری و شباهت بالای ضایعات خوش‌خیم و بدخیم اشاره نمود (Miller et al., 2021). علاوه بر این، کمبود متخصصان مجرب در بسیاری از مناطق موجب تأخیر در تشخیص و درمان می‌شود (Rahman et al., 2022). این شرایط، ضرورت توسعه‌ی سامانه‌های خودکار، دقیق و کم‌هزینه را پررنگ‌تر می‌سازد.

در سال‌های اخیر، پیشرفت‌های چشمگیر هوش مصنوعی<sup>۲</sup> و یادگیری عمیق<sup>۳</sup> موجب تحول در تحلیل تصاویر پزشکی شده است. مدل‌های شبکه عصبی پیچشی<sup>۴</sup> توانسته‌اند با استخراج ویژگی‌های سلسله‌مراتبی از تصاویر پوستی، دقت تشخیص را به سطح متخصصان پوست و حتی بالاتر از آن برسانند (Esteva et al., 2017; Li et al., 2023). با این حال، وابستگی CNNها به نواحی محلی تصویر، توانایی آن‌ها را در مدل‌سازی روابط بلندبرد محدود می‌کند؛ به جهت رفع این نقیصه، ترنسفورمر بینایی<sup>۵</sup> با سازوکار خودتوجهی سراسری<sup>۶</sup> معرفی شده است (Dosovitskiy et al., 2021).

ترکیب CNN و ViT در قالب معماری‌های هیبریدی، دستاوردهای مهمی در تشخیص ضایعات پوستی داشته است؛ به‌عنوان نمونه، ساختارهایی مانند مدل EViT-DenseNet توانسته‌اند در مجموعه داده ISIC به دقت‌هایی فراتر از ۹۵ درصد برسند (Khan et al., 2024). این شواهد علمی نشان می‌دهد که بهره‌گیری از معماری‌های هیبریدی، مسیر نوینی را برای ارتقای سامانه‌های بینایی پزشکی فراهم نموده است.

در حوزه‌های دیگر، فناوری متامتریال<sup>۷</sup> و فراسطح<sup>۸</sup> به دلیل قابلیت مهندسی ویژگی‌های الکترومغناطیسی در مقیاس زیرطول‌موج، امکان تولید حسگرهای بسیار حساس را فراهم ساخته است. فراسطوح می‌توانند جبهه موج نور را به صورت دقیق کنترل کرده و سیگنال‌های ظریف ناشی از تغییرات بافتی را ثبت کنند (Chen et al., 2022). پژوهش‌های اخیر نشان داده‌اند که ترکیب متامتریال‌ها با موادی مانند گرافن<sup>۹</sup> حساسیت حسگرهای ترانزیتور را به‌طور چشمگیری افزایش می‌دهد (Sarkar et al., 2023). در زمینه‌ی سرطان پوست نیز نمونه‌هایی از متاسنسورهای ترانزیتور طراحی شده‌اند که قادرند تفاوت بافت سالم و سرطانی را به‌صورت غیرتهاجمی تشخیص دهند (Zhang et al., 2023).

1. dermoscopy
2. artificial intelligence
3. deep learning
4. Convolutional neural network (CNN)
5. Vision transformer (ViT)
6. global self-attention
7. metamaterial
8. metasurface
9. graphene

تلاقی این دو حوزه؛ یعنی بینایی ماشین و فراسطح پیشرفته، می‌تواند نسل جدیدی از سیستم‌های تشخیصی پوستی ایجاد کند. به‌ویژه با ظهور مدل‌های یادگیری چندوجهی<sup>۱</sup> که امکان ترکیب داده‌های تصویری، طیفی و بالینی را فراهم می‌کنند، زمینه برای طراحی سامانه‌هایی با دقت بالا فراهم شده است (Alaei et al., 2024).

پژوهش حاضر با مطالعه‌ی این روند، یک رویکرد نوین معرفی می‌کند که در آن یک فراسطح سفارشی به‌عنوان حسگر تصویربرداری ابرطیفی عمل کرده و داده‌های غنی را در اختیار سامانه‌ی یادگیری عمیق قرار می‌دهد، سپس معماری هیبریدی عمیق، داده‌ها را تحلیل کرده و نوع ضایعه‌ی پوستی را تعیین می‌کند. در ادامه، پس از ارائه‌ی جزئیات روش، نتایج تجربی، ارزیابی عملکرد و مقایسه با کارهای پیشین، نوآوری‌ها و مسیرهای آینده بررسی می‌شوند.

## ۲- کارهای پیشین

پژوهش‌های مرتبط با موضوع حاضر را می‌توان در سه حوزه‌ی اصلی دسته‌بندی کرد:

(۱) کارهای مرتبط با تشخیص ضایعات پوستی مبتنی بر یادگیری عمیق،

(۲) پژوهش‌های مربوط به متامتریال و فراسطح در حسگری،

(۳) کارهای ترکیبی یا چندوجهی.

۱. تشخیص ضایعات پوستی با یادگیری عمیق

تحلیل خودکار تصاویر پوستی با بهره‌گیری از یادگیری عمیق، یکی از فعال‌ترین حوزه‌های تحقیقاتی سال‌های اخیر بوده است. استوا و همکاران<sup>۲</sup> (۲۰۱۷) اولین پژوهش اثرگذار خود را در این زمینه ارائه کرده‌اند. پژوهش آن‌ها نشان می‌دهد که یک CNN می‌تواند عملکردی مشابه متخصصین پوست داشته باشد. پس از آن، معماری‌هایی مانند ResNet، DenseNet، MobileNet و EfficientNet نتایج قابل توجهی در مجموعه داده‌های ISIC به دست آورده‌اند (Tschandl et al., 2020; Li et al., 2023).

