Journal of Soft Computing and Information Technology (JSCIT)
Babol Noshirvani University of Technology, Babol, Iran
Journal Homepage: jscit.nit.ac.ir
DOI: 10.22034/jscit.2026.542434.2162
Volume 15, Number 1, Spring 2026, pp. 17-28
Received: 08/24/2025, Revised: 12/27/2025, Accepted: 12/29/2025

Test-Time Augmentation with Multi-Scale Strategies and Test-Time Dropout

Emadeddin Asghari, Hedieh Sajedi *
School of Mathematics, Statistics and Computer Science, College of Science, University of Tehran, Tehran, Iran.
emadeddinasghari@ut.ac.ir, * hhsajedi@ut.ac.ir

Corresponding author's address: Hedieh Sajedi, School of Mathematics, Statistics and Computer Science, College of Science, University of Tehran, Tehran, Iran.

Abstract- Skin lesion segmentation, despite advances in deep learning models, still suffers from performance degradation when faced with challenges such as variations in image quality, lesion heterogeneity, and distribution shifts in test data. Since deep learning models are typically trained once in research environments and, after being deployed in clinical systems, retraining or fine-tuning them with new data is not easily feasible due to the need for labeled data, time, and substantial computational resources, this study proposes a lightweight, architecture-free approach to enhance the inference performance of the baseline EGE-UNet model. The proposed method applies two complementary test-time techniques, Test-Time Augmentation (TTA) and Test-Time Dropout (TTD), along with their multi-scale variant, as an additional inference layer without modifying the pretrained network. The core idea is to generate multiple augmented or rescaled versions of the input image and execute the pretrained model multiple times with active Dropout, producing more stable and reliable predictions through appropriate averaging. Experiments on the ISIC2017 and ISIC2018 datasets show that combining TTA and TTD improves segmentation accuracy and reduces prediction variance compared to the baseline, while combining the multi-scale version of TTA and TTD yields sharper boundary delineation in ISIC2017; However, there was no improvement in ISIC2018. Because the proposed strategy requires no retraining and can be applied to any segmentation architecture, it represents a practical and low-overhead solution for improving inference robustness in medical imaging applications.

Keywords- Skin lesion segmentation, Test-Time Augmentation, Test-Time Dropout, Multi-Scale Inference, Ensembling, Medical image analysis.

مجله علمی رایانش نرم و فناوری اطلاعات
دانشگاه صنعتی نوشیروانی بابل
صفحه مجله: jscit.nit.ac.ir
شناسه دیجیتال: 10.22034/jscit.2026.542434.2162
جلد ۱۵، شماره ۱، بهار ۱۴۰۵، صفحه ۱۷-۲۸
دریافت: ۱۴۰۴/۰۶/۰۲، بازنگری: ۱۴۰۴/۱۰/۰۶، پذیرش: ۱۴۰۴/۱۰/۰۸

افزایش داده با تأکید بر استراتژی‌های چندمقیاسی و حذف تصادفی در زمان آزمون

عمادالدین اصغری، هدیه ساجدی*
دانشکده ریاضی، آمار و علوم کامپیوتر، دانشکدگان علوم، دانشگاه تهران، تهران، ایران.
emadeddinasghari@ut.ac.ir, * hhsajedi@ut.ac.ir

* نشانی نویسنده مسئول: هدیه ساجدی، دانشکده ریاضی، آمار و علوم کامپیوتر، دانشکدگان علوم، دانشگاه تهران، تهران، ایران.

چکیده- بخش‌بندی ضایعات پوستی با وجود پیشرفت مدل‌های یادگیری عمیق همچنان در مواجهه با چالش‌هایی مانند تغییر کیفیت تصویر، ناهمگونی ضایعات و اختلاف توزیع داده‌های آزمون دچار کاهش دقت می‌شود. از آن‌جا که مدل‌های یادگیری عمیق معمولاً یک‌بار در محیط‌های پژوهشی آموزش داده می‌شوند و پس از استقرار در سامانه‌های بالینی، بازآموزی یا تنظیم مجدد آن‌ها با داده‌های جدید به‌دلیل نیاز به داده برچسب‌دار، زمان و منابع محاسباتی به‌سادگی انجام نمی‌شود، در این پژوهش روشی سبک و بدون تغییر معماری برای بهبود عملکرد مرحله استنتاج مدل پایه EGE-UNet ارائه می‌شود. در این روش، دو روش مکمل شامل تقویت زمان آزمون (TTA) و حذف تصادفی زمان آزمون (TTD) و نسخه چندمقیاسی TTA، به‌عنوان لایه‌ای افزوده بر خروجی مدل پیش‌آموزش‌دیده اعمال شدند تا پایداری پیش‌بینی‌ها و مقاومت مدل در برابر تغییرات ورودی افزایش یابد. ایده اصلی، ایجاد چند نسخه افزوده‌شده یا مقیاس‌یافته از تصویر و اجرای چندباره مدل با Dropout فعال است که با میانگین‌گیری مناسب، خطای پیش‌بینی کاهش می‌یابد. ارزیابی‌ها روی مجموعه‌داده‌های ISIC2017 و ISIC2018 نشان می‌دهد که ترکیب TTA و TTD موجب بهبود دقت و کاهش نوسان نسبت به مدل پایه می‌شود و ترکیب نسخه چندمقیاسی TTA و TTD در ISIC2017 مرزبندی دقیق‌تری تولید می‌کند؛ هرچند در ISIC2018 بهبودی حاصل نشد. روش پیشنهادی بدون نیاز به بازآموزی، قابل تعمیم به سایر معماری‌های بخش‌بندی بوده و در کاربردهای حساس پزشکی، گزینه‌ای عملی و کم‌هزینه برای افزایش دقت مرحله استنتاج است.

واژه‌های کلیدی: بخش‌بندی ضایعات پوستی، تقویت زمان آزمون، حذف تصادفی زمان آزمون، استنتاج چندمقیاسی، عدم قطعیت تصادفی، تحلیل تصویر پزشکی.

۱- مقدمه

در سال‌های اخیر، مدل‌های بخش‌بندی مبتنی بر یادگیری عمیق به‌ویژه در حوزه تحلیل ضایعات پوستی پیشرفت چشمگیری داشته‌اند. با این حال، عملکرد این مدل‌ها در مرحله‌ی استنتاج به‌شدت تحت‌تأثیر عواملی همچون تنوع شرایط تصویربرداری، تغییر کیفیت، نویز، و اختلاف توزیع داده‌ها قرار می‌گیرد. این مسئله در کاربردهای حساس پزشکی اهمیت بیشتری می‌یابد؛ زیرا کوچک‌ترین خطا در مرزبندی ضایعات می‌تواند موجب تشخیص نادرست یا تصمیم‌گیری اشتباه بالینی شود. از سوی دیگر، بازآموزی مدل‌ها یا اصلاح معماری آن‌ها در محیط‌های واقعی معمولاً امکان‌پذیر نیست؛ زیرا به داده‌های برچسب‌دار، منابع محاسباتی قابل‌توجه، و زمان طولانی نیاز دارد.

به‌منظور رفع این محدودیت، روش‌هایی مانند تقویت زمان آزمون (TTA)[1] [۱] و حذف تصادفی زمان آزمون (TTD)[2] [۲] مطرح شده‌اند که با ایجاد نسخه‌های اصلاح‌شده از تصویر یا انجام چندین عبور تصادفی از مدل، تنوع ورودی و خروجی را افزایش داده و پایداری پیش‌بینی را بهبود می‌بخشند. با وجود کارایی هر یک از این روش‌ها به‌صورت مستقل، تاکنون ترکیب آن‌ها و همچنین توسعه نسخه چندمقیاسی آن‌ها برای بخش‌بندی ضایعات پوستی به‌طور نظام‌مند بررسی نشده است.

علاوه بر این، مدل EGE-UNet [۳] با وجود عملکرد مناسب، همچنان در مواجهه با تصاویر ناهمگون، مرزهای نامشخص و اختلاف توزیع میان داده‌های آموزش و آزمون دچار نوسان در پیش‌بینی می‌شود؛ مشکلی که می‌توان بدون تغییر معماری و تنها با تقویت مرحله استنتاج آن را برطرف کرد.

در این پژوهش، یک لایه‌ی استنتاجی سبک معرفی می‌شود که کاملاً مستقل از فرآیند آموزش بوده و بر روی خروجی مدل پیش‌آموزش‌دیده EGE-UNet اعمال می‌گردد. نوآوری اصلی کار حاضر در دو بخش خلاصه می‌شود:

(۱) ترکیب هم‌زمان TTA و TTD برای ایجاد مجموعه‌ای از پیش‌بینی‌های متنوع و کاهش عدم‌قطعیت در مرحله‌ی استنتاج، و

(۲) بسط این ترکیب به حالت چندمقیاسی (MSTTA + TTD) برای بهره‌گیری هم‌زمان از ساختارهای بزرگ‌مقیاس و جزئیات ریز تصویر.

در هر دو رویکرد، مدل پایه بدون هیچ تغییری ثابت نگه داشته می‌شود و تنها نسخه‌های افزوده‌شده، مقیاس‌یافته یا عبورهای تصادفی Dropout تولید شده و خروجی‌های مربوطه با یکدیگر ادغام می‌شوند. هدف آن است که بدون هیچ‌گونه هزینه آموزش مجدد، عملکرد مدل در مواجهه با تصاویر واقعی و خارج از توزیع بهبود یابد.

