



Statistical Reliability Assessment of Deep Learning Architectures for Tomato Leaf Disease Classification

Shahin Rafiee 

Department of Biosystems Mechanical Engineering, Faculty of Agriculture and Natural Resources, University of Tehran,
Karaj, Iran, Email: shahinrafiee@ut.ac.ir

Article Info

Article type: Research Article

Article history:

Received: Feb. 24, 2026

Revised: Apr. 17, 2026

Accepted: June. 10, 2024

Published online: Autumn 2026

Keywords:

*Transfer Learning,
Convolutional Neural
Networks,
Model Robustness,
Overfitting,
Plant Health Monitoring.*

ABSTRACT

Accurate and timely detection of tomato leaf diseases is essential for reducing crop losses, improving productivity, and supporting sustainable agriculture. Despite the growing success of deep learning approaches in this field, most existing studies report results based on a single training run and primarily focus on peak accuracy, while neglecting statistical stability. This limitation weakens the reliability and real-world applicability of many proposed models. To address this gap, this study introduces a stability-oriented evaluation framework for a comprehensive comparison of deep learning architectures for tomato leaf disease classification. A baseline convolutional neural network trained from scratch was systematically compared with state-of-the-art transfer learning models from the ResNet, GoogLeNet, EfficientNet, and DenseNet families using a standard benchmark dataset. To ensure a fair and realistic assessment, all models were trained under 25 fully independent runs with random initialization. Performance was evaluated using multiple metrics, including accuracy, precision, recall, F1-score, Matthews correlation coefficient, area under the ROC curve, and the standard deviation of results across repeated runs. The results demonstrate that transfer learning models consistently outperform the baseline CNN not only in terms of average accuracy but also in statistical stability. Among all evaluated architectures, DenseNet-121 achieved the most reliable performance, with a mean accuracy of 0.996 and the lowest standard deviation (0.0011). Qualitative analysis of confusion matrices further confirmed reduced inter-class misclassifications. These findings highlight the importance of multi-run stability analysis for selecting dependable deep learning models in practical smart agriculture systems.

Cite this article: Rafiee, Sh., (2026) Statistical Reliability Assessment of Deep Learning Architectures for Tomato Leaf Disease Classification, *Iranian Journal of Biosystem Engineering*, 57 (3),25-53.
<https://doi.org/10.22059/ijbse.2026.411881.665640>

© The Author(s).

Publisher: The University of Tehran Press.

DOI: <https://doi.org/10.22059/ijbse.2026.411881.665640>



EXTENDED ABSTRACT

Introduction

Tomato (*Solanum lycopersicum*) is one of the most important agricultural crops worldwide and plays a critical role in food security and agricultural sustainability. However, tomato production is continuously threatened by a wide range of fungal, bacterial, and viral diseases that can significantly reduce yield and quality if not detected at early stages. Timely and accurate disease diagnosis is therefore essential for effective crop management, reduction of economic losses, and minimization of excessive pesticide usage. Conventional disease identification based on visual inspection relies heavily on expert knowledge, is time-consuming, subjective, and often unreliable, particularly when different diseases exhibit visually similar symptoms.

Recent advances in precision agriculture have accelerated the adoption of artificial intelligence and computer vision techniques for automated plant disease detection. Deep learning, and especially convolutional neural networks (CNNs), has demonstrated strong capability in learning discriminative features directly from leaf images. Nevertheless, many existing studies primarily report results from a single training run and focus mainly on accuracy, while overlooking the statistical stability and reliability of model performance. This limitation raises concerns regarding the robustness and practical applicability of proposed models. To address this gap, the present study provides a systematic comparative evaluation of a baseline CNN trained from scratch and several transfer learning architectures, with a particular emphasis on performance consistency across multiple independent runs.

Materials and Methods

The experiments were conducted using the tomato subset of the publicly available PlantVillage dataset, which includes RGB images of healthy leaves and nine common tomato diseases, namely Bacterial Spot, Early Blight, Late Blight, Leaf Mold, Septoria Leaf Spot, Two-Spotted Spider Mites, Target Spot, Tomato Mosaic Virus, and Tomato Yellow Leaf Curl Virus. After data refinement, the dataset was divided into training, validation, and test sets following a standard split protocol.

A total of six deep learning models were implemented using the PyTorch framework. These models included a custom Baseline CNN trained from scratch and five transfer learning architectures—ResNet-18, ResNet-34, ResNet50, EfficientNet-B0, and DenseNet-121—initialized with ImageNet pre-trained weights. All input images were resized to 224×224 pixels and normalized using ImageNet statistics. Data augmentation techniques, including random rotations and horizontal flips, were applied to enhance generalization.

To ensure statistical reliability and reduce the influence of random weight initialization, each model was trained and evaluated under 25 fully independent runs. Model performance was assessed using accuracy, precision, recall, F1-score, Matthews correlation coefficient (MCC), area under the ROC curve (AUC), and confusion matrices. Mean values and standard deviations across the repeated runs were reported to capture both performance level and stability.

Results and Discussion

The experimental results revealed a pronounced performance gap between the baseline model and transfer learning architectures. The Baseline CNN exhibited clear signs of overfitting and high variability across runs, leading to inferior generalization and unstable classification behavior. In contrast, all transfer learning models achieved substantially higher accuracy and more consistent performance.

Among the evaluated architectures, DenseNet-121 emerged as the most reliable model, achieving the highest average accuracy and F1-score while exhibiting the lowest standard deviation across the 25 independent runs. This indicates a high degree of statistical stability and robustness to random initialization. Qualitative analysis of confusion matrices further showed that DenseNet-121 significantly reduced misclassifications among visually similar disease classes, demonstrating superior class-level discrimination. While other transfer learning models such as ResNet and EfficientNet also delivered strong performance, their variability across runs was slightly higher compared to DenseNet-121.

Conclusions

This study presents a comprehensive and statistically reliable evaluation of deep learning architectures for tomato leaf disease classification. The findings confirm that transfer learning is crucial for achieving both high accuracy and robust generalization in agricultural image analysis. The baseline CNN trained from scratch failed to deliver consistent and dependable results, whereas pre-trained deep architectures showed clear advantages. DenseNet-121, in particular, provided the best balance between classification accuracy and statistical stability across repeated executions. The main contribution of this work lies in highlighting the importance of multi-run stability analysis and qualitative confusion matrix evaluation as essential criteria for

selecting reliable models in real-world smart agriculture applications. Future research will explore extending this framework to field-acquired datasets and investigating transformer-based models with an emphasis on improving generalization and stability.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of Generative AI and AI-assisted technologies in the writing process

there is nothing to disclose.

Data availability statement

Data available on request from the authors.

Acknowledgements

The authors would like to express their sincere appreciation to the Vice-Chancellor for Research of the University of Tehran for the moral support provided during the implementation of this study. The authors also thank the respected reviewers for their constructive structural and scientific comments.

Ethical considerations

The authors avoided data fabrication, falsification, and plagiarism, and any form of misconduct.

Conflict of interest

The authors declare no conflict of interest.

ارزیابی آماری پایداری معماری‌های یادگیری عمیق در تشخیص بیماری‌های برگ گوجه‌فرنگی

شاهین رفیعی^۱^۱. گروه مهندسی میانیک بیوسیستم، دانشکده کشاورزی و منابع طبیعی، دانشگاه تهران، کرج، ایران، رایانامه: Shahinrafiee@ut.ac.ir

اطلاعات مقاله

چکیده

نوع مقاله: مقاله پژوهشی

تاریخ دریافت: ۱۴۰۴/۱۲/۵

تاریخ بازنگری: ۱۴۰۵/۱/۲۸

تاریخ پذیرش: ۱۴۰۵/۳/۲۰

تاریخ انتشار: پاییز ۱۴۰۵

واژه‌های کلیدی:

یادگیری انتقالی،

شبکه‌های عصبی پیچشی،

تعمیم پذیری،

بیش‌برازش،

پایش سلامت گیاه

تشخیص دقیق و به‌موقع بیماری‌های برگ گوجه‌فرنگی نقش اساسی در کاهش خسارات، افزایش بهره‌وری و دستیابی به کشاورزی پایدار دارد. بخش عمده‌ای از مطالعات در زمینه یادگیری عمیق در این حوزه، تنها به یک اجرای منفرد بسنده کرده و پایداری آماری را نادیده گرفته‌اند؛ موضوعی که اعتبار و کاربردپذیری عملی این روش‌ها را محدود می‌سازد. هدف این پژوهش، پر کردن این شکاف از طریق ارائه یک چارچوب ارزیابی جامع و پایدار برای مقایسه معماری‌های مختلف یادگیری عمیق در تشخیص بیماری‌های برگ گوجه‌فرنگی است. در این راستا، یک شبکه عصبی پیچشی پایه آموزش‌دیده از ابتدا، در کنار چندین مدل پیشرفته مبتنی بر یادگیری انتقالی از خانواده‌های ResNet، GoogLeNet، EfficientNet و DenseNet بر روی مجموعه داده استاندارد مرجع مورد ارزیابی تطبیقی قرار گرفت. برای تضمین مقایسه‌ای واقع‌بینانه، تمامی مدل‌ها تحت ۲۵ اجرای کاملاً مستقل با مقداردهی اولیه تصادفی آموزش داده شدند و شاخص‌های متنوعی شامل دقت، صحت، بازخوانی، امتیاز F1، ضریب همبستگی متیوز، مساحت زیر منحنی ROC و انحراف معیار نتایج، محاسبه و تحلیل شد. نتایج نشان داد که مدل‌های مبتنی بر یادگیری انتقالی، علاوه بر دستیابی به دقت بالاتر، از پایداری آماری به مراتب بیشتری نسبت به شبکه پایه برخوردارند. در میان آن‌ها، DenseNet-121 با دقت میانگین ۰/۹۹۶ و کمترین انحراف معیار (۰/۰۰۱۱)، پایدارترین عملکرد را داشت. تحلیل کیفی ماتریس‌های اغتشاش نیز کاهش معنادار خطاهای بین‌کلاسی را تأیید کرد. این یافته‌ها نشان می‌دهد که ارزیابی چندجانبه و توجه به پایداری آماری، گامی ضروری برای انتخاب مدل‌های قابل اعتماد در سامانه‌های هوشمند تشخیص بیماری گیاهان و توسعه کاربردهای عملی در کشاورزی آینده است.

استناد: رفیعی؛ شاهین، (۱۴۰۵) ارزیابی آماری پایداری معماری‌های یادگیری عمیق در تشخیص بیماری‌های برگ گوجه‌فرنگی، مجله مهندسی بیوسیستم ایران/ایران،

<https://doi.org/10.22059/ijbse.2026.411881.665640> ۵۷-۵۳، (۳)، ۲۵-۵۳

© نویسندگان.

ناشر: مؤسسه انتشارات دانشگاه تهران.

DOI: <https://doi.org/10.22059/ijbse.2026.411881.665640>

مقدمه

کشاورزی به عنوان ستون فقرات اقتصاد جهانی، نقش حیاتی در تامین امنیت غذایی ایفا می‌کند؛ اما همواره با چالش‌های گوناگونی روبرو بوده است. یکی از مهم‌ترین این چالش‌ها، خسارات ناشی از بیماری‌های گیاهی به ویژه در محصولات مانند گوجه‌فرنگی است. گوجه‌فرنگی، به‌عنوان یک محصول با ارزش تجاری بالا، متداوم در معرض خطر تهاجم انواع بیماری‌زاهای قارچی، باکتریایی و ویروسی قرار دارد. شیوع این بیماری‌ها نه تنها به شدت کمیت و کیفیت محصول را کاهش می‌دهد، بلکه منجر به کاهش ۲۰ درصدی مصرف سموم با تشخیص زودهنگام، افزایش هزینه‌های تولید و مصرف بی‌رویه آفت‌کش‌ها می‌شود؛ مسائلی که به تبع آن پیامدهای نامطلوب زیست‌محیطی و خطرات بهداشتی برای انسان به همراه دارد. بنابراین، تشخیص سریع، دقیق و زودهنگام بیماری‌ها به‌عنوان اولین و مهم‌ترین خط دفاعی در مدیریت پایدار کشاورزی تلقی می‌شود و عاملی کلیدی در کاهش ضایعات محصول، افزایش بهره‌وری و تضمین امنیت غذایی به شمار می‌رود.

روش‌های سنتی تشخیص بیماری، که به‌صورت عمده بر مشاهدات بصری و قضاوت متخصصان گیاهپزشکی تکیه دارند، با چالش‌های جدی از جمله ذهنیت، هزینه‌های بالا، زمان‌بر بودن و خطای انسانی مواجه هستند. این چالش‌ها به‌ویژه زمانی پیچیده‌تر می‌شود که علائم بیماری‌های مختلف (مانند لکه باکتریایی و سوختگی زودرس) به شدت مشابه یکدیگر هستند و تشخیص آن‌ها حتی برای چشم‌های کارآزموده نیز دشوار به نظر می‌رسد. در دنیای امروز که با پدیده‌هایی مانند تغییرات اقلیمی، کاهش منابع آبی و نیاز فزاینده به افزایش تولید غذا مواجه هستیم، نمی‌توان به روش‌های سنتی و کند گذشته اتکا کرد. این ضرورت، منجر به ظهور کشاورزی هوشمند شده است که در آن از فناوری‌های پیشرفته، شامل هوش مصنوعی و بینایی رایانه، برای بهینه‌سازی و اتوماسیون فرآیندهای کشاورزی بهره‌برداری می‌شود. در قلب این تحولات، بینایی رایانه به ماشین‌ها این قابلیت را می‌دهد که با "دیدن" و تجزیه و تحلیل تصاویر دیجیتال، اطلاعات معناداری استخراج کنند. هنگامیکه این قابلیت با توانمندی‌های یادگیری عمیق، به‌خصوص شبکه‌های عصبی پیچشی، ترکیب می‌شود، امکان توسعه سامانه‌های خودکار برای شناسایی و طبقه‌بندی الگوهای پیچیده و ظریف مرتبط با بیماری‌های گیاهی فراهم می‌آید. این فناوری قابلیت آن را دارد که تشخیص بیماری‌ها را از یک فرآیند واکنشی به یک سیستم نظارتی پیشگیرانه و مستمر تبدیل کند.

دستیابی به مدلی که علاوه بر دقت مناسب، از پایداری آماری و قابلیت اطمینان بالا برخوردار باشد، نقشی کلیدی در عملیاتی‌سازی سامانه‌های تشخیص بیماری در کشاورزی هوشمند ایفا می‌کند. چنین مدلی می‌تواند به‌صورت یک برنامه کاربردی مخصوص تلفن همراه در اختیار کشاورزان قرار گیرد یا بر روی سامانه‌های پایش و ربات‌های مزرعه و گلخانه مستقر شود تا به‌صورت بلادرنگ، غیرمخرب و قابل اعتماد سلامت گیاهان را پایش نماید. این امر زمینه‌ساز مدیریت هدفمند و به‌موقع بیماری‌ها، کاهش مصرف بی‌رویه سموم و در نهایت، افزایش پایداری اقتصادی و زیست‌محیطی تولید محصولات کشاورزی خواهد بود. از این منظر، پژوهش حاضر نه تنها یک تحلیل آکادمیک، بلکه گامی کاربردی در جهت توسعه سامانه‌های هوشمند کشاورزی و تحقق امنیت غذایی پایدار محسوب می‌شود.

طی سال‌های اخیر، یادگیری عمیق به‌عنوان رویکردی محوری در پایش سلامت گیاه مطرح شده و معماری‌هایی مانند شبکه عصبی پیچشی، مدل‌های باقیمانده (ResNet)، EfficientNet و ترنسفورمرهای بینایی توانسته‌اند عملکرد چشمگیری در تشخیص بیماری‌های برگ ارائه دهند (افشاری‌پور و شمسی، ۱۴۰۱؛ شیشه‌چی، ۱۴۰۳؛ He & Tong, 2025؛ Ferentinos, 2018). مرور جامع انجام شده در سال ۲۰۲۵ با تمرکز بر طبقه‌بندی، تشخیص و بخش‌بندی بیماری‌های گوجه‌فرنگی نشان می‌دهد که این معماری‌ها ظرفیت توسعه سامانه‌های سریع، مقیاس‌پذیر و مقرون‌به‌صرفه را دارا هستند. پژوهش‌ها در سه حوزه اصلی طبقه‌بندی، تشخیص شیء و بخش‌بندی دسته‌بندی می‌شوند (Das et al., 2025a; Das et al., 2025b) و در تمامی این حوزه‌ها، استفاده از مدل‌های از پیش آموزش‌دیده و یادگیری انتقالی نقش کلیدی دارد. در طبقه‌بندی تصاویر بیماری، نخستین مطالعات با استفاده از GoogLeNet، VGG، AlexNet و ResNet انجام شد. با پیدایش EfficientNet، مسیر معماری‌های مقیاس‌پذیر و کم‌هزینه‌تر هموار شد.

محققین، گزارش کرده‌اند که ResNet، EfficientNet و مدل‌های ترنسفورمرمحور معمولاً در محیط‌های کنترل‌شده به دقت‌های ۹۵٪ تا ۹۹٪ می‌رسند، هرچند در داده‌های واقعی کاهش دقت مشاهده می‌شود. این مسئله ضرورت ارزیابی تطبیقی معماری‌ها در شرایط واقعی را برجسته می‌سازد (Das et al., 2025b; Atila et al., 2021).