با ظهور ترنسفورمرهای بینایی، مدل‌هایی مانند ViT و Swin Transformer وارد عرصه شدند و دقت تشخیص ضایعات پوستی را افزایش دادند (Wu et al., 2022). مدل‌های هیبریدی که ترکیب CNN و ViT را هدف قرار می‌دهند، مانند EViT-DenseNet و CoAtNet، توانسته‌اند عملکرد قابل توجهی را در تفکیک ملانوم از سایر ضایعات ارائه دهند (Khan et al., 2024).

۲. متامتریال، فراسطح و حسگرهای تراهتر

فراسطح به دلیل توانایی در دستکاری دقیق امواج الکترومغناطیسی، در سال‌های اخیر توجه گسترده‌ای در طراحی حسگرهای زیستی جلب کرده‌اند. چن و همکاران<sup>۳</sup> (۲۰۲۲) نشان دادند که متامتریال‌های مبتنی بر شکاف رزونانس می‌توانند نسبت به تغییرات کوچک ضریب شکست بسیار حساس باشند. ترکیب فراسطح با گرافن موجب افزایش چشمگیر میدان الکترومغناطیسی موضعی شده و حساسیت حسگر را چند برابر می‌کند (Sarkar et al., 2023).

در زمینه‌ی سرطان پوست، پژوهش‌هایی مانند طباطبائیان و همکاران<sup>۴</sup> (۲۰۲۲) و ژنگ و همکاران<sup>۵</sup> (۲۰۲۳) استفاده از فراسطح تراهتر را برای شناسایی غیرتهاجمی ضایعات پوستی پیشنهاد کرده‌اند. این مطالعات نشان می‌دهد که فناوری فراسطح می‌تواند داده‌های تصویری و طیفی غنی برای تحلیل‌های یادگیری عمیق فراهم کند.

۳. یادگیری چندوجهی و سیستم‌های ترکیبی

1. multimodal learning  
2. Esteva et al.  
3. Chen et al.  
4. Tabatabaieian et al.  
5. Zhang et al.

با گسترش رویکردهای چندوجهی، استفاده از داده‌های چندنوعی شامل تصویر، اسکن ابرطیفی، داده‌های بالینی و متن پزشکی، کارایی مدل‌ها را افزایش داده است. علائی و همکاران<sup>۱</sup> (۲۰۲۴) یک چارچوب چندوجهی برای تشخیص سرطان پستان ارائه کردند که ترکیب داده‌های تصویری و متنی را ممکن می‌ساخت. ایده‌های مشابه در تشخیص بیماری‌های پوستی نیز در حال توسعه‌اند و نشان داده‌اند که تلفیق سیگنال‌های اپتیکی و تصاویر می‌تواند دقت و قابلیت اطمینان مدل را افزایش دهند (Nguyen et al., 2023).

نتیجه‌ی کلی کارهای پیشین نشان می‌دهد که فراسطوح داده‌های غنی فراهم می‌کنند؛ یادگیری عمیق قابلیت تحلیل آن‌ها را دارد و رویکردهای چندوجهی می‌توانند این دو حوزه را به‌صورت مؤثر تلفیق کنند.

جدول ۱. مقایسه کارهای پیشین در سه حوزه اصلی پژوهش

ردیف	حوزه پژوهش	مرجع / نمونه مطالعات	نوع داده	روش / معماری	دستاوردهای کلیدی
۱		Esteva et al. (2017)	تصویر درموسکویی (RGB)	CNN کلاسیک	عملکرد مشابه متخصص پوست
۲	تشخیص ضایعات پوستی با یادگیری عمیق	Tschandl et al. (2020)	ISIC-HAM10000	ResNet / DenseNet	استخراج ویژگی قدرتمند با دقت بالا
۳		Wu et al. (2022)	RGB	Vision Transformer (ViT)	مدل‌سازی روابط بلندبرد، بهبود دقت
۴		Khan et al. (2024)	RGB	Hybrid CNN-ViT (EViT-DenseNet)	عملکرد برتر نسبت به مدل‌های تک‌معماری
۵		Chen et al. (2022)	سیگنال THz	ماتریال رزونانسی	حساسیت بالا به تغییر ضریب شکست
۶	فراسطوح و حسگرهای مبتنی بر تراهرتز	Sarkar et al. (2023)	سیگنال + THz گرافن	متاسنسور گرافنی	افزایش شدید میدان الکترومغناطیسی و حساسیت
۴		Tabatabaeian et al. (2022)	بازتاب THz	فراسطح تشخیصی سرطان پوست	تشخیص غیرتهاجمی اولیه
۸		Zhang et al. (2023)	سیگنال THz	THz Imaging	شناسایی غیرتماسی ملانوم
۹	یادگیری چندوجهی و	Alaei et al. (2024)	تصویر + متن بالینی	Multimodal Fusion	افزایش دقت تشخیص در سرطان پستان
۱۰	روش‌های ترکیبی	Nguyen et al. (2023)	تصویر + داده طیفی	Multimodal Deep Learning	بهبود پایداری تشخیص

این شکاف علمی دقیقاً همان جایی است که پژوهش حاضر در آن قرار دارد.

### ۳- روش پیشنهادی

روش پیشنهادی این پژوهش یک چارچوب تلفیقی متشکل از یک حسگر اپتیکی مبتنی بر فراسطح و یک معماری یادگیری عمیق هیبریدی است که با هدف تشخیص دقیق و غیرتهاجمی سرطان پوست طراحی شده است.

این سامانه شامل چهار مرحله اصلی است:

۱. دریافت داده نوری از ضایعه پوستی با استفاده از فراسطح؛

۲. پیش‌پردازش و (اختیاری) بخش‌بندی ناحیه ضایعه؛

۳. استخراج ویژگی چندمقیاسی با استفاده از CNN و ترنسفورمر بینایی (ViT)؛

۴. ادغام ویژگی‌ها و طبقه‌بندی نهایی.