برای ارزیابی جامع، روش پیشنهادی بر روی دو مجموعه‌داده‌ی مرجع ISIC2017 و ISIC2018 آزمایش شد. نتایج نشان می‌دهد که افزودن این لایه‌ی استنتاجی سبک باعث افزایش پایداری، کاهش نوسان خروجی و بهبود دقت بخش‌بندی می‌شود و این رویکرد می‌تواند بدون بازآموزی، به‌راحتی در کنار سایر معماری‌های بخش‌بندی به‌کار گرفته شود.

هیچ‌یک از مطالعات قبلی ترکیب هم‌زمان TTA و TTD و نسخه چندمقیاسی TTA را به‌عنوان یک لایه استنتاجی بدون بازآموزی روی مدل EGE-UNet بررسی نکرده‌اند.

۲- مدل پایه (EGE-UNet)

در این پژوهش، مدل EGE-UNet به عنوان شبکه پایه برای بخش‌بندی ضایعات پوستی استفاده شده است. این مدل یک معماری سبک و بهینه‌شده بر مبنای U-Net است که با هدف بهبود استخراج ویژگی‌های مرزی و افزایش دقت در نواحی کوچک و ناهمگون طراحی شده است. مطابق مقاله اصلی، معماری شامل سه مؤلفه کلیدی است:

۲-۱- ماژول هدایت‌شده با لبه (EGM)

ماژول هدایت‌شده با لبه (EGM)[3] با استخراج ویژگی‌های لبه‌ای از لایه‌های کم‌عمق‌تر، اطلاعات ساختاری مهم را به مسیر دیکودر منتقل می‌کند. این کار باعث افزایش دقت بخش‌بندی در نواحی با مرزهای ضعیف، محوشده یا پیچیده می‌شود؛ مشکلی که در ضایعات پوستی بسیار شایع است.

۲-۲- بلوک توجه مکانی–کانالی (SCAB)

بلوک توجه مکانی-کانالی (SCAB)[4] با مدل‌سازی وابستگی‌های مکانی و ارتباط میان کانال‌ها، مهم‌ترین الگوهای بافتی را تقویت و نویزهای غیرمرتبط را تضعیف می‌کند.

این ویژگی به‌خصوص در داده‌های ISIC که دارای نوردهی غیرهمگن، سایه، یا تفاوت رنگ‌پوست هستند، اهمیت بالایی دارد.

۲-۳- نظارت عمیق

منظور از نظارت عمیق [5]آن است که در مسیر بالارونده دیکودر، خروجی‌های میانی نیز تحت نظارت قرار می‌گیرند. این کار باعث بهبود همگرایی، تقویت ویژگی‌های چندمقیاسی و بازسازی دقیق ساختارهای ریز می‌شود.

مدل پایه شامل یک مسیر رمزگذار (Encoder) برای استخراج ویژگی‌ها و یک مسیر رمزگشا (Decoder) برای بازسازی نقشه بخش‌بندی است. اتصالات پرشی بین لایه‌های متناظر Encoder و Decoder، همراه با EGM و SCAB، امکان انتقال ویژگی‌ها در چند سطح را فراهم می‌کنند و باعث غنای اطلاعاتی در خروجی نهایی می‌شوند.

در این مطالعه از نسخه پیش‌آموزش‌دیده مدل که بر اساس داده‌های ISIC آموزش داده شده بدون هیچ‌گونه تغییر در وزن‌ها، تنظیمات آموزش یا ساختار شبکه استفاده شده است. مدل پایه کاملاً ثابت نگه داشته شده و تمام روش‌های پیشنهادی تنها در زمان آزمون استنتاج به خروجی مدل اعمال شده‌اند.

۳- مرور روش‌های موجود

۳-۱- تقویت زمان آزمون (TTA)

TTA یکی از پرکاربردترین روش‌ها برای افزایش پایداری و دقت مدل‌های یادگیری عمیق در مرحله‌ی استنتاج است. در این رویکرد، به‌جای آن‌که مدل تنها یک نسخه از تصویر ورودی را پردازش کند، چندین نسخه‌ی تغییریافته از همان تصویر ایجاد شده و مدل روی تک‌تک آن‌ها اجرا می‌شود. این تبدیلات می‌توانند شامل چرخش، وارون‌سازی افقی یا عمودی، برش، تغییر مقیاس یا دیگر تغییرات ظاهری باشند.

برای هر تصویر، خروجی‌های حاصل پس از اعمال معکوس تبدیلات، به فضای اصلی بازگردانده شده و معمولاً به‌صورت میانگین‌گیری ادغام می‌شوند. به این ترتیب، حساسیت مدل نسبت به تغییرات جزئی کاهش یافته و پیش‌بینی نهایی پایدارتر و قابل‌اعتمادتر خواهد بود.

به‌صورت رسمی، اگر $A = \{A_1, A_2, \dots, A_n\}$ مجموعه‌ای از تبدیلات مجاز زمان آزمون باشند که روی تصویر ورودی x اعمال می‌شوند. برای هر تبدیل A_i، ابتدا تصویر تبدیل‌یافته $A_i(x)$ به مدل داده شده و سپس خروجی به کمک تبدیل معکوس A_i^{-1} به دستگاه مختصات تصویر اصلی بازگردانده می‌شود. خروجی حاصل از هر تبدیل به‌صورت رابطه (۱) می‌شود:

$$\hat{y}_i = A_i^{-1} f\big(A_i(x)\big) \qquad (۱)$$

به عنوان مثال و برای روشن شدن مفهوم تبدیلات به‌عنوان نمونه می‌توان $n = 3$ در نظر گرفت و تبدیلات را به‌صورت زیر تعریف کرد:

$$A_1(x) = x$$
$$A_2(x) = HFlip(x)$$
$$A_3(x) = VFlip(x)$$

که در آن x تصویر بدون فلیپ، $HFlip(x)$ فلیپ افقی و $VFlip(x)$ فلیپ عمودی است.

در این حالت، مدل $f(.)$ روی هر یک از تصاویر $A_3(x)$،$A_2(x)$،$A_1(x)$ اجرا می‌شود و خروجی‌ها پس از بازگردانی به مختصات تصویر اصلی (A_i^{-1}) تجمیع می‌گردند.

در نهایت، خروجی نهایی حاصل از تقویت زمان آزمون با میانگین‌گیری از خروجی‌های مربوط به تمامی تبدیلات محاسبه می‌شود:

$$\hat{y}_{TTA}(x) = \frac{1}{n}\sum_{i=1}^{n} \hat{y}_i \qquad (۲)$$

شایان ذکر است که این فرمول‌بندی را می‌توان به‌عنوان یک حالت خاص از چارچوب نظری ارائه‌شده توسط ونگ و همکاران (۲۰۲۱) [۱] در نظر گرفت. در آن مرجع، تقویت زمان آزمون در قالب یک مدل اخذ تصویر فرموله می‌شود که شامل مؤلفه‌ی نویز افزوده‌شده به تصویر و تبدیلات هندسی تصادفی است. در این چارچوب، هر نمونه‌ی مونت‌کارلو به‌صورت رابطه (۳) تعریف شده است:

$$y_n = \tau_{\beta_n}\left(f\left(\theta, {\tau_{\beta_n}}^{-1}(X - e_n)\right)\right) \qquad (۳)$$

در این رابطه، e_n به‌طور صریح نشان‌دهنده‌ی مؤلفه‌ی نویز افزوده‌شده به تصویر در فرآیند اخذ تصویر است. سپس پیش‌بینی نهایی با میانگین‌گیری از خروجی نمونه‌های مونت‌کارلو محاسبه می‌شود:

$$\hat{Y} = E(Y \mid X) \approx \frac{1}{N}\sum_{n=1}^{N} y_n \qquad (۴)$$

در پژوهش حاضر، با صرف‌نظر از مؤلفه‌ی نویز ($e_n = 0$) و با جایگزینی نمونه‌گیری پیوسته از پارامترهای تبدیل β_n با مجموعه‌ای محدود از تبدیلات قطعی زمان آزمون $\{A_i\}$ فرمول‌بندی کلی ارائه‌شده توسط ونگ و همکاران [۱] به‌صورت مستقیم به رابطه (۱) این مقاله تقلیل می‌یابد. در این حالت، پیش‌بینی مدل ابتدا بر روی نسخه‌های تبدیل‌یافته‌ی تصویر انجام شده، سپس خروجی‌ها با اعمال تبدیل معکوس به مختصات تصویر اصلی بازگردانده می‌شوند و در نهایت از طریق میانگین‌گیری تجمیع می‌گردند.

۳-۲- حذف تصادفی زمان آزمون (TTD)

حذف تصادفی زمان آزمون (TTD) یکی از روش‌های رایج برای افزایش پایداری پیش‌بینی و برآورد عدم‌قطعیت مدل‌های یادگیری عمیق در مرحله‌ی استنتاج است. در این رویکرد، برخلاف حالت استاندارد که لایه‌های Dropout تنها در مرحله‌ی آموزش فعال هستند، این لایه‌ها در زمان آزمون نیز فعال نگه داشته می‌شوند. بدین ترتیب، با هر بار اجرای مدل روی یک ورودی ثابت، به دلیل اعمال ماسک‌های تصادفی Dropout، خروجی متفاوتی تولید می‌شود.