1. *Solanum lycopersicum*

2. Accuracy

3. Vision Transformers

برای تقویت تمرکز مدل بر نواحی بیمار و کاهش تأثیر پس‌زمینه، معماری‌های توجه‌محور توسعه یافته‌اند. در تحقیقی در سال ۲۰۲۳، شبکه توجه تکراری چندکاناله معرفی شد که در داده‌های واقعی به دقت ۹۷٪ تا ۹۸٪ دست یافت. مدل‌های ترکیبی مانند RIC-Net که ساختارهای اینسپشن و باقیمانده را با بلوک‌های توجه آدغام می‌کنند نیز عملکرد بالاتری نسبت به شبکه عصبی پیچشی ساده ارائه داده‌اند (Zhao et al., 2022). همچنین در پژوهشی در سال ۲۰۲۴، نشان داده شد که در شرایط کمبود داده، روش‌های یادگیری با نمونه‌های اندک قادر به ارائه دقت‌های رقابتی هستند (Zhang et al., 2023; Rezaei et al., 2024).

در حوزه طبقه‌بندی، استفاده از شبکه‌های مختلف همچنان ادامه دارد: در سال ۲۰۱۷، AlexNet و SqueezeNet در گلخانه آزمایش شد؛ در سال ۲۰۲۰، شبکه عصبی پیچشی اختصاصی با دقت ۹۱٪/۲٪ ارائه شد؛ در همان سال، با یادگیری انتقالی به ۹۷٪/۱۰٪ رسید؛ چهار سال بعد، شبکه عصبی پیچشی، AlexNet و VGG16 مقایسه شد و تحقیقی دیگری در سال ۲۰۲۴، نشان داد که در تشخیص چهار سبزی، MobileNet با دقت ۸۴٪/۴۹٪ برتر از VGG16 و CNN عمل می‌کند (Durmus et al., 2017; Agarwal et al., 2020; Hong et al., 2020; Shanthi et al., 2024; Qabulio et al., 2024).

در سال‌های اخیر، ورود ترنسفورمر بینایی تحول جدیدی ایجاد کرد و گزارش شد که ViT-Base با دقت ۹۹٪/۱۷٪ در دهکده گیاه^۵ و عملکرد قوی در مجموعه داده واقعی TomatoEbola، مدل‌های VGG19، InceptionV3، ResNet50، EfficientNetB2 و DMCCNN را پشت سر گذاشته است. این یافته‌ها نقش ViT را در تشخیص بیماری‌های گیاهی پررنگ‌تر کرده‌اند (Shehu et al., 2025).

یادگیری انتقالی با بهره‌گیری از مدل‌های آموزش‌دیده بر ایملجنت‌نیز به‌طور گسترده برای کاهش نیاز به داده و سخت‌افزار استفاده شده است. برای مثال، Ahmadi (۲۰۲۵) با EfficientNet-B0 روی یک لپ‌تاپ فاقد GPU توانست دقت ۸۵٪ را برای طبقه‌بندی ۱۰ بیماری گوجه‌فرنگی ثبت کند (Ahmadi, 2025). همچنین در تحقیق دیگری، با ریزتنظیم^۶ مدل VGG، دقت ۹۵٪/۷۱٪ به دست آوردند (Paymode & Malode, 2022).

در حوزه تشخیص شیء، برخلاف طبقه‌بندی که تنها نوع بیماری را مشخص می‌کند، مدل‌های تشخیص شیء قادر به تعیین محل دقیق ضایعه هستند؛ قابلیت مهم برای کاربردهای رباتیک و سم‌پاشی هدفمند (Zhao et al., 2019). در تحقیقی از Faster R-CNN برای شناسایی همزمان ۹ آفت استفاده شد. مدل‌های دو مرحله‌ای مانند Faster/Mask R-CNN معمولاً دقت بالایی دارند (Fuentes et al., 2022)؛ اما مدل‌های تک‌مرحله‌ای خانواده YOLO (YOLOv8 & YOLOv5/v7) و EfficientDet برای سرعت بیشتر توسعه یافته‌اند (Sun et al., 2024; Kumar et al., 2024).

در پژوهشی مجموعه داده واقعی TomDet معرفی شد و نشان داد که مدل Mask R-CNN در میان مدل‌های غیر YOLO بهترین تعادل بین mAP و امتیاز F1 ارائه می‌دهد و YOLOv9 با امتیاز F1 نزدیک ۹۸٪ پیش‌تاز است. همچنین در دسته‌های دشوار مانند ویروس موزائیک، مدل‌های طبقه‌بندی مانند ConvNeXt و ViT-Large با افت امتیاز F1 تا حدود ۵۶٪ مواجه می‌شوند، در حالی که مدل‌های تشخیصی به دلیل تمرکز بر ضایعه، پایداری بیشتری دارند (Ruhad et al., 2025).

در پژوهشی، تشخیص بیماری‌های رایج گل‌کلم با استفاده از مدل LeNet امکان‌پذیر شد و مدل توانست چهار طبقه شامل گیاه سالم و گیاه با بیماری کپک پودری، پوسیدگی سیاه و پوسیدگی نرم باکتریایی را با دقت نزدیک به ۱۰۰٪ شناسایی کند (پوردربانی و سبزی، ۱۴۰۲). در مطالعه‌ای دیگر، تشخیص دو بیماری سفیدک پودری و آنتراکنوز برگ خیار با استفاده از شبکه عصبی مصنوعی و پردازش تصویر انجام گرفت و از ترکیب ویژگی‌های بافتی، رنگی و شکل برای شناسایی استفاده شد (حسینی و همکاران، ۱۳۹۶). همچنین در پژوهشی بر روی برگ‌های درخت سیب، استخراج ویژگی بر اساس بافت، رنگ و تبدیل موجک و طبقه‌بندی با شبکه عصبی مصنوعی، توانست بیماری‌های لکه سیاه، آلترناریا و مینوز را با دقت ۱۰۰٪ تشخیص دهد (عمرانی م همکاران، ۱۳۹۲).

در زمینه یادگیری انتقالی، محققین گزارش نموده‌اند که معماری‌های پیش‌آموزش‌دیده مانند MobileNet، EfficientNet و

1. Attention-Based Architectures
2. Multi-Channel Recurrent Attention Network
3. Attention Blocks
4. Few-Shot Learning
5. PlantVillage
6. ImageNet
7. Fine-Tuning

ResNet عملکرد تشخیص بیماری‌های برگ درخت سیب را ارتقا می‌دهد (روستایی و نوروزی، ۱۴۰۳). در پژوهشی، تشخیص بیماری سفیدک پودری هلو با پردازش تصویر و شبکه عصبی انجام شد (سلیمانی و همکاران، ۱۳۹۵).

در سال ۱۴۰۲، ترکیب الگوریتم بهینه‌ساز پروانه و ماشین بردار پشتیبان (SVM) برای تشخیص سه بیماری پوسیدگی سیاه، اسکا و لکه برگ به دقت ۹۸/۷۵٪ دست یافت (نجف‌آبادی‌ها و همکاران، ۱۴۰۲)، و در سال ۱۴۰۱، ویژگی‌های LBP، GLCM، GLRLM و روش‌های SVM برای تشخیص بیماری‌های برگ انگور استفاده شد (نصیری و خجسته‌نژاد، ۱۴۰۱). رویکردهای طیف‌سنجی، تشخیص بیماری سوختگی زودرس گوجه‌فرنگی قبل از ظهور علائم از طریق طیف‌سنجی مرئی/فروسرخ و شبکه عصبی مصنوعی امکان‌پذیر شد (Azadshahraki et al., 2022).

محققین توسعه روش‌های یادگیری عمیق در تشخیص بیماری‌های گیاهی را بررسی نمودند و میانگین دقت مدل‌های یادگیری عمیق حدود ۹۵/۶٪ گزارش گردید. از منظر تبدیل موجک و شبکه‌های عمیق، مطالعه‌ای در سال ۲۰۲۲ روش جدیدی برای طبقه‌بندی بیماری‌های گیاهان مبتنی بر تبدیل موجک ارائه داد و عملکرد شبکه‌های AlexNet، GoogLeNet، ResNet و شبکه عمیق ساده را مقایسه کرد (Karimi & Naderifar, 2022؛ کریمی و نادری‌فر، ۱۴۰۱). در سال ۱۴۰۱، تشخیص بیماری‌های سطح برگ برنج با پردازش تصویر و انتخاب ویژگی‌های رنگی و بافتی انجام شد (پیمان و همکاران، ۱۳۹۵). همچنین در سال ۱۳۹۵، تشخیص بیماری‌های برگ گوجه‌فرنگی با یادگیری عمیق و پردازش تصویر تأیید شد (آق‌آتابای و همکاران، ۱۳۹۵).