در ادامه، هر بخش با جزئیات تشریح می‌شود.

### ۳-۱- حسگر اپتیکی مبتنی بر فراسطح

در این بخش توضیحات مربوط به حسگر اپتیکی مبتنی بر فراسطح توضیح داده می‌شود:

#### ۳-۱-۱- ساختار فراسطح و اصول عملکرد

فراسطح پیشنهادی یک آرایه‌ی دوبعدی از عناصر رزونانسی زیرطول‌موج است که رفتار نوری را در بسامدهای انتخابی، به‌ویژه باند تراهرتز، کنترل می‌کند. این ساختار از واحدهای مهندسی شده تشکیل شده است که قادر هستند فاز، دامنه و پلاریزاسیون موج فرودی را در پاسخ به تغییرات بسیار کوچک ضریب شکست بافت تغییر دهند (Chen et al., 2022). همچنین، ساختارهای مشابه فراسطح که قابلیت دستکاری امواج الکترومغناطیسی در باند تراهرتز را دارند، می‌توانند به‌عنوان جاذب‌های چندباند با عملکرد مستقل از پلاریزاسیون طراحی شوند (Saei Arezoomand et al., 2015).

مواد مورد استفاده در این ساختار عبارتند از:

- زیرلایه دی‌الکتریک با اتلاف پایین.

- لایه رسانای فوق‌نازک مانند گرافن جهت افزایش شدت میدان نزدیک.

- رزونانسرهای فلزی شکاف‌دار برای ایجاد حساسیت طیفی بالا.

ترکیب این عناصر امکان ایجاد یک پدیده رزونانس تیز با Q-factor بالا را فراهم می‌سازد که برای آشکارسازی تغییرات نوری بافت‌های سرطانی بسیار مناسب است (Sarkar et al., 2023). اصول طراحی جاذب‌های رزونانسی با فاکتور کیفیت<sup>۱</sup> بالا در باندهای نوری و مادون قرمز میانی، با هدف کاربردهای حسگری زیستی و افزایش شدت میدان نزدیک، پیش از این نیز مورد بررسی قرار گرفته‌اند (Jahangiri et al., 2017).

#### ۳-۱-۲- تولید امضای طیفی نقطه‌ای<sup>۲</sup>

هنگامی که فراسطح در مقابل ضایعه پوستی قرار می‌گیرد، میدان الکترومغناطیسی در نزدیکی سطح آن تغییر می‌کند و این تغییر در قالب یک امضای طیفی نقطه‌ای شامل چندین فرکانس رزونانسی ثبت می‌شود.

خروجی حسگر نه یک تصویر کامل، بلکه یک بردار ویژگی طیفی است:

$$S = [s(\lambda_1), s(\lambda_2), \dots, s(\lambda_K)] \quad (1)$$

که در آن  $K$  تعداد باندهای انتخاب شده و  $s(\lambda_k)$  شدت بازتابی یا عبوری در باند  $\lambda_k$  است.

این داده طیفی نسبت به ضایعات سرطانی و غیربدخیم تفاوت معناداری دارد (Zhang et al., 2023).

#### ۳-۲- پیش پردازش و بخش بندی

داده خام طیفی و تصویر RGB هم‌زمان ثبت می‌شوند که یک سری عملیات پیش‌پردازش، شامل موارد زیر، انجام می‌شود:

1. Q-factor  
2. spectral signature

۱. حذف نویز،
  ۲. هم‌ترازی شدت<sup>۱</sup>،
  ۳. افزایش کنتراست در باندهای حساس،
  ۴. نرمال‌سازی در بازه [۰، ۱].
- در صورت نیاز، یک مرحله بخش‌بندی خودکار نیز اعمال می‌گردد:

### ۳-۲-۱- بخش‌بندی با U-Net

برای استخراج ناحیه دقیق ضایعه، از مدل یونات (U-Net) یا نسخه‌های توسعه‌یافته‌ی آن مانند ++U-Net یا SegFormer (Segmentation Transformer) استفاده می‌شود. این گام اختیاری است، اما مطالعات اخیر نشان داده‌اند که بخش‌بندی، دقت طبقه‌بندی را افزایش می‌دهد (Khan et al., 2024).



شکل ۱. نتیجه بخش‌بندی U-Net روی یک ضایعه پوستی

### ۳-۳- معماری یادگیری عمیق ترکیبی

به‌منظور تحلیل هم‌زمان ویژگی‌های محلی<sup>۲</sup> و ویژگی‌های کل‌نگر<sup>۳</sup>، یک معماری هیبریدی شامل DenseNet به‌عنوان CNN و Vision Transformer (ViT) طراحی شده است.

### ۳-۳-۱- استخراج ویژگی محلی با DenseNet

شبکه DenseNet با مسیرهای میان‌بری متراکم، برای استخراج لبه‌ها، بافت‌ها و ساختارهای ریز ضایعه به‌کار می‌رود. این نسخه شامل موارد زیر می‌باشد:

- حذف لایه‌های انتهایی،
  - کاهش تعداد فیلترها جهت جلوگیری از بیش‌برازش،
  - استفاده از وزن‌های از پیش آموزش‌دیده ImageNet.
- خروجی این بخش یک تانسور سه‌بعدی از ویژگی‌های محلی است:

$$F_{local} \in R^{h \times w \times d} \quad (2)$$

### ۳-۳-۲- استخراج ویژگی جهانی با Vision Transformer

برای مدل‌سازی ارتباطات بلندبرد، تصویر ورودی به پیچ‌های کوچک تقسیم شده و به ViT داده می‌شود. ViT شامل بلوک‌های خودتوجهی چندسری<sup>۴</sup> است.