در TTD، به‌جای اتکا به یک اجرای منفرد از مدل، شبکه چندین بار با Dropout فعال روی همان تصویر ورودی اجرا می‌شود و هر اجرا یک نمونه‌ی تصادفی از توزیع پیش‌بینی مدل را فراهم می‌کند. سپس پیش‌بینی نهایی با تجمیع این خروجی‌ها، معمولاً به‌صورت میانگین‌گیری، محاسبه می‌شود. این فرایند موجب کاهش نوسانات تصادفی و افزایش پایداری پیش‌بینی نهایی می‌گردد.

به‌صورت رسمی، اگر T تعداد اجرای تصادفی مدل با Dropout فعال در زمان آزمون باشد و خروجی مدل در اجرای t-ام با (x) نمایش داده شود، پیش‌بینی نهایی حاصل از حذف تصادفی زمان آزمون به‌صورت رابطه(۵) محاسبه می‌شود:

$$\hat{y}_{TTD}(x) = \frac{1}{T}\sum_{t=1}^{T}\hat{y}^{(t)}(x) \qquad (۵)$$

این فرمول‌بندی مبتنی بر چارچوب نظری ارائه‌شده توسط گال و قهرمانی (۲۰۱۶) [۲] است. در آن مرجع، نشان داده می‌شود که فعال‌سازی Dropout در زمان آزمون معادل یک تقریب بیزی از استنتاج مدل بوده و امیدریاضی توزیع پیش‌بینی مدل را می‌توان با میانگین‌گیری از چندین اجرای تصادفی شبکه برآورد کرد. به‌طور مشخص، این رابطه در معادله (۶)، مقاله‌ی گال و قهرمانی (۲۰۱۶) [۲] به‌صورت رابطه (۶) بیان شده است:

$$E(y^* \mid x^*) \approx \frac{1}{T}\sum_{t=1}^{T}\widehat{y^*}(x^*, W_t) \qquad (۶)$$

که در آن W_t نشان‌دهنده‌ی تحقق تصادفی وزن‌های شبکه ناشی از اعمال ماسک‌های Dropout در اجرای t-ام است. در عمل، این فرمول معادل اجرای مکرر شبکه با فعال بودن Dropout در مرحله‌ی آزمون و برآورد پیش‌بینی نهایی از طریق میانگین‌گیری خروجی‌های حاصل از این اجراهای تصادفی است.

به‌عنوان مثال، در روش حذف تصادفی زمان آزمون (TTD)، برای یک تصویر ورودی ثابت، مدل چندین بار با فعال بودن لایه‌های Dropout اجرا می‌شود. به دلیل اعمال ماسک‌های تصادفی متفاوت در هر اجرا، خروجی‌های حاصل اندکی با یکدیگر متفاوت خواهند بود. خروجی نهایی مطابق رابطه (۳) با میانگین‌گیری از این پیش‌بینی‌ها محاسبه می‌شود که منجر به کاهش نوسانات تصادفی و افزایش پایداری پیش‌بینی می‌گردد.

۳-۳- تقویت زمان آزمون چندمقیاسی (MSTTA)

تقویت زمان آزمون چندمقیاسی (MSTTA)[۶] در واقع بسط‌یافته‌ی روش تقویت زمان آزمون (TTA) است. همان‌طور که در بخش TTA اشاره شد، مجموعه‌ی تبدیلات زمان آزمون می‌تواند شامل فلیپ، چرخش، برش و تغییر مقیاس باشد. در MSTTA، به‌طور خاص، مؤلفه‌ی «تغییر مقیاس» به‌صورت صریح و نظام‌مند در قالب چند مقیاس گسسته از پیش تعیین‌شده اعمال می‌شود و برای هر مقیاس نیز همان تبدیلات TTA (مانند فلیپ‌ها) اجرا می‌گردد. سپس تمام پیش‌بینی‌های حاصل، پس از بازگردانی به اندازه و مختصات تصویر اصلی، به‌صورت میانگین‌گیری تجمیع می‌شوند.

به‌صورت رسمی، فرض کنید $S = \{s_1, s_2, \dots, s_k\}$ مجموعه‌ی مقیاس‌های زمان آزمون باشد و R_k عملگر تغییر اندازه‌ی تصویر مطابق مقیاس s_k و ${R_k}^{-1}$ عملگر بازگردانی خروجی به اندازه‌ی اصلی تصویر باشد. همچنین $A = \{A_1, A_2, \dots, A_n\}$ مجموعه‌ی تبدیلات زمان آزمون (مانند فلیپ افقی/عمودی) باشد و A_i^{-1} تبدیل معکوس متناظر برای بازگرداندن خروجی به مختصات تصویر اصلی در نظر گرفته می‌شود. همچنین $f(.)$ مدل پایه (شبکه‌ی از پیش‌آموزش‌دیده) است که خروجی پیش‌بینی را تولید می‌کند.

در این صورت، خروجی مدل متناظر با تبدیل A_i در مقیاس s_k به‌صورت رابطه (۷) تعریف می‌شود:

$$\hat{y}_{i,k} = {R_k}^{-1}\left(A_i^{-1}\left(f\left(A_i(R_k(x))\right)\right)\right) \qquad (۷)$$

در نهایت، خروجی نهایی تقویت زمان آزمون چندمقیاسی با میانگین‌گیری از تمامی پیش‌بینی‌های حاصل از مقیاس‌ها و تبدیلات مختلف محاسبه می‌شود:

$$\hat{y}_{MSTTA}(x) = \frac{1}{nK}\sum_{k=1}^{K}\sum_{i=1}^{n}\hat{y}_{i,k} \qquad (۸)$$

به‌عنوان مثال، اگر $K = 3$ مقیاس $S = \{0.75,1,1.25\}$ و $n = 3$ تبدیل زمان آزمون در نظر گرفته شود، تبدیلات را می‌توان به‌صورت زیر تعریف کرد:

$$A_1(x) = x$$
$$A_2(x) = HFlip(x)$$
$$A_3(x) = VFlip(x)$$

که در آن x تصویر بدون فلیپ، $HFlip(x)$ فلیپ افقی و $VFlip(x)$ فلیپ عمودی را نشان می‌دهد. در این حالت، برای هر تصویر ورودی در مجموع $nK = 9$ پیش‌بینی مختلف به‌دست می‌آید. پس از بازگردانی تمامی خروجی‌ها به اندازه‌ی تصویر اصلی، خروجی نهایی MSTTA از طریق میانگین‌گیری این ۹ پیش‌بینی محاسبه می‌شود.

۴-۳- سازگاری در زمان آزمون

در سازگاری در زمان آزمون [۷]، مدل در مرحله‌ی استنتاج خود را با داده‌های جدید سازگار می‌کند. برخلاف TTA یا TTD، این رویکردها بخشی از میان‌لایه‌های شبکه را به‌صورت محدود و کنترل‌شده به‌روزرسانی می‌کنند. مطالعات مروری اخیر نشان داده‌اند که سازگاری در زمان آزمون یکی از مؤثرترین استراتژی‌ها برای مقابله با تغییر توزیع داده‌ها است و طیف گسترده‌ای از روش‌های

مبتنی بر اصلاح آمار نرمال‌سازی، کاهش آنتروپی و یا تطبیق ویژگی را شامل می‌شود [۴]. روش تنت[۸] [۵] نمونه‌ی شناخته‌شده‌ای است که تنها پارامترهای خطی (گاما و بتا) در لایه‌های نرمال‌سازی دسته‌ای که به آن‌ها پارامترهای Affine گفته می‌شود در مرحله‌ی آزمون به‌روزرسانی می‌شوند تا آنتروپی خروجی کاهش یابد. این روش بدون نیاز به داده‌ی برچسب‌دار، مدل را در محیط آزمون سازگار می‌سازد.

در سال‌های اخیر، نسخه‌های سبک‌تری نیز معرفی شده‌اند که با به‌روزرسانی بسیار محدود لایه‌های اولیه، پایداری مدل را در داده‌های خارج از توزیع افزایش می‌دهند. به‌عنوان نمونه، روش LTTA نشان داده است که می‌توان با کمترین هزینه‌ی محاسباتی و تنها با به‌روزرسانی لایه‌های ابتدایی، سازگاری مؤثری به‌دست آورد [۶].

چالش اصلی این رویکردها، حساسیت به نرخ یادگیری و اندازه دسته در مرحله آزمون است.

۳-۵- مدل‌سازی تجمعی خودکار در زمان آزمون

در رویکرد مدل‌سازی تجمعی خودکار در زمان آزمون[۹]، چندین پیش‌بینی از ورودی‌های کمی متفاوت یا نسخه‌های نویزی الگوبرداری می‌شود. خروجی‌ها نه‌تنها تجمیع می‌شوند، بلکه وزن‌دهی آن‌ها بر اساس میزان اعتماد مدل انجام می‌گیرد. این روش علاوه بر بهره‌گیری از تنوع ورودی، از وابستگی میان پیش‌بینی‌های متوالی نیز استفاده می‌کند و در نتیجه خروجی نهایی پایدارتر است.