اگرچه پژوهش‌های متعددی به‌کارگیری مدل‌های یادگیری عمیق را در تشخیص بیماری‌های گیاهی گزارش کرده‌اند، بخش عمده این مطالعات ارزیابی عملکرد مدل‌ها را به یک اجرای منفرد و اغلب با تمرکز صرف بر بیشینه دقت محدود ساخته‌اند. چنین رویکردی، پایداری آماری، قابلیت تکرارپذیری و حساسیت مدل‌ها نسبت به عواملی مانند مقداردهی اولیه تصادفی وزن‌ها، ترتیب نمونه‌های آموزشی و نوسانات فرایند یادگیری را نادیده می‌گیرد؛ در حالی‌که این عوامل، به‌ویژه در شبکه‌های عمیق و در شرایط محدودیت داده‌های برچسب‌خورده کشاورزی، می‌توانند به تغییرات معنادار و حتی متناقض در نتایج منجر شوند.

از این‌رو، علی‌رغم پیشرفت‌های چشمگیر، همچنان خلأ یک ارزیابی مقایسه‌ای نظام‌مند؛ آماری و تکرارپذیر از پایداری و قابلیت اطمینان معماری‌های مختلف یادگیری عمیق در شرایط کاملاً یکسان در ادبیات پژوهش، به‌ویژه در مقایسه‌ی مدل‌های آموزش‌دیده از پایه با معماری‌های پیشرفته مبتنی بر یادگیری انتقالی مشهود است.

افزون بر این، بسیاری از مطالعات پیشین، یا به بررسی جداگانه معماری‌ها بسنده کرده‌اند و یا بدون تحلیل انحراف معیار و رفتار مدل در اجراهای تکرارشونده، به مقایسه دقت متوسط پرداخته‌اند (He et al., 2016; Tan & Le, 2019). همچنین، تحلیل طبقه‌های دشوار و خطاهای متقابل ناشی از شباهت‌های بافتی و رنگی بیماری‌ها، نظیر همپوشانی میان لکه باکتریایی و سوختگی زودرس، اغلب به‌صورت محدود و غیرنظام‌مند انجام شده است (Shanthi et al., 2024). در سطح ملی نیز، بخش قابل توجهی از پژوهش‌ها معطوف به محصولات دیگر یا روش‌های سنتی بوده و بررسی جامع بیماری‌های برگ گوجه‌فرنگی با تمرکز بر پایداری مدل‌ها کمتر مورد توجه قرار گرفته است. در کنار اهمیت دقت مدل‌ها، باید توجه داشت که شرایط واقعی مزرعه با چالش‌هایی همچون تغییرات شدید نور، وجود سایه‌ها، حرکت برگ‌ها در اثر باد، لرزش دوربین در پهپادها یا تراکتورها، و پیچیدگی پس‌زمینه شامل خاک، علف‌های هرز و سایر اجزای مزرعه همراه است. مجموعه داده‌هایی مانند دهکده گیاه اگرچه برای آموزش اولیه مدل‌ها بسیار رایج هستند؛ اما تصاویر آن در محیط‌های کنترل‌شده و با پس‌زمینه یکنواخت ثبت شده‌اند و این امر باعث می‌شود بسیاری از مدل‌های آموزش‌دیده بر روی آن در مواجهه با داده‌های واقعی مزرعه عملکردی به‌مراتب پایین‌تر ارائه دهند. مرور مطالعاتی که از دهکده گیاه استفاده کرده‌اند نیز نشان می‌دهد که اگرچه نتایج آنها از نظر دقت چشمگیر است؛ اما اغلب به مسئله پایداری و قابلیت تعمیم مدل‌ها توجه کافی نشده است. بنابراین، بررسی پایداری عملکرد مدل‌ها در شرایط کنترل‌شده، همانند آنچه در این پژوهش با اجرای ۲۵ باره هر مدل انجام شد، می‌تواند بینش ارزشمندی درباره قابلیت اعتماد و پایداری مدل‌ها ارائه دهد و زمینه مقایسه منسجم‌تری با مطالعات پیشین فراهم سازد.

پژوهش حاضر با هدف پر کردن این خلأها، به ارزیابی جامع و تطبیقی مجموعه‌ای متنوع از معماری‌های یادگیری عمیق شامل یک شبکه پیچشی آموزش‌دیده از پایه و چندین مدل شاخص مبتنی بر یادگیری انتقالی از خانواده‌های باقیمانده؛ اینسپشن^۱ و چگال^۲، بر روی

1. Residual
2. Inception
3. Dense

یک مجموعه داده معتبر^۱ و تحت شرایط آزمایشی کاملاً یکسان می‌پردازد. نوآوری اصلی این مطالعه، فراتر از مقایسه صرف دقت، بر تحلیل آماری پایداری، قابلیت تعمیم و رفتار مدل‌ها در اجراهای مستقل استوار است؛ به گونه‌ای که هر معماری در ۲۵ اجرای کاملاً مستقل آموزش داده شده و مجموعه‌ای از شاخص‌های عملکردی برای دستیابی به قضاوتی واقع‌بینانه و قابل اتکا استخراج و تحلیل می‌شود. این رویکرد، دیدگاهی فراتر از گزارش یک نتیجه بهینه منفرد ارائه داده و چارچوبی مناسب برای انتخاب مدل‌های قابل اعتماد در کاربردهای واقعی کشاورزی هوشمند فراهم می‌آورد.

مواد و روش‌ها

توصیف داده‌ها و نحوه تقسیم مجموعه داده‌ها

برای ارزیابی پایداری آماری معماری‌های یادگیری عمیق در مسئله طبقه‌بندی بیماری‌های برگ گوجه‌فرنگی، از زیرمجموعه مربوط به گوجه‌فرنگی در پایگاه داده‌ی دهکده گیاه^۲ استفاده شد. این پایگاه داده یکی از شناخته‌شده‌ترین مجموعه داده‌ها در حوزه بینایی کامپیوتری کشاورزی است که شامل تصاویر با کیفیت بالا از برگ‌های سالم و بیمار است. تصاویر این مجموعه در شرایط کنترل شده آزمایشگاهی و با پس‌زمینه یکنواخت جمع‌آوری شده‌اند؛ عاملی که امکان ارزیابی دقیق رفتار آماری مدل‌ها را فراهم می‌کند، گرچه ممکن است عملکرد آن‌ها در شرایط مزرعه با پس‌زمینه‌های پیچیده متفاوت باشد.

این مجموعه شامل تصاویر برگ‌های سالم و آلوده به نه بیماری مختلف قارچی، باکتریایی و ویروسی است. کلاس‌های داده عبارت از باکتریایی لکه‌دار^۳، سوختگی زودرس^۴، سوختگی دیررس^۵، کپک برگ^۶، لکه سپتوریایی برگ^۷، کنه دو نقطه‌ای^۸، لکه هدفی^۹، ویروس موزائیک گوجه‌فرنگی^{۱۰}، ویروس پیچیدگی زرد برگ گوجه‌فرنگی^{۱۱} و کلاس سالم^{۱۲} (شکل ۰۱) است. این تصاویر در مجموع چندین هزار نمونه را شامل می‌شوند که تعداد آن‌ها برای تحلیل پایداری با تکرارهای متعدد (۲۵ بار اجرا) مناسب است.

انتخاب این دیتاست به این دلیل انجام شد که علاوه بر تنوع بیماری‌های قارچی، ویروسی و باکتریایی، امکان مقایسه با مطالعات پیشین را فراهم می‌کند و یک بستر استاندارد برای ارزیابی پایداری آماری و تغییرپذیری عملکرد معماری‌های مختلف شبکه‌های عصبی پیچشی فراهم می‌سازد.

این داده‌ها پس از بررسی کیفیت، پالایش و حذف داده‌های تکراری، بر اساس اصل برابری نمونه‌ها با نسبت ۷۰٪، ۱۵٪ و ۱۵٪ به ترتیب برای آموزش، اعتبارسنجی و آزمون تقسیم شده‌اند. جدول ۱ توزیع تعداد تصاویر هر کلاس از مجموعه تصاویر مرجع و نحوه تقسیم آن‌ها به سه بخش آموزش، اعتبارسنجی و آزمون را نشان می‌دهد. بر اساس این تقسیم، بزرگ‌ترین کلاس مربوط به بیماری ویروس پیچیدگی زرد برگ گوجه‌فرنگی با ۵۳۶۳ تصویر و کوچک‌ترین کلاس مربوط به ویروس موزائیک گوجه‌فرنگی با ۳۸۲ تصویر است؛ این تفاوت ذاتی همانند اغلب مجموعه داده‌های گیاهی، ناشی از فراوانی طبیعی و شرایط ثبت اولیه نمونه‌هاست. چنین ناهمگونی جزئی، با استفاده از روش‌های افزایش داده^{۱۳} شامل چرخش تصادفی، برش و وارون‌سازی تصویر جبران شده است تا مدل‌ها توانایی تعمیم بهتری پیدا کنند.