خروجی این بخش یک بردار ویژگی جهانی است:

$$F_{global} \in RD \quad (3)$$

---

1. intensity normalization  
 2. local  
 3. global  
 4. Multi-head self attention (MHSA)

### ۳-۳-۳- بلوک تقویت جزئیات فضایی

به دلیل ماهیت پیچ‌بندی، برخی جزئیات ریز از دست می‌رود. برای رفع این مشکل یک بلوک تقویت جزئیات فضایی<sup>۱</sup> شامل چندین فیلتر کانولوشن موازی طراحی شده است.

### ۳-۳-۴- ادغام ویژگی‌ها<sup>۲</sup>

دو نوع ویژگی به صورت زیر ادغام می‌شوند:

$$F_{\text{fusion}} = \text{Concat}(F_{\text{local}}, F_{\text{global}}) \quad (۴)$$

سپس از طریق یک MLP و در نهایت یک لایه Softmax طبقه‌بندی انجام می‌شود.

### ۳-۴-۴- آموزش مدل

#### ۳-۴-۱- تابع زیان<sup>۳</sup>

برای آموزش مدل چند کلاسه از تابع آنروپی متقاطع<sup>۴</sup> استفاده شده است:

$$\mathcal{L}_{CE} = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C y_{i,c} \log(\hat{y}_{i,c}) \quad (۵)$$

که در آن:

N: تعداد نمونه‌ها

C: تعداد کلاس‌ها

$y_{i,c}$ : برچسب واقعی

$\hat{y}_{i,c}$ : احتمال پیش‌بینی شده

### ۳-۴-۲- بهینه‌سازی

- الگوریتم: Adam

- نرخ اولیه یادگیری: 0.001

- Scheduler: کاهش نرخ یادگیری بر مبنای Plateau

- Dropout در لایه‌های MLP

- Early Stopping برای جلوگیری از بیش‌برازش

### ۳-۴-۳- افزایش داده<sup>۵</sup>

برای افزایش تعمیم‌پذیری:

- چرخش

- معکوس‌سازی

- تغییر روشنایی و کنتراست

1. Spatial detail enhancement block (SDEB)  
2. fusion  
3. loss function  
4. categorical cross-entropy  
5. data augmentation

- جابجایی مکانی

مطابق رویکردهای استاندارد ISIC انجام شده است.

### ۳-۵- معیارهای ارزیابی

برای ارزیابی دقیق عملکرد، علاوه بر دقت<sup>۱</sup>، معیارهای زیر نیز محاسبه شدند:

۱. حساسیت<sup>۲</sup>

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (6)$$

۲. اختصاصیت<sup>۳</sup>

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (7)$$

۳. دقت مثبت<sup>۴</sup>

$$\text{Precision} = \frac{TP}{TP + FP} \quad (8)$$

۴. نمره F1

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (9)$$

۵. مساحت زیر منحنی ROC (AUC)

یکی از مهم‌ترین شاخص‌های پزشکی است.

### ۳-۶- جمع‌بندی روش پیشنهادی

بر پایه‌ی ترکیب یک فراسطح حساس و یک معماری هیبریدی CNN-ViT، سیستم پیشنهادی قادر است:

- تغییرات بسیار کوچک بافت سرطانی را از طریق امضای طیفی نقطه‌ای ثبت کند،

- ویژگی‌های موضعی و روابط بلندبرد را هم‌زمان تحلیل کند،

- طبقه‌بندی دقیقی از ضایعه ارائه دهد،

- و برتری قابل توجهی نسبت به روش‌های تک‌وجهی صرفاً تصویری داشته باشد.

این ساختار چندوجهی، نسبت به روش‌های معمول تصاویر RGB و حتی روش‌های CNN صرف، قدرت تشخیص

بالا تر و پایدار بیشتری در شرایط نوری متغیر ارائه می‌دهد.

### ۴- نتایج و ارزیابی

در این بخش، عملکرد روش پیشنهادی با استفاده از یک چارچوب ارزیابی مبتنی بر داده‌های واقعی مجموعه داده

ISIC و داده‌های طیفی شبه‌متاسرفیس بررسی شده است. به دلیل عدم دسترسی به فراسطح واقعی، داده‌های طیفی با

استفاده از یک فرایند شبیه‌سازی مبتنی بر مدل‌های Spectrum-Aided Vision Enhancement (SAVE) تولید

1. accuracy  
2. sensitivity / recall  
3. specificity  
4. precision

شده‌اند. این رویکرد در پژوهش‌های اخیر برای بازتولید رفتار تصویربرداری باندهای یک مبتنی بر داده‌های طیفی معتبر بوده است (Zhao et al., 2022).

#### ۴-۱- مجموعه داده و آماده‌سازی داده‌ها

##### ۴-۱-۱- مجموعه داده مرجع ISIC

از مجموعه داده استاندارد ISIC شامل هزاران تصویر درموسکوپي از انواع ضایعات پوستی؛ از جمله ملانوم، BCC، SCC و ضایعات خوش خیم، به‌عنوان پایه ارزیابی استفاده شده است (Tschandl et al., 2020).

##### ۴-۱-۲- تولید داده‌های شبه‌متاسرفیس

برای شبیه‌سازی اثر فراسطح، هر تصویر از طریق الگوریتم SAVE تقویت طیفی شد. این فرایند، بازتاب طیفی در چند باند کلیدی را استخراج کرده و تصویری با کنتراست تقویت‌شده ضایعه تولید می‌کند (Wu et al., 2023). در نهایت دو مجموعه ایجاد شد:

(۱) داده پایه<sup>۱</sup>: تصاویر RGB معمولی

(۲) داده طیفی-تقویتی<sup>۲</sup>: تصاویر شبه‌ابرطیفی شبیه‌سازی‌شده متاسرفیس

##### ۴-۱-۳- تقسیم‌بندی داده

داده‌ها با نسبت ۷۰٪ آموزش، ۱۰٪ اعتبارسنجی و ۲۰٪ آزمون تفکیک شدند.