به‌طور خاص، کویین و همکاران (۲۰۲۵) [۷] نشان دادند که این روش می‌تواند نوسان خروجی مدل را به‌طور تقریبی «تا حدود ۱۵٪» کاهش دهد و پایداری عملکرد را در شرایطی که داده‌ها از توزیع اصلی فاصله دارند، بهبود دهد.

۳-۶- پالایش هدایت‌شده با عدم قطعیت

در رویکرد پالایش هدایت‌شده با عدم قطعیت[۱۰]، مدل ابتدا یک پیش‌بینی اولیه همراه با نقشه‌ای از عدم‌قطعیت تولید می‌کند تا مشخص شود کدام بخش‌های تصویر با اطمینان کمتری تحلیل شده‌اند. سپس نواحی با عدم‌قطعیت بالا که معمولاً بیانگر پیش‌بینی‌های مبهم یا مرزهای نامطمئن هستند شناسایی شده و به‌صورت محلی پالایش می‌شوند. این پالایش می‌تواند از طریق بازپیش‌بینی نواحی پرریسک یا تنظیم ویژگی‌های میانی شبکه انجام شود.

مطالعات اخیر نیز کارایی این ایده را تأیید کرده‌اند. برای مثال، ونگ و همکاران (۲۰۲۳) [۸] استفاده از هم‌ترازی و یکنواخت‌سازی ویژگی‌ها نشان دادند که اطلاعات عدم‌قطعیت می‌تواند به مدل کمک کند تا با داده‌های جدید سازگارتر شود. همچنین، یو و همکاران (۲۰۲۳) [۹] با استفاده از استخراج نمونه‌های مشابه بدون‌برچسب و ترکیب آن‌ها با پیش‌بینی اولیه مدل، نشان دادند که پالایش هدایت‌شده توسط عدم‌قطعیت، خروجی نهایی را پایدارتر و قابل‌اعتمادتر می‌کند.

به‌طور کلی، نتایج این خط پژوهشی نشان می‌دهد که استفاده از عدم‌قطعیت به‌عنوان سیگنال هدایت‌کننده می‌تواند خطاهای موضعی را کاهش داده، مرزبندی‌ها را دقیق‌تر کند و توان مدل را در مواجهه با داده‌های متغیر یا خارج از توزیع افزایش دهد.

۳-۷- انطباق مبتنی بر حافظه برای نرمال‌سازی دسته‌ای (MemBN)

انطباق مبتنی بر حافظه برای نرمال‌سازی دسته‌ای (MemBN)[۱۱] به ان مفهوم است که در شرایطی مانند اندازه‌ی دسته‌ی بسیار کوچک، نرمال‌سازی دسته‌ای دچار ناپایداری می‌شود. روش MemBN [۱۰] برای رفع این مشکل، برای هر لایه‌ی نرمال‌سازی دسته‌ای یک «حافظه آماری» ایجاد می‌کند که میانگین و واریانس چند دسته‌ی اخیر را نگه می‌دارد. هنگام پردازش ورودی جدید، این مقادیر با آمار دسته‌ی فعلی ترکیب می‌شوند و برآوردی پایدارتر از میانگین و واریانس نرمال‌سازی دسته‌ای حاصل می‌شود. نتایج تجربی نشان داده‌اند که MemBN دقت و ثبات مدل را در داده‌های جریانی و دسته‌های کوچک افزایش می‌دهد.

۳-۸- انطباق مبتنی بر پرامپت بصری (VPA)

در روش انطباق مبتنی بر پرامپت بصری (VPA)[۱۲] [۱۱] وزن‌های اصلی مدل ثابت نگه داشته می‌شوند و تنها یک «پرامپت بصری» کوچک و قابل‌آموزش به ورودی افزوده می‌شود. این پرامپت جریان ویژگی‌ها و میانگین و واریانس نرمال‌سازی دسته‌ای را با داده‌های جدید هماهنگ می‌کند. روش‌های اخیر نشان داده‌اند که با تعداد بسیار کمی پارامتر اضافی می‌توان مدل را در زمان آزمون تطبیق داد، بدون آن‌که خطر فراموشی دانش یا هزینه‌ی زیاد ایجاد شود.

۳-۹- بهینه‌سازی سبک در زمان آزمون

در رویکرد بهینه‌سازی سبک در زمان آزمون[۱۳] تنها بخش کوچکی از پارامترهای مدل (معمولاً لایه‌های نرمال‌سازی دسته‌ای) با چند گام بهینه‌سازی کم‌هزینه به‌روز می‌شوند. روش تنت از این دسته است و با کمینه‌سازی آنتروپی خروجی مدل را وادار می‌کند

تصمیمات مطمئن‌تری بگیرد. این روش محاسبات اندکی دارد و در مواجهه با داده‌های نویزی یا خارج از توزیع مؤثر است.

۴- روش پیشنهادی

در این پژوهش، با هدف بهبود دقت و پایداری پیش‌بینی در مدل EGE-UNet [۳]، تمرکز صرفاً بر تقویت مرحله استنتاج قرار داده شده است؛ به‌طوری‌که هیچ تغییری در معماری مدل پایه و هیچ‌گونه بازآموزی انجام نمی‌شود. ایده اصلی آن است که با بهره‌گیری از دو رویکرد، TTA و TTD، رفتار مدل در زمان آزمون پایدارتر و مقاوم‌تر گردد.

در این چارچوب، روش‌های TTA و TTD و همچنین نسخه چندمقیاسی ترکیب‌شده آن‌ها به عنوان یک لایه افزوده استنتاجی بر خروجی مدل پیش‌آموزش‌دیده اعمال می‌شوند. این روش‌ها با ایجاد نسخه‌های افزوده‌شده یا مقیاس‌یافته از تصویر ورودی و اجرای چندباره مدل با Dropout فعال، مجموعه‌ای متنوع از پیش‌بینی‌ها تولید می‌کنند که با ادغام آن‌ها، یک برآورد نهایی دقیق‌تر، پایدارتر و کم‌نوسان‌تر حاصل می‌شود.

این طراحی امکان می‌دهد که عملکرد مدل در مواجهه با داده‌های واقعی و ناهمگون بهبود یابد، بدون آن‌که هزینه آموزش مجدد، تنظیم پارامترها یا تغییر در ساختار شبکه لازم باشد. دو نسخه پیشنهادی این رویکرد، شامل ترکیب TTA+TTD و نسخه چندمقیاسی آن، در ادامه به‌طور کامل تشریح می‌شوند.

۴-۱- ترکیب TTA و TTD

در نسخه‌ی نخست روش پیشنهادی، تقویت زمان آزمون (TTA) و حذف تصادفی زمان آزمون (TTD) به‌صورت هم‌زمان مورد استفاده قرار می‌گیرند. TTA حساسیت مدل نسبت به تبدیلات هندسی را کاهش می‌دهد، در حالی که TTD با میانگین‌گیری روی چند اجرای تصادفی، نوسانات ناشی از Dropout را تعدیل می‌کند. در این پژوهش، این دو رویکرد در قالب یک چارچوب مشترک در مرحله‌ی استنتاج ترکیب شده‌اند.

برای هر تصویر ورودی x، ابتدا مجموعه‌ای از تبدیلات زمان آزمون $A = \{A_1, A_2, \dots, A_n\}$ شامل تبدیل بدون فلیپ، فلیپ افقی و فلیپ عمودی اعمال می‌شود و چند نسخه‌ی تقویت‌شده از تصویر به‌دست می‌آید. سپس هر نسخه‌ی تقویت‌شده T بار با Dropout فعال از مدل عبور داده می‌شود و مجموعه‌ای از پیش‌بینی‌های تصادفی تولید می‌گردد.

برای هر تبدیل ثابت A_i ابتدا خروجی‌های حاصل از T اجرای تصادفی میانگین‌گیری می‌شوند (مرحله‌ی TTD). سپس تبدیل معکوس A_i^{-1} اعمال می‌شود تا خروجی‌ها به مختصات تصویر اصلی بازگردند. در نهایت، خروجی‌های حاصل از تبدیلات مختلف با یکدیگر ادغام می‌شوند (مرحله‌ی TTA) و پیش‌بینی نهایی به‌دست می‌آید.

در رابطه (۹)، ابتدا برای هر تبدیل A_i خروجی مدل روی T اجرای تصادفی با Dropout فعال میانگین‌گیری می‌شود . سپس با اعمال تبدیل معکوس A_i^{-1}، خروجی‌ها به مختصات تصویر اصلی بازگردانده شده و در نهایت نتایج مربوط به همه‌ی تبدیلات با میانگین‌گیری تجمیع می‌شوند. این فرمول‌بندی، چارچوب ترکیبی پیشنهادی این پژوهش برای استنتاج پایدارتر بدون بازآموزی مدل پایه است.

به‌صورت رسمی، خروجی ترکیبی TTA+TTD به‌صورت زیر تعریف می‌شود:

$$y_{TTA+TTD} = \frac{1}{n}\sum_{i=1}^{n} A_i^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\hat{y}^{(t)}\left(A_i(x)\right)\right) \qquad (۹)$$

این فرمول‌بندی موجب می‌شود مدل هم نسبت به تبدیلات هندسی ورودی مقاوم‌تر شود و هم از طریق میانگین‌گیری روی مسیرهای تصادفی ناشی از Dropout، پیش‌بینی‌ای پایدارتر و قابل‌اعتمادتر ارائه دهد؛ بدون آن‌که تغییری در ساختار مدل پایه ایجاد شود.