رویه آزمایشی و تحلیل آماری

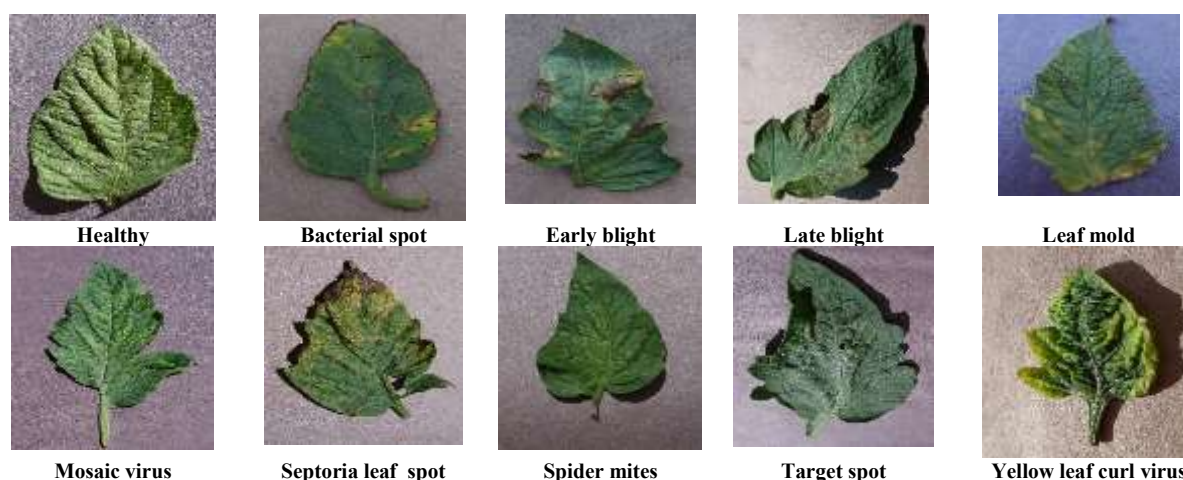
این پژوهش با هدف تشخیص خودکار بیماری‌های برگ گوجه‌فرنگی از روی تصاویر و بررسی کارایی معماری‌های گوناگون یادگیری عمیق در این مسئله انجام شده است. طراحی پژوهش به گونه‌ای انجام گرفت که بتوان دقت، پایداری و قابلیت تعمیم مدل‌ها را به صورت کمی

۱. داده‌ها از وبسایت PlantVillage استخراج شده است.

2. PlantVillage
3. Bacterial Spot
4. Early Blight
5. Late Blight
6. Leaf Mold
7. Septoria Leaf Spot
8. Spider Mite
9. Target Spot
10. Tomato Mosaic Virus
11. Yellow Leaf Curl Virus
12. Healthy
13. Data Augmentation

ارزیابی کرد (Mohanty et al., 2016).

شکل ۲ جریان کار چارچوب پیشنهادی از ورود داده تا ارزیابی نهایی مدل‌ها را نشان می‌دهد. تمامی مدل‌ها تحت شرایط یکسان و با تنظیمات ۱۵ دوره آموزش^۱ و ۲۵ اجرای کاملاً مستقل با مقداردهی اولیه تصادفی مورد آموزش قرار گرفتند تا پایداری آماری تحلیل شود. در هر اجرا، از همان نسبت تقسیم داده‌های (۷۰، ۱۵ و ۱۵) استفاده شد تا متغیرهای خارج از کنترل حداقل شوند. مدل‌های مورد آزمایش شامل: شبکه‌های پیچشی پایه، ResNet-18، ResNet-34، ResNet50، EfficientNet-B0، GoogLeNet و DenseNet-121 است و تمامی مدل‌ها غیر از شبکه‌های پیچشی پایه با استفاده از وزن‌های ایمپجنت^۲ اولیه‌سازی و با تابع خطای آنتروپی متقاطع^۳ و بهینه‌ساز آدام^۴ (نرخ یادگیری: ۰/۰۰۱) آموزش داده شدند. ارزیابی عملکرد با استفاده از معیارهای دقت، صحت^۵، بازخوانی^۶ و امتیاز F1 و نیز معیارهای مقاوم شامل MCC و منحنی‌های ROC/AUC انجام شد. افزون بر این، پایداری آماری مدل‌ها در ۲۵ اجرای مستقل مورد بررسی قرار گرفت (Ferentinos, 2018; Zhang et al., 2020; Ghanbari et al., 2022; Reimers and Gurevych, 2017).

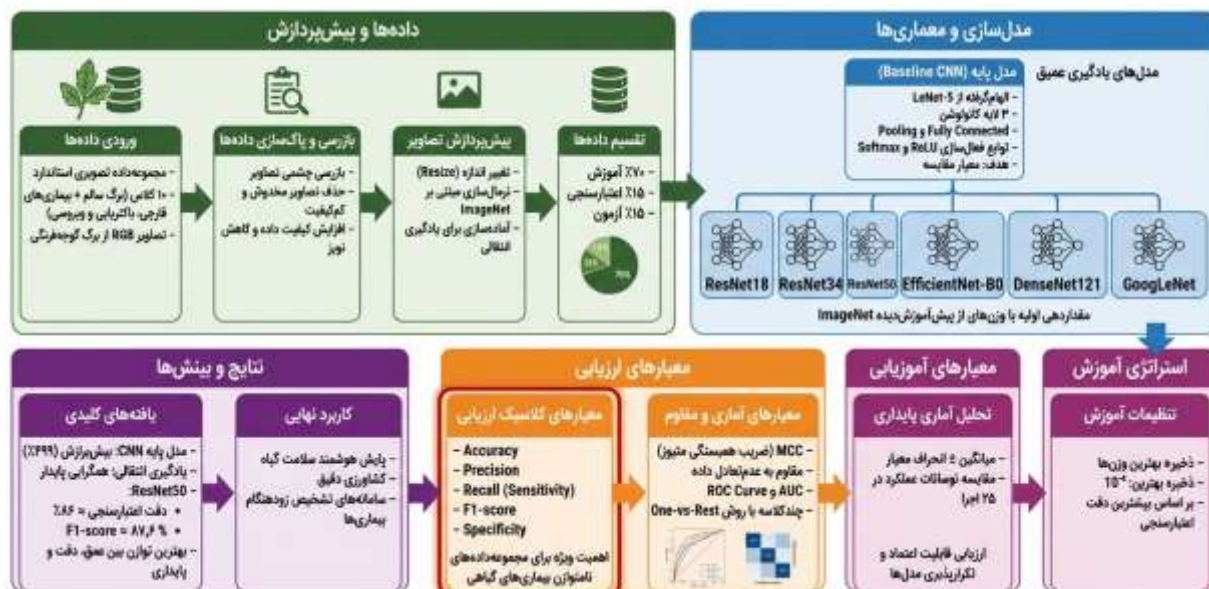


شکل ۱- برگ‌های کلاس‌های دهگانه گوجه فرنگی مورد مطالعه

جدول ۱- توزیع مجموعه تصاویر و نحوه تقسیم به آموزش، اعتبارسنجی و آزمون

کلاس	کل تصاویر	آموزش	اعتبارسنجی	آزمون
لکه باکتریایی	۲۱۳۶	۱۴۹۵	۳۲۰	۳۲۱
سوختگی زودرس	۱۰۰۹	۷۰۶	۱۵۱	۱۵۲
سالم	۱۵۹۸	۱۱۱۸	۳۳۹	۲۴۱
سوختگی دیررس	۱۹۱۹	۱۳۴۳	۲۸۷	۲۸۹
کپک برگ	۹۵۷	۶۶۹	۱۴۳	۱۴۵
ویروس موزاییک	۳۸۲	۲۶۷	۵۷	۵۸
لکه برگ سپتوریا	۱۷۸۲	۱۲۴۷	۲۶۷	۲۶۸
کنه تارتن	۱۶۷۶	۱۱۷۳	۲۵۱	۲۵۲
لکه هدف	۱۴۰۴	۹۸۲	۲۱۰	۲۱۲
ویروس پیچیدگی زرد برگ	۵۳۶۳	۳۷۵۴	۸۰۴	۸۰۵