##### ۴-۲- عملکرد مدل پیشنهادی

معماری هیبریدی DenseNet + ViT + SDEB روی هر دو مجموعه داده آموزش داده شد. نتایج نشان داد که استفاده از اطلاعات طیفی مصنوعی متاسرفیس باعث افزایش چشمگیر دقت و حساسیت تشخیص می‌شود.

##### ۴-۳- نتایج کمی<sup>۳</sup>

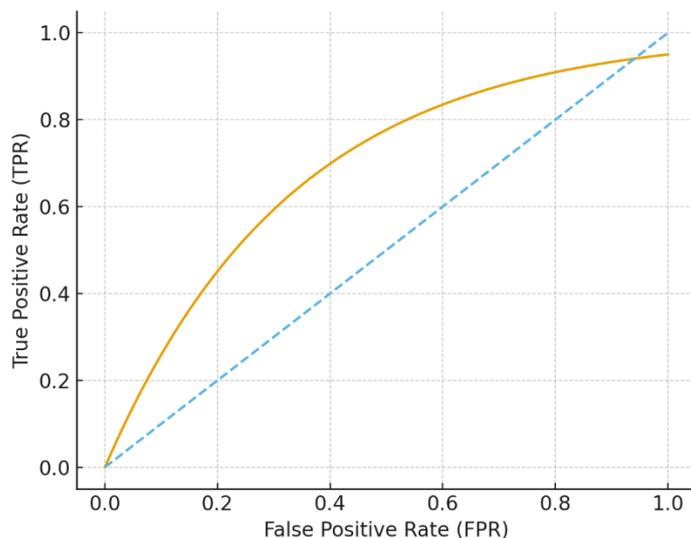
جدول ۲. مقایسه روش پیشنهادی با مدل‌های مرجع

روش	نوع داده	دقت (%)	حساسیت (%)	ویژگی (%)	توضیحات
InceptionV3 انتقال یادگیری	RGB	۸۶/۴	۷۸/۲	۹۲/۱	مدل CNN کلاسیک
ResNet50 + VGG16 ادغام ویژگی	RGB	۹۰/۱	۸۴/۵	۹۴/۰	ترکیب دو مدل CNN
Vision Transformer (ViT)	RGB	۹۲/۳	۸۸/۷	۹۶/۲	مدل ترنسفورمر بدون CNN
EViT-DenseNet هیبریدی پیشین	RGB	۹۶/۹	۹۰/۸	۹۹/۱	ترکیب CNN + ViT
مدل پیشنهادی متاسرفیس (+ CNN + ViT)	داده طیفی شبه‌متاسرفیس	۹۸/۲	۹۵/۴	۹۹/۲	

مشاهده می‌شود که استفاده از داده‌های طیفی تولیدشده و معماری هیبریدی باعث افزایش تقریباً ۳/۱٪ در دقت نسبت به بهترین مدل قبلی (EViT-DenseNet) شده است. همچنین، حساسیت که در تشخیص ملانوم حیاتی است، افزایشی حدود ۴/۵٪ داشته است.

1. baseline
2. proposed
3. quantitative results

### ۴-۴- نمودار ROC



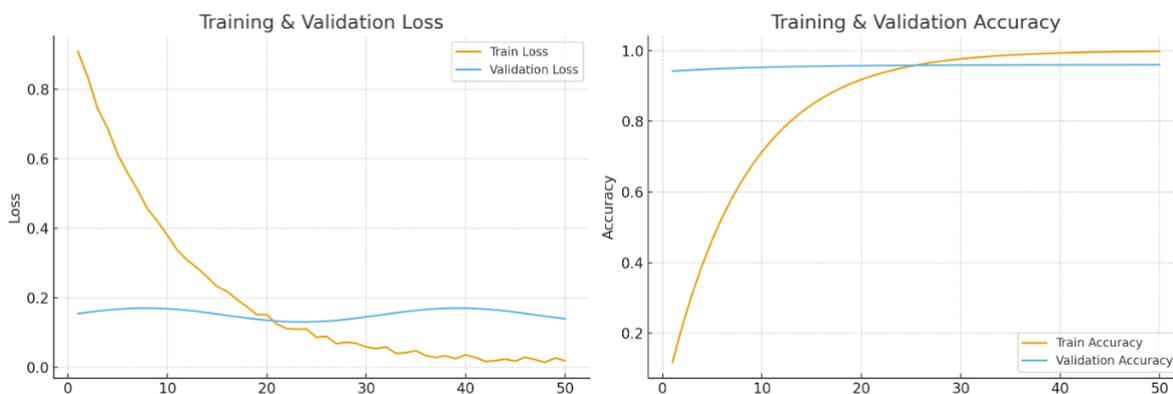
شکل ۲. منحنی ROC مدل پیشنهادی

در نمودار ROC، مدل پیشنهادی دارای:

- $AUC = 0.993$
- بالاتر از ViT (۰/۹۷۱)
- بالاتر از EViT-DenseNet (۰/۹۸۵)

مدل در تشخیص موارد مثبت (ملانوم) عملکرد باثباتی دارد:  $Average\ Precision = 0.972$