۴-۲- ترکیب TTD و TTA چندمقیاسی

نسخه‌ی دوم روش پیشنهادی، بسطی از ترکیب تقویت زمان آزمون و حذف تصادفی زمان آزمون (TTA+TTD) است که در آن، علاوه بر اعمال تبدیلات هندسی بر روی تصویر ورودی، از چندین مقیاس مختلف نیز در مرحله‌ی آزمون استفاده می‌شود. به‌عبارت دیگر، همان چارچوب ترکیب TTA و TTD این‌بار بر روی نسخه‌های تغییراندازه‌یافته‌ی تصویر در مقیاس‌های گوناگون اعمال می‌گردد تا حساسیت مدل نسبت به تغییرات مقیاس کاهش یابد.

در این روش، هر تصویر ورودی ابتدا در مجموعه‌ای از مقیاس‌های از پیش تعیین‌شده تغییر اندازه داده می‌شود (در این پژوهش: ۰/۷۵، ۱ و ۱/۲۵ برابر اندازه‌ی اصلی) . سپس برای هر مقیاس، فرایند ترکیب TTA و TTD به‌طور مستقل اجرا می‌شود؛ بدین معنا که ابتدا بر روی نسخه‌ی مقیاس‌یافته‌ی تصویر، تبدیلات هندسی مشخصی مانند فلیپ افقی و عمودی اعمال می‌گردد و برای هر نسخه، مدل با Dropout فعال در زمان آزمون چندین بار اجرا می‌شود تا مجموعه‌ای از پیش‌بینی‌های تصادفی حاصل شود.

در هر مقیاس، خروجی‌های حاصل ابتدا روی اجراهای تصادفی

میانگین‌گیری می‌شوند (TTD)، سپس با اعمال تبدیلات معکوس و بازگردانی مقیاس به اندازه‌ی تصویر اصلی منتقل شده و در نهایت با میانگین‌گیری روی همه‌ی تبدیلات و مقیاس‌ها تجمیع می‌گردند. بنابراین، رابطه (۱۰) بسط مستقیم رابطه قبل به حالت چندمقیاسی است.

اگر $S = \{s_1, s_2, \dots, s_k\}$ مجموعه‌ی مقیاس‌ها، R_k عملگر تغییر اندازه در مقیاس s_k و R_k^{-1} عملگر بازگردانی به اندازه‌ی اصلی تصویر باشد. خروجی ترکیبی MSTTA+TTD به‌صورت رابطه(۱۰) تعریف می‌شود:

$$y_{MSTTA+TTD} = \frac{1}{nK}\sum_{i=1}^{n}\sum_{k=1}^{K} R_k^{-1}\left[A_i^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\hat{y}^{(t)}(A_i(R_k(x)))\right)\right] \quad (۱۰)$$

در این رابطه، A_i تبدیلات تقویت (مانند فلیپ‌ها)، $\hat{y}^{(t)}$ خروجی مدل در اجرای t-ام با Dropout فعال در زمان آزمون است.

این طراحی موجب می‌شود مدل، علاوه بر بهره‌مندی از تنوع ناشی از Dropout و تبدیلات هندسی، نسبت به تغییرات اندازه‌ی اشیاء در تصویر نیز مقاوم‌تر گردد و بتواند هم ساختارهای کلی و هم جزئیات ریز را با دقت بیشتری شناسایی کند.

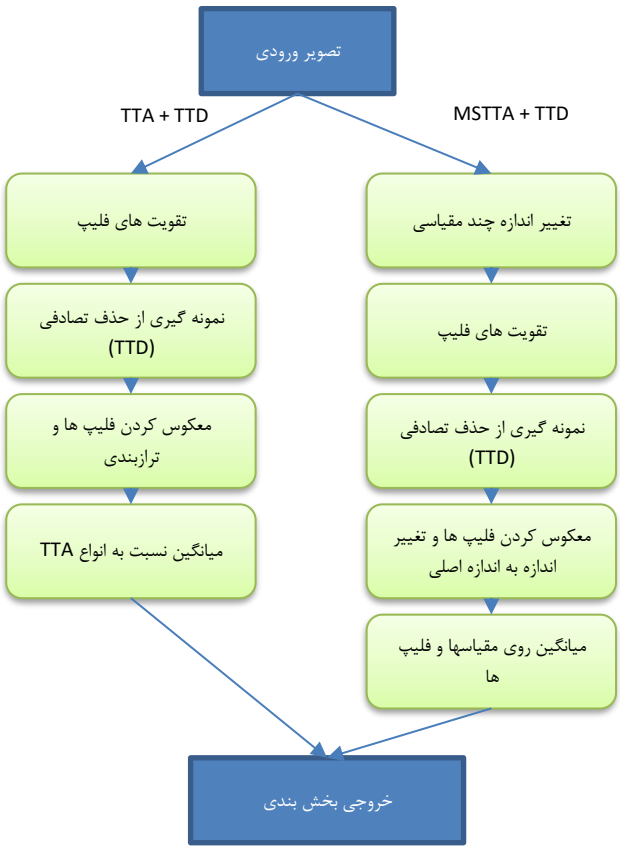
به‌عبارت دیگر، ترکیب MSTTA و TTD نوعی مدل‌سازی تجمعی در زمان آزمون ایجاد می‌کند که به‌طور هم‌زمان از داده‌های متنوع در چند مقیاس و از قدرت کاهش عدم‌قطعیت از طریق میانگین‌گیری چندگانه بهره می‌برد. نتایج تجربی نشان می‌دهند که این روش، با هزینه‌ی محاسباتی نسبتاً اندک و بدون نیاز به تغییر در ساختار مدل پایه، دقت و پایداری پیش‌بینی‌های EGE-UNet را بهبود داده و قابلیت تعمیم به سایر معماری‌های بخش‌بندی تصویر را نیز داراست.

در این پژوهش، دو راهکار تقویتی برای بهبود عملکرد مدل در مرحله‌ی آزمون به‌کار گرفته شده است:

(۱) ترکیب (TTA+TTD) و (۲) نسخه‌ی چندمقیاسی همین ترکیب (MSTTA+TTD).

برای جلوگیری از تکرار و همچنین ارائه‌ی یک دید یکپارچه از فرآیند استنتاج، مراحل هر دو روش در قالب یک نمودار بلوکی مشترک در شکل ۱ نمایش داده شده‌اند.

در این ساختار، ابتدا تصویر ورودی در دو مسیر موازی پردازش می‌شود.



شکل ۱: نمای کلی فرآیند استنتاج پیشنهادی شامل دو مسیر TTA+TTD و MSTTA+TTD. در مسیر TTA+TTD، تقویت‌های فلیپ و حذف تصادفی در زمان آزمون اعمال شده و پس از معکوس‌سازی تبدیلات، خروجی‌ها تجمیع می‌شوند. در مسیر MSTTA+TTD، علاوه بر این مراحل، پردازش چندمقیاسی (۰٫۷۵×، ۱× و ۱٫۲۵×) انجام شده و نتایج پس از بازگردانی به اندازه‌ی اصلی میانگین‌گیری می‌گردند تا خروجی نهایی بخش‌بندی تولید شود.

در مسیر نخست (TTA+TTD)، تصویر در سه حالت مختلفِ بدون فلیپ، فلیپ افقی و فلیپ عمودی مورد استفاده قرار می‌گیرد. هر نسخه‌ی افزوده‌شده چندین بار با Dropout فعال از مدل عبور داده می‌شود تا نمونه‌های تصادفی متعددی تولید شود. خروجی‌های به‌دست‌آمده پس از اعمال فلیپ معکوس، با یکدیگر میانگین‌گیری شده و یک پیش‌بینی پایدار حاصل می‌شود.

در مسیر دوم (MSTTA+TTD)، ابتدا تصویر در سه مقیاس مختلف (۰٫۷۵×، ۱٫۰×، ۱٫۲۵×) تغییر اندازه می‌یابد و سپس همان چرخه‌ی افزوده‌سازی، اجرای چندباره‌ی مدل با Dropout و بازگردانی فلیپ‌ها و اندازه به شکل اولیه تکرار می‌شود. نهایتاً خروجی‌های مربوط به همه‌ی مقیاس‌ها و فلیپ‌ها با هم ادغام می‌شوند.

ترکیب این دو مسیر باعث می‌شود مدل در برابر انواع تغییرات فضایی، مقیاسی و نوسانات ناشی از Dropout مقاوم‌تر شده و در

مرحله‌ی آزمون پیش‌بینی‌های پایدارتر و قابل‌اعتمادتری ارائه دهد. نمودار زیر، جریان کامل این فرآیند را به‌صورت یکپارچه نشان می‌دهد.

۵- ارزیابی و نتایج

روش پیشنهادی ما بر روی دو مجموعه داده‌ی عمومی و معتبر بخش‌بندی ضایعات پوستی، یعنی [۱۲] و [۱۳] ISIC2017 و ISIC2018 [۱۴] و [۱۵]، مورد ارزیابی قرار گرفت.