1. Epoch
2. ImageNet
3. Cross-Entropy
4. Adam
5. Precision
6. Recall



شکل ۲. نمودار چارچوب تشخیص بیماری‌های برگ گوجه‌فرنگی با استفاده از مدل‌های یادگیری عمیق و تحلیل آماری پایدار

مدل شبکه پیچشی پایه^۱

این مدل با الهام از ساختار کلاسیک $L \times N \times 3$ طراحی شد تا به عنوان معیار پایه برای سنجش عملکرد مدل‌های عمیق تر عمل کند. شبکه شامل سه لایه کانولوشن^۲ با کرنل 3×3 ، دو لایه پولینگ^۳ و دو لایه کاملاً متصل^۴ بود. از تابع فعال‌سازی $ReLU$ و تابع خروجی $Softmax$ برای تفکیک چند کلاسه استفاده شد. این معماری به دلیل تعداد پارامتر پایین، سرعت آموزش زیاد، ولی دقت پایین‌تر در شناسایی الگوهای بیماری انتخاب شد تا بتوان مقیاس پیشرفت سایر مدل‌ها را نسبت به آن سنجید. ویژگی‌های این مدل شامل سادگی ساختار و توضیح‌پذیری بالا، مناسب برای مقایسه اولیه و فاقد مکانیزم اجتناب از ناپدید شدن گرادیان است (اسدی، ۱۴۰۴، Odusami et al., 2021). در این پژوهش تمرکز بر مدل‌های شبکه پیچشی پایه با وزن کمتر و قابلیت تکرارپذیری بالاتر قرار گرفته است تا تحلیل پایداری آماری با دقت بیشتری انجام شود.

۱. معماری شبکه‌های باقیمانده^۵

معماری شبکه‌های باقیمانده توسط هی و همکاران در سال ۲۰۱۶ معرفی شد و یکی از تحولات بنیادین در یادگیری عمیق به شمار می‌رود. ایده اصلی این معماری، استفاده از اتصالات میان‌بر^۶ است که امکان یادگیری نگاشت باقیمانده را فراهم می‌کند. این سازوکار موجب بهبود جریان گرادیان در شبکه‌های بسیار عمیق شده و مشکل گرادیان ناپدیدشونده را به طور مؤثر کاهش می‌دهد. شبکه‌های باقیمانده نشان داد که افزایش عمق شبکه، در صورت استفاده از اتصالات باقیمانده، نه تنها باعث افت عملکرد نمی‌شود، بلکه می‌تواند دقت را نیز بهبود دهد و به همین دلیل، پایه‌گذار استفاده گسترده از شبکه‌های بسیار عمیق در بینایی ماشین مدرن شد (He et al., 2016).

۲. مدل ResNet-18

ResNet-18 یکی از نسخه‌های کم‌عمق‌تر خانواده ResNet است که شامل ۱۷ لایه کانولوشنی و یک لایه تمام‌اتصال می‌باشد. این مدل از بلوک‌های باقیمانده پایه^۷ استفاده می‌کند و دارای حدود ۱۱/۷ میلیون پارامتر قابل آموزش است. به دلیل پیچیدگی محاسباتی کمتر و ساختار ساده‌تر، ResNet-18 گزینه‌ای مناسب برای کاربردهایی با محدودیت منابع محاسباتی محسوب می‌شود. با وجود سادگی، این مدل توانایی یادگیری ویژگی‌های مؤثر را داشته و به ویژه در چارچوب یادگیری انتقالی، عملکرد قابل اعتمادی از خود نشان می‌دهد (He et al., 2016).

1. Baseline CNN
2. Convolution
3. Max Pooling
4. Dense
5. Residual Networks – ResNet
6. Skip Connections
7. Basic Residual Blocks

۳. مدل ResNet-34

مدل ResNet-34 نسخه‌ای عمیق‌تر از ResNet-18 است که با ۳۴ لایه، ظرفیت بیشتری برای استخراج ویژگی‌های سطح بالا فراهم می‌کند. این مدل همچنان از بلوک‌های باقیمانده پایه استفاده می‌کند؛ اما عمق بیشتر آن امکان یادگیری الگوهای پیچیده‌تری مانند ساختارهای بافتی و روابط فضایی دقیق‌تر را فراهم می‌سازد. تعداد پارامترهای ResNet-34 حدود ۲۱/۸ میلیون است و استفاده هم‌زمان از اتصالات باقیمانده و نرمال‌سازی دسته‌ای باعث حفظ پایداری آموزش و جلوگیری از افت عملکرد در شبکه‌های عمیق‌تر می‌شود (He et al., 2016).

۴. مدل ResNet50

ResNet50 یکی از پرکاربردترین اعضای خانواده ResNet است که از بلوک‌های باقیمانده بطری‌شکل^۱ بهره می‌برد. در این بلوک‌ها، کانولوشن‌های ۱×۱ برای کاهش و سپس بازیابی ابعاد کانالی استفاده می‌شوند که ضمن حفظ توان بازنمایی، کارایی محاسباتی را افزایش می‌دهند. این مدل دارای حدود ۲۵/۶ میلیون پارامتر قابل آموزش است و توانایی بالایی در یادگیری ویژگی‌های انتزاعی و سطح بالا دارد. ResNet50 به‌طور گسترده به‌عنوان شبکه پایه در بسیاری از سامانه‌های پیشرفته بینایی ماشین و یادگیری انتقالی به‌کار گرفته می‌شود (He et al., 2016).

۵. مدل EfficientNet-B0

EfficientNet توسط تان و لی در سال ۲۰۱۹ معرفی شد و رویکردی نوین برای مقیاس‌بندی شبکه‌های کانولوشنی ارائه داد. نوآوری اصلی این معماری، استفاده از مقیاس‌بندی مرکب^۲ است که در آن عمق، عرض شبکه و وضوح ورودی به‌صورت هم‌زمان و متوازن تنظیم می‌شوند. نسخه پایه EfficientNet-B0 دارای حدود ۵/۳ میلیون پارامتر است و با وجود اندازه کوچک، دقت رقابتی بالایی نسبت به بسیاری از معماری‌های عمیق‌تر ارائه می‌دهد. این ویژگی باعث شده EfficientNet-B0 گزینه‌ای مناسب برای کاربردهای با منابع محاسباتی محدود، مانند سامانه‌های لبه و دستگاه‌های همراه، باشد (Tan and Le, 2019).

۶. مدل GoogLeNet

معماری گوگل‌نت^۳ با نام اینسپشن نسخه یک^۴ شناخته می‌شود، توسط سزگدی و همکاران در سال ۲۰۱۵ معرفی شد. نوآوری اصلی این شبکه، معرفی ماژول‌های اینسپشن است که امکان استخراج هم‌زمان ویژگی‌ها در مقیاس‌های مختلف (مانند ۱×۱، ۳×۳ و ۵×۵) را در یک لایه فراهم می‌کنند. GoogLeNet با وجود عمق ۲۲ لایه، تنها حدود ۵ تا ۷ میلیون پارامتر دارد که این امر عمدتاً به استفاده گسترده از کانولوشن‌های ۱×۱ برای کاهش بعد ویژگی‌ها بازمی‌گردد. این معماری توازن مناسبی بین دقت و هزینه محاسباتی برقرار می‌کند، هرچند ساختار پیچیده ماژول‌های اینسپشن می‌تواند فرآیند بازآموزی دقیق را چالش‌برانگیزتر سازد (Szegedy et al., 2015).

۷. مدل DenseNet-121

DenseNet توسط هانگ و همکاران در سال ۲۰۱۷ معرفی شد و مبتنی بر مفهوم اتصال متراکم^۵ است. در این معماری، خروجی هر لایه به‌صورت مستقیم به ورودی تمام لایه‌های بعدی متصل می‌شود، به‌گونه‌ای که هر لایه به مجموعه‌ای غنی از ویژگی‌های قبلی دسترسی دارد. مدل DenseNet-121 با وجود عمق زیاد، تنها حدود ۸ میلیون پارامتر دارد که نشان‌دهنده بهره‌وری بالای آن از نظر استفاده از پارامترها است. این معماری جریان گرادینان را بهبود می‌بخشد، استفاده مجدد از ویژگی‌ها را تقویت می‌کند و برای داده‌هایی با الگوهای ظریف و پیچیده، عملکرد بسیار مناسبی دارد. با این حال، به‌دلیل اتصال متراکم لایه‌ها، مصرف حافظه در مرحله آموزش افزایش می‌یابد (Huang et al., 2017).