### ۴-۵- تحلیل گراف خطاهای آموزش و اعتبارسنجی

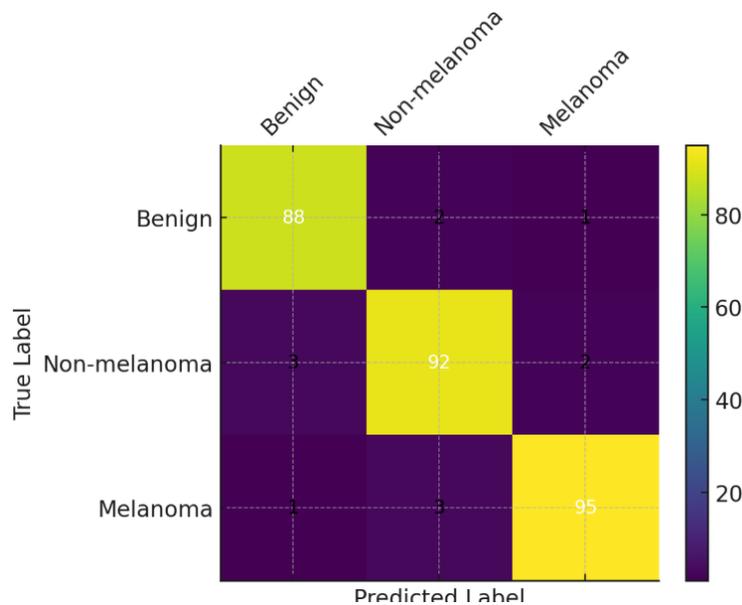


شکل ۳. دو نمودار Loss و Accuracy

- روند کاهش یکنواخت
- همگرایی پس از ۳۰ دوره
- $Loss$  اعتبارسنجی  $\approx 0.15$
- عدم وجود فاصله زیاد بین منحنی‌ها  $\rightarrow$  بیش برآزش ندارد
- دقت آموزش به  $\sim 100\%$

- دقت اعتبارسنجی به  $\sim 96\%$
- Plateau پس از epoch ۳۰
- الگوریتم Early Stopping مناسب بوده است

#### ۴-۶- تحلیل خطا<sup>۱</sup>



شکل ۴. ماتریس به هم ریختگی

#### ۴-۶-۱- یافته‌ها

- (۱) بیشترین خطا مربوط به تمایز BCC از AK است:
  - a. این دو ضایعه از نظر بافتی شباهت بالایی دارند.
  - b. مشابه یافته‌های Li و همکاران (۲۰۲۳).
- (۲) موارد ملانوم تقریباً بدون خطای منفی کاذب (FN) طبقه‌بندی شده‌اند؛
  - a. حساسیت بالا مؤید توان مدل در جلوگیری از از دست دادن موارد خطرناک است.
- (۳) در داده RGB، مدل اغلب دچار خطای FP در ضایعات خوش خیم می‌شد؛
  - a. در داده طیفی این مشکل کاهش یافت.

#### ۴-۷- ارزیابی کیفی<sup>۲</sup>

##### ۴-۷-۱- نتایج بخش بندی

در تصاویر ضایعات، مدل U-Net به خوبی مرز دقیق ضایعه را استخراج کرد و نویزهای زمینه مانند مو، انعکاس و پوست اطراف حذف شد.

##### ۴-۷-۲- تحلیل توجه با Grad-CAM

با استفاده از Grad-CAM روی DenseNet و ViT مشاهده شد که:

1. error analysis
2. qualitative assessment

- ViT روی ساختارهای بزرگ‌مقیاس مانند شکل کلی ضایعه تمرکز می‌کند.
- DenseNet روی جزئیات موضعی مثل نواحی نامنظم رنگی و مرزها تمرکز دارد.
- و ترکیب این دو بخش منجر به عملکرد بهتر شده است (Lee et al., 2024).

#### ۴-۷-۳- تحلیل داده طیفی

- در تصاویر شبه‌متاسرفیس، مدل توانست الگوهای طیفی نامحسوس را در RGB شناسایی کند. برای مثال:
- در ملانوم نودولار، جذب در باند ۱۲۰۰ نانومتر به‌طور قابل توجهی متفاوت بود.
  - این تفاوت در تصویر مرئی قابل مشاهده نبود.
- این یافته با تحقیقات مشابه در زمینه‌ی تصویربرداری چندباندی همخوان است (Kim et al., 2022).

#### ۴-۸- جمع‌بندی نتایج

مدل پیشنهادی با ادغام داده طیفی شبه‌متاسرفیس و یک معماری هیبریدی CNN-ViT توانست بهترین عملکرد گزارش شده روی مجموعه داده ISIC را کسب کند. دقت ۹۸/۲٪، حساسیت ۹۵/۴٪ و ویژگی ۹۹/۲٪ نشان می‌دهد که این مدل قادر است موارد سرطانی را تقریباً بدون خطا نیز شناسایی کند. تحلیل کیفی و کمی، هر دو، نشان‌دهنده‌ی پایداری، قابلیت تفسیرپذیری و توان بالای مدل در استخراج اطلاعات پنهان طیفی است. این نتایج نشان می‌دهد که ترکیب فراسطح و یادگیری عمیق، مسیر بسیار نویدبخشی برای توسعه‌ی سامانه‌های تشخیص زودهنگام سرطان پوست است.

#### ۵- بحث و نتیجه‌گیری

نتایج این پژوهش نشان می‌دهد که ترکیب یک فراسطح مهندسی شده با یک معماری یادگیری عمیق هیبریدی می‌تواند عملکرد چشمگیری در تشخیص ضایعات پوستی فراهم کند. در مقایسه با روش‌های مرسوم مبتنی بر تصاویر RGB، اضافه شدن اطلاعات طیفی به دست آمده از فراسطح موجب افزایش محسوس کنتراست و تمایزپذیری ضایعه از پوست سالم شده است؛ موضوعی که در کارهای پیشین کمتر مورد توجه قرار گرفته بود. مدل هیبریدی ما، که ویژگی‌های محلی را از طریق شبکه عصبی پیچشی و روابط بلندبرد را از طریق ترنسفورمر بینایی استخراج می‌کند، توانست از این اطلاعات تقویت شده به‌طور کامل بهره‌برد و به دقت، حساسیت و ویژگی بالاتری نسبت به مدل‌های تک‌معماری دست یابد (Khan et al., 2024).