مجموعه داده‌های ISIC2017 و ISIC2018 شامل به ترتیب ۲۱۵۰ و ۲۶۹۴ تصویر درموسکوپی است که هر تصویر به‌همراه یک ماسک بخش‌بندی دودویی دقیق ارائه شده است. ما مجموعه داده‌ها را به طور تصادفی به مجموعه‌های آموزش و آزمون با نسبت ۷:۳ تقسیم می‌کنیم. تمام آزمایش‌ها بر روی یک پردازنده گرافیکی NVIDIA RTX 4050 انجام شده است.

پیش از آموزش، تمامی تصاویر با رزولوشن ثابت ۲۵۶×۲۵۶ پیکسل بازنمونه‌برداری شده‌اند. تمامی مدل‌ها در چارچوب PyTorch [۱۶] پیاده‌سازی شده و مدل پایه EGE-UNet با بهینه‌ساز AdamW [۱۷] و نرخ یادگیری اولیه ۰/۰۰۱ آموزش داده شد. در طول ۳۰۰ دوره آموزش، برای ارزیابی روش پیشنهادی، از معیارهای میانگین تقاطع بر اتحاد (mIoU) و امتیاز شباهت دایس (DSC) استفاده شد.

آموزش مدل‌ها با اندازه دسته برابر ۸ انجام شد. همچنین، برای بهبود قابلیت تعمیم‌پذیری مدل، روش‌های افزایش داده شامل چرخش تصادفی (افقی، عمودی و زاویه‌ای) به کار گرفته شدند.

۵-۱- نتایج کمی

نتایج کمی نشان می‌دهد که به‌کارگیری TTA و TTD نسبت به مدل پایه EGE-UNet در مجموعه‌داده ISIC2017 در جدول ۱ باعث بهبود ٪۰/۳ در DSC و ٪۰/۱ در mIoU و در مجموعه‌داده ISIC2018 در جدول ۲ منجر به افزایش ٪۰/۴۲ در DSC و ٪۰/۳ در mIoU شده است.

با این حال افزودن چند مقیاس متفاوت ۰/۷۵، ۱ و ۱/۲۵ برابر اندازه اصلی در TTA (MSTTA+TTD) در مجموعه‌داده ISIC2017 حتی نسبت به ترکیب اول نیز بهتر عمل کرده و نسبت به این ترکیب بهبود ٪۰/۶ در DSC و ٪۱/۱ در mIoU را نشان داده است و نسبت به مدل پایه بهبود ٪۰/۸ در DSC و ٪۱/۴ در mIoU را نشان داده است؛ با این حال، در ISIC2018 این روش با افت نسبی عملکرد همراه بوده و نسبت به مدل پایه کاهش ٪۰/۶ در DSC و ٪۱ در mIoUرا و نسبت به ترکیب اول کاهش ٪۰/۸۹ در DSC و ٪۱/۵ در mIoU تجربه کرده است.

اگرچه این تغییرات درصدی در برخی موارد کوچک به نظر می‌رسند، اما از نظر بالینی معنادار هستند. برخلاف رویکردهای موجود که بر تغییرات معماری یا طراحی توابع هزینه پیچیده متکی هستند، روش‌های پیشنهادی کاملاً در مرحله استنتاج اعمال شده و هیچ نیازی به بازآموزی مدل ندارند؛ بنابراین می‌توان آن‌ها را بدون هزینه محاسباتی آموزش، به مدل‌های از پیش آموزش‌دیده اضافه کرد.

جدول ۱: مقایسه‌ی عملکرد روش‌های مختلف بخش‌بندی ضایعه پوستی روی مجموعه‌داده ISIC2017 بر اساس معیارهای mIoU و DSC. نتایج شامل مدل‌های مرجع و همچنین مدل پایه EGE-UNet و نسخه‌های تقویت‌شده‌ی آن با راهبردهای TTA+TTD و MSTTA+TTD بوده و بهترین مقادیر هر معیار مشخص شده‌اند.

مدل	mIoU(%)	DSC(%)
UNet [18]	76.98	86.99
UTNetV2 [19]	77.35	87.23
TransFuse [20]	79.21	88.40
MobileViTv2 [21]	78.72	88.09
MobileNetv3 [22]	77.69	87.44
UNeXtS [23]	78.26	87.80
MALUNet [24]	78.78	88.13
EGE-UNet [3]	79.81	88.77
EGE-UNet+TTA+TTD	**80.05**	**88.92**
EGE-UNet+MSTTA+TTD	**80.99**	**89.49**

جدول ۲: مقایسه‌ی عملکرد روش‌های مختلف بخش‌بندی ضایعه پوستی روی مجموعه‌داده ISIC2018 بر اساس معیارهای mIoU و DSC. نتایج شامل مدل‌های مرجع و همچنین مدل پایه EGE-UNet و نسخه‌های تقویت‌شده‌ی آن با راهبردهای TTA+TTD و MSTTA+TTD بوده و بهترین مقادیر هر معیار مشخص شده‌اند.

مدل	mIoU(%)	DSC(%)
UNet [18]	77.86	87.55
UNet++ [25]	78.31	87.83
AttUNet [26]	78.43	87.91
UTNetV2[19]	78.97	88.25
SANet [27]	79.52	88.59
TransFuse [20]	80.63	89.27
MobileViTv2 [21]	79.88	88.81
MobileNetv3 [22]	78.55	87.98
UNeXtS [23]	79.09	88.33
MALUNet [24]	80.25	89.04
EGE-UNet [3]	80.94	89.46
EGE-UNet+TTA+TTD	**81.36**	**89.72**
EGE-UNet+MSTTA+TTD	**80.06**	**88.92**

افزون بر این، نتایج نشان می‌دهد که حتی پیشرفته‌ترین معماری‌ها نیز می‌توانند از مدل‌سازی عدم قطعیت در زمان آزمون و رفتاری مشابه مدل‌سازی تجمعی که از طریق TTA و TTD ایجاد می‌شود، بهره‌مند شوند. در عمل، چنین بهبودهای سبک و کم‌هزینه‌ای می‌توانند نقش مهمی در افزایش قابلیت اطمینان سیستم‌های تصویربرداری پزشکی در شرایط واقعی، به‌ویژه در محیط‌های بالینی با منابع محدود، ایفا کنند.

برای ارزیابی عملکرد مدل‌های مختلف در بخش‌بندی ضایعات پوستی، دو معیار mIoU و DSC بر روی مجموعه داده‌های ISIC2017 و ISIC2018 محاسبه شد.

نمودارهای میله‌ای در شکل ۲ و ۳، که برای مقایسه بصری و روشن‌تر مقادیر از جدول به نمودار تبدیل شده‌اند، مقادیر این دو معیار را برای مدل‌های پیشین و دو روش ترکیبی پیشنهادی، به‌ترتیب بر روی داده‌های ISIC2017 و ISIC2018، نشان می‌دهند.

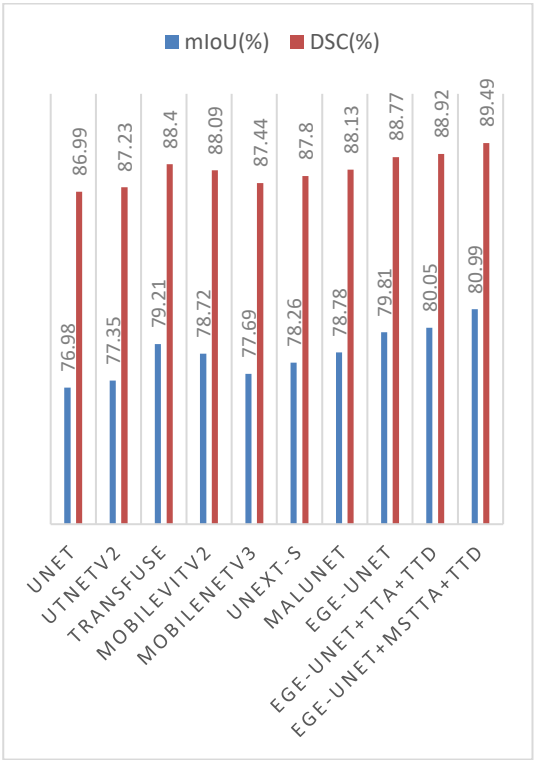
به منظور بررسی پایداری روش‌های ترکیبی، آزمایش مشابهی بر روی مجموعه داده ISIC2018 انجام شد. نتایج mIoU و DSC در نمودارهای زیر برای هر مدل قابل مشاهده است و عملکرد پایدار روش‌های ترکیبی را در مجموعه داده‌های مختلف تأیید می‌کند.