پیکربندی آموزش مدل‌ها^۶

پیکربندی دقیق مراحل آموزش نقشی اساسی در ارتقای کارایی مدل‌ها و پیشگیری از بروز بیش‌برازش دارد. در این پژوهش، هدف اصلی

1. Bottleneck Blocks
2. Compound Scaling
3. GoogLeNet
4. Inception v1
5. Dense Connectivity
6. Model Training Configuration

ایجاد توازن میان سرعت همگرایی، پایداری فرایند یادگیری و توانایی تعمیم مدل‌ها به داده‌های دیده‌نشده بود. از آن‌جا که هر مدل در ۲۵ اجرای مستقل و با مقداردهی اولیه تصادفی آموزش یافت (Ferentinos, 2018)، مجموعه‌ای یکسان از پارامترهای عمومی آموزش در جدول ۲ ارائه شد تا شرایط مقایسه میان معماری‌های مختلف کاملاً منصفانه و یکپارچه حفظ شود.

به‌منظور تضمین مقایسه عادلانه میان تمامی مدل‌ها، تمامی ابرپارامترها یکسان در نظر گرفته شدند. در همه آزمایش‌ها از بهینه‌ساز Adam با نرخ یادگیری اولیه 10^{-4} و اندازه Batch، ۳۲ استفاده شد. ضرایب کاهشی ممان‌ها به ترتیب برابر $\beta_1 = 0.9$ و $\beta_2 = 0.999$ تنظیم شده، مقدار $\epsilon = 10^{-8}$ به‌منظور پایداری عددی به‌کار رفت (Kingma & Ba, 2015). انتخاب Adam بر اساس گزارش‌های پیشین مبنی بر همگرایی سریع و کارایی مناسب آن در مجموعه‌داده‌های کوچک و متوسط، از جمله Mohanty et al. در سال ۲۰۱۶، انجام شده است.

جدول ۲. پارامترهای عمومی آموزش مدل‌های استفاده شده در تحقیق

پارامتر	مقدار و توضیح	منبع
بهینه‌ساز ^۱	روش بهینه‌سازی، Adam با نرخ یادگیری اولیه 10^{-4} ، $\beta_1 = 0.9$ ، $\beta_2 = 0.999$ ، و $\epsilon = 10^{-8}$ برای ثبات گرادینانت انتخاب شد؛ علت انتخاب Adam، سرعت همگرایی بالا در داده‌های کوچک بود (Kingma & Ba, 2015).	Mohanty et al., 2016
تابع هزینه ^۲	Cross-Entropy چندکلاسه، مناسب برای مسائل طبقه‌بندی تصویر با چند کلاس مستقل.	Tan & Le, 2019
تعداد دوره آموزش ^۳	۱۵ دوره آموزش برای هر مدل؛ آموزش تا دوره آموزش ۱۵ جهت ثبت کامل منحنی‌های Loss و Accuracy بدون توقف انجام می‌شود. بهترین دوره آموزش براساس منحنی LOSS تعیین می‌شود.	Quan et al., 2023
معیار توقف زودهنگام ^۴	غیرفعال؛ به‌صورت عمدی اجرا نشد تا روند کامل همگرایی هر مدل و رفتار Overfitting مشاهده شود.	Zhao et al., 2022
اندازه دسته ^۵	۳۲ نمونه در هر دسته؛ مقدار میان‌زا برای تعادل بین Memory GPU و دقت گرادینانت.	Ferentinos, 2018
تفکیک داده‌ها	۷۰، ۱۵ و ۱۵٪ به ترتیب برای آموزش، اعتبارسنجی و آزمون؛ ارزیابی با میانگین ۲۵ اجرا مستقل.	Ghanbari et al., 2022

پیش‌پردازش و آماده‌سازی ورودی‌ها

برای یکسان‌سازی داده‌ها با شبکه‌های عصبی پیچشی (CNN) و تسهیل فرایند همگرایی، ابتدا عملیات پیش‌پردازش به‌صورت استاندارد انجام شد. اندازه‌ی تمام تصاویر به 224×224 پیکسل تغییر داده شد و در مدل EfficientNet-B0 نیز همان اندازه‌ی ورودی اصلی مبتنی بر مجموعه‌داده‌ی ایمجنت حفظ گردید. سپس رنگ‌های تصویر نرمال‌سازی شدند؛ به این معنا که هر کانال رنگی RGB براساس میانگین و انحراف معیار مجموعه‌داده‌ی ایمجنت تنظیم شد تا مدل بتواند به‌درستی از وزن‌های پیش‌آموزش‌داده‌شده بهره‌برد. به‌منظور افزایش تنوع داده‌ها و کاهش بیش‌برازش، مجموعه‌ای از تبدیلات تصادفی (افزایش داده) به‌کار گرفته شد؛ که شامل چرخش تا $\pm 20^\circ$ درجه، وارون‌سازی افقی، تغییر روشنایی تا $\pm 15\%$ ، تغییر اشباع تا $\pm 10\%$ ، کاهش رزولوشن و اعمال نویز گوسی برای شبیه‌سازی شرایط واقعی گلخانه‌ها بود (Zhao et al., 2022). در ادامه، داده‌ها برای مراحل آموزش و اعتبارسنجی به‌صورت تصادفی اما بدون هیچ‌گونه همپوشانی میان کلاس‌ها تقسیم شدند تا از نشت داده جلوگیری شود. این تفکیک دقیق تضمین می‌کند که مدل صرفاً براساس داده‌های واقعی و مستقل ارزیابی شود و هیچ نمونه‌ای هم‌زمان در هر دو بخش آموزش و اعتبارسنجی قرار نگیرد.

تنظیمات یادگیری انتقالی

برای به‌کارگیری یادگیری انتقالی، تمامی مدل‌های ازپیش‌آموزش‌دیده شامل ResNet و EfficientNet با وزن‌های اولیه مبتنی بر مجموعه‌داده ایمجنت مورد استفاده قرار گرفتند (Too et al., 2019). در این فرایند، لایه‌های ابتدایی شبکه که وظیفه استخراج ویژگی‌های عمومی را بر عهده دارند، در مرحله نخست منجمد شدند تا الگوهای پایدار و مستقل از دامنه حفظ گردند. در مقابل، لایه‌های بالادستی یا

1. Optimizer
2. Loss Function
3. Epochs
4. Early Stopping
5. Batch Size
6. Train
7. Validation

بخش طبقه‌بند مجدداً آموزش داده شدند. به‌طور مشخص، در مدل ResNet50، سه بلوک پایانی شبکه باز نگه داشته شد تا امکان یادگیری ویژگی‌های اختصاصی مجموعه داده هدف فراهم شود. این راهبرد آموزش مرحله‌ای که در مطالعات معتبر پیشین نیز گزارش شده است، موجب بهبود معنادار عملکرد مدل در کاربردهای کشاورزی و تصویربرداری گیاهی می‌شود (Too et al., 2019; Chungcharoen et al., 2022).

در مدل EfficientNet-B0، بهره‌گیری از سازوکار مقیاس‌بندی مرکب^۱ باعث شد اندازه دسته‌های آموزشی و نرخ یادگیری به‌صورت خودتنظیم و سازگار با پیچیدگی مدل انتخاب شود (Tan & Le, 2019). نتیجه این رویکرد، همگرایی سریع‌تر و ایجاد فلات در مقدار خطای اعتبارسنجی^۲ در حدود دوره آموزش هفتم بود (Zhao et al., 2022).

معیارهای ارزیابی آماری

برای ارزیابی جامع و مقایسه‌ای عملکرد مدل‌های یادگیری عمیق، علاوه بر معیارهای مرسوم دقت، صحت، بازخوانی و امتیاز F1، از دو شاخص آماری پیشرفته‌تر استفاده شد: ضریب همبستگی متیوز^۳ (MCC) و مساحت زیر منحنی مشخصه عملکرد گیرنده^۴ (AUC). این شاخص‌ها به دلیل حساسیت بالا به عدم تعادل کلاس‌ها و ارائه دیدگاهی متوازن از عملکرد کلی مدل، انتخاب شدند (Matthews, 1975).

ضریب همبستگی متیوز (MCC)

در این پژوهش، ضریب همبستگی متیوز (MCC) ابتدا برای هر کلاس به‌صورت یک مسئله دودویی «کلاس مورد نظر در برابر سایر کلاس‌ها» محاسبه شده است و تعداد نمونه‌های کلاس C که به اشتباه در سایر کلاس‌ها قرار گرفته‌اند، هستند. مقدار MCC برای هر کلاس رابطه یک محاسبه می‌شود:

$$MCC_c = \frac{TP_c \times TN_c - FP_c \times FN_c}{\sqrt{(TP_c + FP_c)(TP_c + FN_c)(TN_c + FP_c)(TN_c + FN_c)}} \quad \text{رابطه ۱}$$

که در آن:

TP_c : تعداد نمونه‌های کلاس C که به‌درستی پیش‌بینی شده‌اند.

TN_c : تعداد نمونه‌های سایر کلاس‌ها که به‌درستی به عنوان «غیر از C» پیش‌بینی شده‌اند.