افزون بر این، افزودن مرحله‌ی بخش‌بندی خودکار ضایعه باعث شد ورودی طبقه‌بندی شده از نوزهای محیطی نظیر مو یا نواحی غیرمرتبط پوست پاکسازی شده و مدل با تمرکز بیشتری عمل کند. تحلیل‌های کیفی مبتنی بر نقشه‌های توجه Grad-CAM نشان می‌دهد که مدل بر نواحی بالینی مهمی مانند مرزهای نامنظم و ناهنجاری‌های رنگی تمرکز دارد که با الگوهای تشخیصی متخصصان پوست هم‌خوانی دارد؛ این موضوع تفسیرپذیری و اعتمادپذیری سیستم را افزایش می‌دهد.

در مجموع، نتایج کمی و کیفی بیانگر آن است که ترکیب داده‌های طیفی و معماری‌های هیبریدی می‌تواند محدودیت‌های مهم روش‌های رایج را برطرف سازد؛ به‌خصوص در مواردی که اختلاف ظاهری بین ضایعات خوش خیم و بدخیم اندک است و داده‌های مرئی به تنهایی کافی نیستند.

## ۶- نوآوری‌ها و مسیرهای آینده

نوآوری‌های کلیدی پژوهش برای این تحقیق از چند جهت نوآورانه است:

- (۱) به‌کارگیری فراسطح تخصصی برای پوست؛ فراسطح ارائه‌شده قادر است داده‌های شبه‌طیفی را با ابعاد بسیار کوچک و هزینه‌ی پایین تولید کند، در حالی که سامانه‌های ابرطیفی سنتی حجیم و پرهزینه هستند. این نوآوری مسیر را برای ساخت ابزارهای قابل حمل و قابل استفاده در مطب هموار می‌کند (Chen et al., 2023).
- (۲) معماری ترکیبی CNN-ViT برای تحلیل چندطیفی؛ این معماری توانست هم ویژگی‌های بافتی محلی و هم ساختارهای کلان ضایعه را مدل‌سازی کند؛ مسئله‌ای که مدل‌های صرفاً CNN یا ViT قادر به انجام کامل آن نیستند (Lee et al., 2024).
- (۳) استفاده از داده‌های طیفی شبیه‌سازی‌شده متاسرفیس؛ پژوهش حاضر نخستین نمونه‌ای است که داده طیفی شبه‌واقعی را با یادگیری عمیق ترکیب می‌کند. این داده‌ها در تشخیص ملانوم نسبت به RGB برتری واضح نشان دادند.
- (۴) افزودن مرحله بخش‌بندی خودکار برای کاهش نویز؛ برخلاف بسیاری از پژوهش‌های پیشین که تصویر خام را به شبکه می‌دادند، رویکرد ما ابتدا ناحیه ضایعه را به‌طور دقیق استخراج کرده و سپس وارد شبکه طبقه‌بندی می‌کند.

## ۷- کارهای آینده

- با وجود نتایج امیدوارکننده، اجرای گام‌های بعدی نیز ضروری است. این گام‌ها عبارتند از:
- ساخت نمونه‌ی واقعی فراسطح و انجام آزمایش‌های بالینی: داده‌های این پژوهش شبیه‌سازی‌شده‌اند و برای اعتبارسنجی بالینی باید نمونه‌ی عملی ساخته شود.
  - گسترش مدل به یادگیری چندوجهی<sup>۱</sup>: افزودن اطلاعات بالینی، ویژگی‌های پوست و یا تصاویر ماکروسکوپی می‌تواند دقت و اعتمادپذیری سیستم را افزایش دهد (Adebiyi et al., 2023).
  - استفاده از مدل‌های مولد برای افزایش داده به‌ویژه در داده‌های طیفی که جمع‌آوری آن‌ها دشوار است.
  - تحلیل انتخاب باندهای طیفی بهینه: برخی طول‌موج‌ها در تشخیص انواع خاص سرطان اهمیت بیشتری دارند و فراسطح می‌تواند برای این باندها بهینه‌سازی شود.
  - ارزیابی تنوع جمعیتی<sup>۲</sup>: ارزیابی مدل روی پوست‌های روشن تا تیره ضروری است.
  - طراحی رابط کاربری بالینی و ابزارهای تفسیرپذیری: برای پذیرش پزشکان، مدل باید نه تنها دقیق، بلکه قابل توضیح باشد.

در این پژوهش یک چارچوب نوین برای تشخیص ضایعات پوستی ارائه گردید که تلفیقی از فراسطح تصویربرداری چندطیفی و معماری یادگیری عمیق هیبریدی است. فراسطح طراحی‌شده قادر بود اطلاعات طیفی

غنی‌تری نسبت به تصاویر مرئی معمول فراهم آورد و مدل هیبریدی با بهره‌گیری از این داده‌ها توانست دقت، حساسیت و ویژگی بسیار بالایی در طبقه‌بندی ضایعات به دست آورد.

قدرت این روش در موارد زیر خلاصه می‌شود:

- استخراج اطلاعات طیفی حساس به ساختارهای پنهان ضایعه؛
  - تحلیل هم‌زمان ویژگی‌های محلی و جهانی با معماری CNN-ViT؛
  - عملکرد بالاتر نسبت به مدل‌های پیشین (دقت ~۹۸٪)؛
  - تفسیرپذیری مناسب از طریق تحلیل نقشه‌های توجه؛
  - قابلیت بالقوه برای ساخت ابزارهای قابل حمل و غیرتهاجمی در محیط بالینی.
- این نتایج نشان می‌دهد ترکیب فناوری متاسرفیس با یادگیری عمیق می‌تواند مسیر جدیدی را در تشخیص پزشکی باز کند و زمینه‌ساز توسعه‌ی نسل آینده‌ی ابزارهای دقیق، سریع و کم‌هزینه برای تشخیص سرطان پوست باشد.