۵-۲- نتایج کیفی

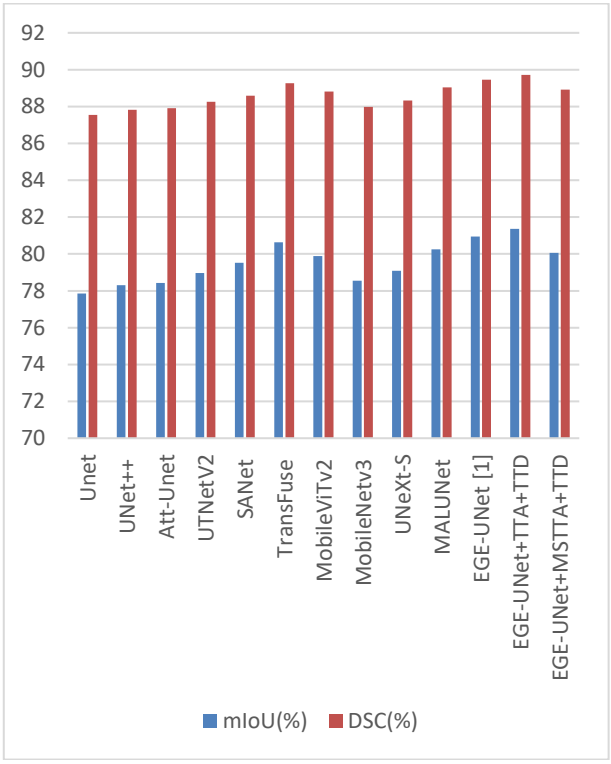
برای ارزیابی بصری روش‌های پیشنهادی، مقایسه کیفی بر روی دو مجموعه‌داده ISIC2017 و ISIC2018 انجام شد. در هر یک از تصاویر شکل ۴ و ۵، هر ردیف شامل شش ستون است که به‌ترتیب عبارت‌اند از:

۱) تصویر اصلی درموسکوپی
۲) ماسک مرجع[۱۴]
۳) خروجی ماسک مدل پایه EGE-UNet
۴) خروجی ماسک ترکیب TTA و TTD
۵) خروجی ماسک ترکیب MSTTA و TTD

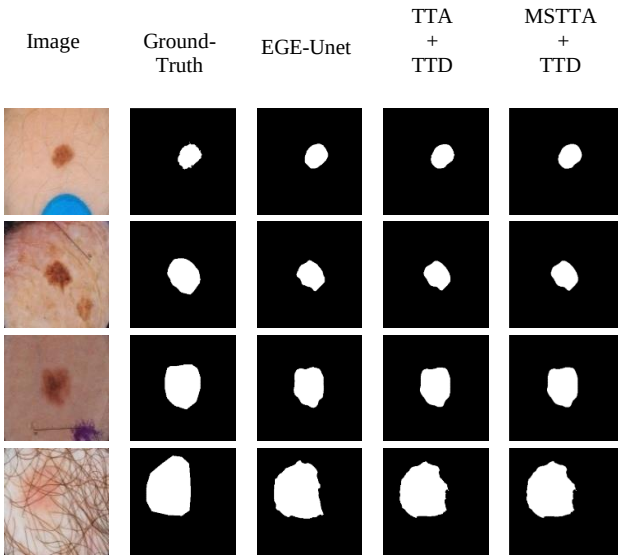
همان‌طور که مشاهده می‌شود، ترکیب اول (TTA+TTD) در هر دو مجموعه‌داده ISIC2017 و ISIC2018 عملکرد بهتری نسبت به مدل پایه داشته و مرزهای ضایعه را دقیق‌تر و پایدارتر بازسازی کرده است. در حالی که ترکیب دوم (MSTTA+TTD) در مجموعه‌داده ISIC2017بهبود بیشتری به همراه داشته و توانسته جزئیات ظریف و ضایعات کوچک را بهتر حفظ کند، اما در مجموعه‌داده ISIC2018 باعث کاهش نسبی عملکرد نسبت به ترکیب اول شده است، که احتمالاً به تفاوت ویژگی‌های تصویری و تنوع ضایعات در این مجموعه‌داده بازمی‌گردد.



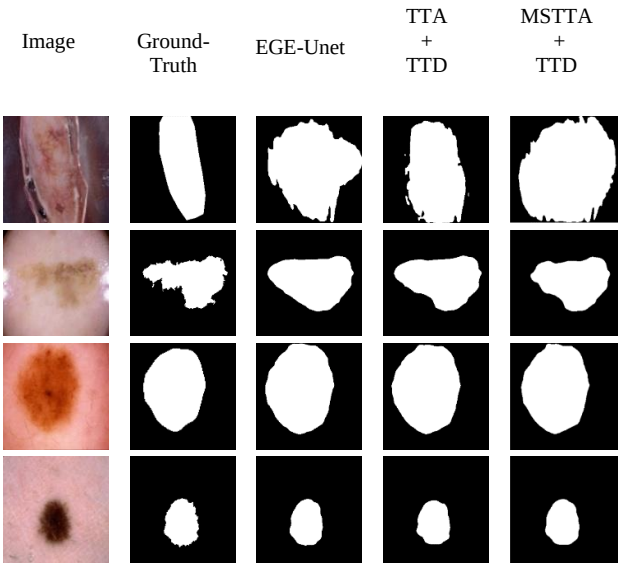
شکل ۲: مقایسه‌ی عملکرد روش‌های مختلف بخش‌بندی ضایعه پوستی روی مجموعه‌داده ISIC2017 بر اساس معیارهای mIoU و DSC. در این شکل، عملکرد مدل پایه EGE-UNet و نسخه‌های تقویت‌شده‌ی آن با راهبردهای TTA+TTD و MSTTA+TTD در کنار سایر مدل‌های مرجع نمایش داده شده است.



شکل ۳: مقایسه‌ی عملکرد روش‌های مختلف بخش‌بندی ضایعه پوستی روی مجموعه‌داده ISIC2018 بر اساس معیارهای mIoU و DSC. نتایج مربوط به مدل EGE-UNet و نسخه‌های مبتنی بر TTA+TTD و MSTTA+TTD در کنار سایر روش‌های مقایسه‌ای گزارش شده‌اند.



شکل ۴: نتایج کیفی بخش‌بندی ضایعه پوستی روی مجموعه‌داده ISIC2017. از چپ به راست: (۱) تصویر ورودی، (۲) ماسک مرجع (Ground-Truth)، (۳) خروجی مدل EGE-UNet، (۴) خروجی روش پیشنهادی TTA+TTD و (۵) خروجی روش پیشنهادی MSTTA+TTD. نواحی روشن، ضایعه‌ی پیش‌بینی‌شده را نشان می‌دهند.



شکل ۵: نتایج کیفی بخش‌بندی ضایعه پوستی روی مجموعه‌داده ISIC2018. از چپ به راست: (۱) تصویر ورودی، (۲) ماسک مرجع (Ground-Truth)، (۳) خروجی مدل EGE-UNet، (۴) خروجی روش پیشنهادی TTA+TTD و (۵) خروجی روش پیشنهادی MSTTA+TTD. نواحی روشن، ضایعه‌ی پیش‌بینی‌شده را نشان می‌دهند.

این نتایج کیفی، که در شرایط چالش‌برانگیز شامل اشکال نامنظم ضایعه، مرزهای محو و عوامل مزاحم مانند مو و نیز پایدار بوده‌اند، با یافته‌های کمی گزارش‌شده در بخش‌های قبل همخوانی دارد و نقش مثبت استراتژی‌های پیشنهادی را در بهبود دقت و پایداری بخش‌بندی ضایعات پوستی تأیید می‌کند.

از نظر هزینه محاسباتی، روش پیشنهادی (MSTTA+TTD) نسبت به روش پیشنهادی (TTA+TTD) به‌طور متوسط زمان استنتاج را حدود سه برابر افزایش داد. با این حال، در کاربردهای بالینی که دقت بالاتر اولویت دارد، این افزایش زمان قابل توجیه است. همچنین هزینه محاسباتی روش‌های پیشنهادی نسبت به روش پایه از آنجا که روش پیشنهادی تنها در مرحله استنتاج اعمال می‌شود، افزایش هزینه محاسباتی صرفاً ناشی از اجرای چندباره مدل است و هیچ‌گونه هزینه آموزشی جدیدی تحمیل نمی‌کند. این افزایش خطی و قابل کنترل بوده و با توجه به کاربردهای حساس پزشکی، حتی بهبودهای کمتر از یک درصد نیز توجیه‌پذیر هستند.

۶- نتیجه‌گیری و کارهای آینده

در این پژوهش، رویکردی برای ارتقاء عملکرد مدل EGE-UNet در بخش‌بندی ضایعات پوستی معرفی شد. این رویکرد بر پایه ترکیب دو استراتژی ساده ولی مؤثر، یعنی TTA و TTD، بنا شده است و توانست دقت و پایداری پیش‌بینی‌ها را در مرحله استنتاج افزایش دهد. افزون بر این، نسخه‌ای توسعه‌یافته از روش پیشنهادی نیز مورد بررسی قرار گرفت که در آن، رویکرد چندمقیاسی به ترکیب TTA و TTD افزوده شد. نتایج نشان داد که این نسخه در مجموعه‌داده ISIC2017موجب بهبود بیشتر عملکرد شد، در حالی که در مجموعه‌داده ISIC2018 با کاهش نسبی دقت همراه بود. این موضوع نشان می‌دهد که روش‌های چندمقیاسی می‌توانند نسبت به ماهیت داده‌ها حساس باشند و نیازمند طراحی تطبیقی دقیق‌تری هستند.

مزیت اصلی رویکرد پیشنهادی در سادگی پیاده‌سازی و استقلال از ساختار معماری است؛ به‌طوری که بدون نیاز به بازآموزی مدل یا تغییر در اجزای آن، به سادگی می‌تواند به مدل‌های از پیش آموزش‌دیده افزوده شود. این ویژگی، ادغام روش در سیستم‌های موجود و کاربرد آن در محیط‌های بالینی را تسهیل می‌کند.