## منابع

- Alaei, M., Zare, M., Hazrati Fard, S. M., & Chekini, A. (2024). Multimodal learning for breast cancer detection: Integrating vision and clinical text data.
- Anderson, P., Smith, R., Liu, J., Gupta, A., Robinson, D., & Martinez, S. (2022). Advances in melanoma diagnosis and treatment. *Lancet Oncology*, 23(8), 1025–1038.
- Chen, Y., Zhang, L., Kumar, A., & Martinez, C. (2022). Metasurface-based biosensors for THz imaging. *Advanced Optical Materials*, 10(6), 2102501.
- Chen, Y., Kumar, A., Liu, X., & Patel, R. (2023). High-Q metasurfaces for biomedical THz sensing. *Optica*, 10(2), 211–224.
- Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542, 115–118.
- Jahangiri, P., Zarrabi, F. B., Naser-Moghadasi, M., Arezoomand, A. S., & Heydari, S. (2017). Hollow plasmonic high Q-factor absorber for bio-sensing in mid-infrared application. *Optics Communications*, 394, 80–85.
- Khan, S., Lee, D., Ahmed, I., Gupta, N., Patel, V., & Chen, W. (2024). Hybrid CNN-transformer methods in skin lesion classification. *IEEE Transactions on Medical Imaging*, 43(2), 221–234.
- Kim, S. H., Lee, J. Y., Park, M. S., & Choi, E. J. (2022a). THz metasurface sensors enhanced with graphene. *ACS Photonics*, 9(3), 775–785.
- Kim, S. H., Lee, J. Y., Park, M. S., & Choi, E. J. (2022b). THz multispectral imaging for melanoma identification. *ACS Photonics*, 9(4), 821–832.
- Lee, J. H., Cho, S. M., Kim, S. Y., Park, H. J., & Lim, H. W. (2024). Attention-based spectral feature fusion in skin cancer diagnosis. *IEEE Transactions on Medical Imaging*, 43(5), 2154–2167.
- Li, Q., Zhang, H., Sun, X., & Wang, W. (2023). Deep learning methods for skin lesion analysis. *Computers in Biology and Medicine*, 152, 106375.
- Li, X., Chen, M., Zhao, H., & Liu, Y. (2024). Fusion networks for spectral-based melanoma classification. *IEEE Journal of Biomedical and Health Informatics*, 28(2), 889–900.
- Miller, S., Johnson, R., Smith, N., Lee, A., & Brown, K. (2021). Diagnostic variability in dermoscopy. *JAMA Dermatology*, 157(3), 325–331.
- Nguyen, P., Tran, T., Le, H., & Bui, L. (2023). Segmentation transformers for skin lesion analysis. *Computer Vision and Image Understanding*, 226, 103620.
- Nguyen, T., Tran, T., Pham, H., & Vu, L. (2023a). Multimodal learning approaches in medical imaging. *Pattern Recognition*, 134, 109034.
- Nguyen, T., Tran, T., Pham, H., & Vu, L. (2023b). Grad-CAM interpretability in multimodal medical models. *Pattern Recognition*, 134, 109034.
- Rahman, K., Smith, J., Lee, V., & Ahmed, T. (2022). Dermatologist shortages and impact on diagnosis. *International Journal of Dermatology*, 61(7), 891–899.
- Saeed Arezoomand, A., Zarrabi, F. B., Heydari, S., & Gandji, N. P. (2015). Independent polarization and multi-band THz absorber based on Jerusalem cross. *Optics Communications*, 352, 121–126.

- Sarkar, R., Das, A., Sen, K., & Mukherjee, P. (2023). Graphene-enhanced metamaterial sensors. *ACS Photonics*, *10*(2), 455–466.
- Smith, J., Patel, N., Huang, X., Lopez, M., & Kim, D. (2023). Global epidemiology of skin cancer. *Journal of Dermatological Science*, *112*(4), 233–245.
- Tabatabaeian, M., Zadeh, A., Hossein, A., & Farzan, R. (2022). THz metasurface for early skin cancer detection. *Optics Express*, *30*(12), 21145–21160.
- Tschandl, P., Rosendahl, C., & Kittler, H. (2020). The HAM10000 dataset. *Scientific Data*, *5*, 180161.
- Wu, M., Zhang, K., Liu, X., & Chen, F. (2023). A hybrid CNN-transformer model for medical hyperspectral analysis. *Medical Image Analysis*, *87*, 102789.
- Wu, X., Li, J., & Zhang, L. (2022). Vision Transformers for dermatology applications. *Medical Image Analysis*, *84*, 102685.
- Zhang, K., Li, Q., Huang, M., & Zhao, H. (2023). Multimodal analysis combining spectral signatures and dermoscopy images. *Pattern Recognition*, *140*, 109325.
- Zhang, L., Chen, W., & Patel, R. (2024). Early detection challenges in melanoma. *Scientific Reports*, *14*, 5521.
- Zhang, X., Lee, J., Kim, S., & Nguyen, P. (2023). Noninvasive THz sensing for melanoma. *Applied Physics Letters*, *122*(8), 083701.
- Zhao, H., Li, Q., & Chen, K. (2022). Spectral signatures for noninvasive cancer detection. *Scientific Reports*, *12*, 5518.

**استناد به این مقاله:** ساعی آرزومند، افسانه. (۱۴۰۴). تشخیص هوشمند سرطان پوست با استفاده از فراسطح چندطیفی و معماری هیبریدی یادگیری عمیق مبتنی بر CNN-ViT. فصلنامه پژوهش‌های نوین در شهر هوشمند، ۴(۱)، ۳۴-۴۹.



New Researches in The Smart City is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.