با وجود نتایج مثبت، محدودیت‌هایی نیز وجود دارد. آزمایش‌ها صرفاً بر روی دو مجموعه‌داده ISIC2017 و ISIC2018 انجام شد که اگرچه معیارهای شناخته‌شده‌ای در حوزه درموسکوپی هستند، اما به‌طور کامل نماینده‌ی چالش‌های واقعی نیستند. در همین راستا، انتشار مجموعه‌داده‌های جدیدتر مانند ISIC2024، که شامل تصاویر گسترده‌تر و متنوع‌تری است، فرصت مناسبی برای ارزیابی جامع‌تر این رویکرد فراهم می‌کند. همچنین، در صورت در دسترس بودن مجموعه‌های جدیدتر مانند ISIC2025، بررسی عملکرد روش بر روی آن‌ها می‌تواند دید دقیق‌تری از میزان تعمیم‌پذیری ارائه دهد.

برای کارهای آینده، چند مسیر پژوهشی قابل توجه است: نخست، طراحی الگوریتم‌های انتخاب مقیاس هوشمند که به جای استفاده از مقیاس‌های ثابت، متناسب با ویژگی‌های هر تصویر بهترین مقیاس را انتخاب کنند. دوم، ترکیب این رویکرد با روش‌های برآورد عدم قطعیت، تا علاوه بر بهبود دقت، میزان اطمینان پیش‌بینی‌ها نیز در اختیار متخصصان قرار گیرد. سوم، ارزیابی عملی روش در محیط‌های بالینی و سامانه‌های کمک‌تشخیص، به‌ویژه در شرایطی که منابع محاسباتی محدود است یا نیاز به پردازش سریع وجود دارد. در نهایت، گسترش این روش به حوزه‌های دیگر تصویربرداری پزشکی، مانند رادیولوژی یا پاتولوژی دیجیتال، می‌تواند گام مهمی برای تعمیم کاربرد آن باشد.

۷- دسترسی به داده‌ها و کدها

داده مجموعه‌هایی که از یافته‌های این مطالعه پشتیبانی می‌کنند، در پیوندهای زیر در دسترس هستند:

https://challenge.isicarchive.com/data/#2017
https://challenge.isicarchive.com/data/#2018

همچنین، این مجموعه داده‌ها را می‌توان از آدرس زیر نیز دانلود کرد:

https://drive.google.com/file/d/1oMzh0z-Ut2vuYpyqPcN_ql_SSgNdLX5t/view?usp=drive_link

کدهای این مقاله در آدرس زیر موجود است:

https://github.com/emadeddindata/IT_EGE-UNet

مراجع

[1] G. Wang, W. Li, *et al.*, “Aleatoric uncertainty estimation with test-time augmentation for medical image segmentation with convolutional neural networks,” *Neurocomputing*, vol. 429, pp. 66–73, 2021. DOI: 10.1016/j.neucom.2020.11.103

[2] Y. Gal and Z. Ghahramani, “Dropout as a Bayesian approximation: Representing model uncertainty in deep learning,” in *Proc. 33rd Int. Conf. Mach. Learn. (ICML)*, 2016, pp. 1050–1059. DOI: 10.48550/arXiv.1506.02142

[3] R. Cao, K. Ma, *et al.*, “EGE-UNet: An efficient group enhanced U-Net for skin lesion segmentation,” *Comput. Biol. Med.*, vol. 142, p. 105197, 2022. DOI: 10.1016/j.compbiomed.**2022.105197**

[4] J. Liang, R. He, and T. Tan, “A comprehensive survey on test-time adaptation under distribution shifts,” *Int. J. Comput. Vis.*, vol. 133, no. 1, pp. 31–64, 2024. DOI: **10.1007/s11263-023-01925-4**

[5] D. Wang, E. Shelhamer, S. Liu, B. Olshausen, and T. Darrell, “Tent: Fully test-time adaptation by entropy minimization,” in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2021. DOI: **10.48550/arXiv.2006.10726**

[6] X. Zhang, Y. Qin, J. Huang, S. Yang, and M. Xu, “LTTA: Lightweight test-time adaptation via updating early layers and uncertainty estimation,” in *Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2024. DOI: **10.48550/arXiv.2311.08252**

[7] S. Qin, W. Zhang, and Z. Li, “SelfEnsemble for test-time adaptation,” *Neurocomputing*, vol. 653, p. 131038, Nov. 2025. DOI: **10.1016/j.neucom.2025.131038**

[8] J. Wang, Y. Liu, C. Gong, and Y. Qiao, “Feature alignment and uniformity for test-time adaptation,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2023, pp. 19452–19461. DOI: **10.1109/CVPR52729.2023.01865**

[9] J. Yu, H. Liu, and S. Wang, “Train/test-time adaptation with retrieval,” *arXiv preprint*, 2023. DOI: **10.48550/arXiv.2303.14333**

[10] J. Kang, N. Kim, J. Ok, and S. Kwak, “MemBN: Robust test-time adaptation via batch norm with statistics memory,” in Proc. Eur. Conf. Comput. Vis. (ECCV), 2024. DOI: 10.1007/978-3-031-72664-4_6

[11] J. Sun, M. Ibrahim, M. Hall, I. Evtimov, Z. M. Mao, C. C. Ferrer, and C. Hazirbas, “VPA: Fully test-time visual prompt adaptation,” in Proc. 31st ACM Int. Conf. Multimedia (MM), 2023, pp. 5796–5806. DOI: 10.1145/3581783.3612293

[12] International Skin Imaging Collaboration, “ISIC 2017 challenge dataset.” [Online]. Available: https://challenge.isicarchive.com/data/#2017. DOI: N/A

[13] C. F. Codella et al., “Skin lesion analysis toward melanoma detection: A challenge at the 2017 ISBI,” in Proc. IEEE 15th Int. Symp. Biomed. Imaging (ISBI), Washington, DC, USA, 2018. DOI: 10.1109/ISBI.2018.8363547

[14] International Skin Imaging Collaboration, “ISIC 2018 challenge dataset.” [Online]. Available: https://challenge.isicarchive.com/data/#2018. DOI: N/A

[15] N. Codella et al., “Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the ISIC,” arXiv preprint, 2019. DOI: 10.48550/arXiv.1902.03368

[16] A. Paszke et al., “PyTorch: An imperative style, high-performance deep learning library,” in Adv. Neural Inf. Process. Syst., vol. 32, 2019. DOI: 10.48550/arXiv.1912.01703

[17] I. Loshchilov and F. Hutter, “Decoupled weight decay regularization,” in Proc. Int. Conf. Learn. Represent. (ICLR), 2019. DOI: 10.48550/arXiv.1711.05101

[18] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional networks for biomedical image segmentation,” in Proc. MICCAI, 2015, pp. 234–241. DOI: 10.1007/978-3-319-24574-4_28

[19] X. Gao, M. Zhou, D. Liu, and D. Metaxas, “A multiscale transformer for medical image segmentation,” arXiv preprint, 2022. DOI: 10.48550/arXiv.2203.00131

[20] Y. Zhang, H. Liu, and Q. Hu, “TransFuse: Fusing transformers and CNNs for medical image segmentation,” in Proc. MICCAI, 2021, pp. 14–24. DOI: 10.1007/978-3-030-87231-1_2

[21] S. Mehta and M. Rastegari, “Separable self-attention for mobile vision transformers,” arXiv preprint, 2022. DOI: 10.48550/arXiv.2206.02680

[22] A. Howard et al., “Searching for MobileNetV3,” in Proc. IEEE/CVF Int. Conf. *Comput. Vis. (ICCV)*, 2019, pp. 1314–1324. DOI: **10.1109/ICCV.2019.00140**

[23] J. M. J. Valanarasu and V. M. Patel, “U-NeXt: MLP-based rapid medical image segmentation network,” *arXiv preprint*, 2022. DOI: **10.48550/arXiv.2203.04967**

[24] J. Ruan, S. Xiang, *et al.*, “MALUNet: A multi-attention and lightweight U-Net for skin lesion segmentation,” in *Proc. IEEE Int. Conf.* Bioinf. Biomed. (BIBM), 2022, pp. 1150–1156. DOI: 10.1109/BIBM55620.2022.9995165

[25] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, “U-Net++: A nested U-Net architecture for medical image segmentation,” in *Deep Learning in Medical Image Analysis*, Springer, 2018, pp. 3–11. DOI: **10.1007/978-3-030-00889-5_1**

[26] O. Oktay et al., “Attention U-Net: Learning where to look for the pancreas,” arXiv preprint, 2018. DOI: 10.48550/arXiv.1804.03999

[27] J. Wei et al., “Shallow attention network for polyp segmentation,” in Proc. MICCAI, 2021, pp. 699–708. DOI: 10.1007/978-3-030-87231-1_64

پاورقی‌ها:

[1] Test-Time Augmentation
[2] Test-Time Dropout
[3] Edge-Guided Module
[4] Spatial–Channel Attention Block
[5] Deep Supervision
[6] Multi-Scale Test-Time Augmentation
[7] Test-Time Adaptation
[8] Tent
[9] Self-Ensembling at Test-Time
[10] Uncertainty-Guided Refinement
[11] Memory-based Batch Normalization
[12] Visual Prompt Adaptation
[13] Lightweight Test-Time Optimization