FP_c : تعداد نمونه‌های سایر کلاس‌ها که به اشتباه به عنوان کلاس C پیش‌بینی شده‌اند.

FN_c : تعداد نمونه‌های کلاس C که به اشتباه به سایر کلاس‌ها نسبت داده شده‌اند.

سپس برای به‌دست آوردن یک شاخص کلی از عملکرد مدل در مسئله طبقه‌بندی چندکلاسه، از میانگین وزنی MCC کلاس‌ها استفاده شده است. اگر n_c تعداد نمونه‌های کلاس C در مجموعه آزمون و C تعداد کل کلاس‌ها باشد، MCC نهایی مدل به‌صورت رابطه دو تعریف می‌شود:

$$MCC_{\text{weighted}} = \frac{\sum_{c=1}^C n_c MCC_c}{\sum_{c=1}^C n_c} \quad \text{رابطه ۲}$$

بدین ترتیب، وزن هر کلاس متناسب با فراوانی آن در مجموعه آزمون در نظر گرفته شده و شاخص نهایی نسبت به عدم تعادل داده‌ها حساسیت کنترل‌شده‌ای خواهد داشت.

مساحت زیر منحنی ROC (AUC)

منحنی مشخصه عملکرد گیرنده (ROC) نرخ مثبت‌های واقعی را در برابر نرخ مثبت‌های کاذب برای آستانه‌های تصمیم‌گیری مختلف ترسیم می‌کند. مساحت زیر این منحنی (AUC) معیاری از توانایی مدل در تمایز بین کلاس‌ها است و مستقل از مقدار آستانه طبقه‌بندی عمل می‌کند. مقدار AUC بین ۰/۵ (پیش‌بینی تصادفی) تا ۱/۰ (تمایز کامل) متغیر است. در این پژوهش، برای محاسبه AUC در مسئله چندکلاسه، از رویکرد یک‌در برابر بقیه^۵ (OvR) استفاده شده است. در این رویکرد، برای هر کلاس یک منحنی ROC مستقل محاسبه می‌شود و در نهایت، میانگین وزنی مساحت زیر این منحنی‌ها متناسب با فراوانی هر کلاس به‌عنوان مقدار نهایی AUC گزارش می‌گردد.

1. Compound Scaling
2. Validation Loss
3. Matthews Correlation Coefficient
4. Area Under the ROC Curve
5. One-vs-Rest



این دو شاخص، در کنار معیارهای سنتی، امکان ارزیابی عمیق‌تر پایداری و قابلیت اطمینان مدل‌ها را در شرایط عدم تعادل کلاس‌ها فراهم کرده و مبنای تحلیل مقایسه‌ای نهایی در بخش نتایج قرار گرفته‌اند (Fawcett, 2006).

فرایند آموزش و ثبت نتایج آزمایش

فرایند آموزش برای هر مدل براساس مجموعه‌ای از گام‌های مشخص انجام گرفت. در ابتدا وزن‌های مدل در هر اجرا به صورت تصادفی و مستقل مقداردهی اولیه شدند و برای تضمین تکرارپذیری، یک بذر تصادفی ثابت در هر تکرار تنظیم شد. پس از آغاز آموزش، در پایان هر دوره آموزشی، شاخص‌های عملکرد شامل خطا و دقت در هر دو بخش آموزش و اعتبارسنجی محاسبه گردید.

طراحی آزمایش و تحلیل آماری

به منظور ارزیابی جامع عملکرد مدل‌ها، از معیارهای استاندارد زیر استفاده شد: دقت به عنوان نسبت پیش‌بینی‌های صحیح به کل نمونه‌ها؛ بازخوانی^۲ برای سنجش توانایی مدل در شناسایی صحیح تمام موارد مثبت واقعی (به‌ویژه کلاس‌های مرتبط با بیماری)؛ صحت^۳ برای ارزیابی میزان درستی پیش‌بینی‌های مثبت مدل و در نهایت امتیاز F1 به عنوان میانگین همساز صحت و بازخوانی که تعادل مناسبی بین این دو معیار فراهم می‌آورد (Zhang et al., 2022). هر یک از مدل‌ها به صورت کاملاً مستقل و در ۲۵ اجرای جداگانه با مقداردهی اولیه تصادفی متفاوت آموزش داده شد؛ سپس میانگین و انحراف معیار، معیارهای استاندارد محاسبه شد (مشابه روش به کار رفته در Ferentinos در سال ۲۰۱۸). جدول ۳ فرمول بندی و مشخصات معیارهای استاندارد را ارائه می‌دهد. که در آن TP، FP و FN به ترتیب نشان‌دهنده تعداد موارد درست مثبت، نادرست مثبت و نادرست منفی هستند.

جدول ۳- معیارهای استاندارد برای ارزیابی عملکرد مدل

معیار ارزیابی	معادل انگلیسی	تعریف کوتاه	روند تغییر	توضیح برای پژوهش تشخیص بیماری گیاهی (چرا مهم است؟)
دقت (A)	$Accuracy = \frac{TP + TN}{TP + TN + FP}$	نسبت کل پیش‌بینی‌های صحیح به تعداد کل نمونه‌ها	هر چه بیشتر، بهتر	شاخص کلی، ولی اگر داده‌ها نامتعادل باشند ممکن است گمراه کننده باشد
صحت (P)	$Precision = \frac{TP}{TP + FP}$	از میان تمام نمونه‌هایی که مدل گفته «بیمار است»، چندتا واقعاً بیمار بودند؟	هر چه بیشتر، بهتر	خیلی مهم، یعنی وقتی مدل می‌گوید «این برگ بیمار است»، احتمال واقعا بیمار بودن آن چقدر است؟ (کاهش هشدار اشتباه)
بازخوانی (R)	$Recall = \frac{TP}{TP + FN}$	از میان تمام برگ‌های واقعاً بیمار، مدل چند درصد را درست تشخیص داد؟	هر چه بیشتر، بهتر (بسیار مهم در بیماری گیاهی)	مهم‌ترین شاخص در کشاورزی، هرچه بالاتر باشد یعنی مدل کمتر بیماری را از قلم می‌اندازد (جلوگیری از گسترش بیماری)
امتیاز F1 (F1)	$F1 = 2 \left(\frac{P \times R}{P + R} \right)$	میانگین همساز صحت و بازخوانی	هر چه بیشتر، بهتر	بهترین شاخص کلی وقتی داده‌ها نامتعادل هستند (مثلاً ۸۰٪ سالم و ۲۰٪ بیمار). بالاترین امتیاز F1 یعنی بهترین تعادل
ویژگی خاص	$Specificity = \frac{TN}{TN + FN}$	از میان تمام برگ‌های واقعاً سالم، مدل چند درصد را درست سالم تشخیص داد؟	هر چه بیشتر، بهتر	مهم برای کاهش هزینه سم‌پاشی بی‌مورد روی گیاهان سالم
انحراف معیار (Std)		پراکندگی نتایج در ۲۵ اجرای مستقل	هر چه کمتر، بهتر	نشان‌دهنده پایداری و قابلیت تکرار نتایج است

یافته‌های پژوهش

هدف این بخش، ارائه و تحلیل نتایج حاصل از تشخیص خودکار بیماری‌های برگ گوجه‌فرنگی با استفاده از معماری‌های مختلف یادگیری عمیق است. در این پژوهش، کارایی مدل‌ها بر اساس میانگین نتایج ۲۵ اجرای مستقل و بر روی مجموعه آزمون ارزیابی شده است تا علاوه بر دقت، قابلیت تعمیم و پایداری آماری آن‌ها نیز به صورت واقع‌بینانه بررسی شود.

مقایسه عملکرد کلی معماری‌ها

Error! Reference source not found. ۴ خلاصه‌ای از نوع معماری، تعداد پارامترها، منبع پیش‌آموزش و شاخص‌های عملکرد شامل دقت، بازخوانی و امتیاز F1 را برای تمامی مدل‌های مورد بررسی نشان می‌دهد.

بر اساس نتایج جدول ۴، مدل شبکه پیچشی پایه که بدون استفاده از یادگیری انتقالی آموزش داده شده است، عملکرد ضعیف‌تری نسبت به سایر معماری‌ها دارد و به دقت ۸۸/۴۹٪، بازخوانی ۸۵/۱۰٪ و امتیاز F1 برابر با ۸۶/۱۵٪ دست یافته است. این اختلاف قابل توجه در مقایسه با سایر مدل‌ها نشان می‌دهد که آموزش از پایه، در مواجهه با داده‌های تصویری کشاورزی و تنوع ظاهری بیماری‌ها، با محدودیت جدی در استخراج ویژگی‌های تعمیم‌پذیر روبه‌روست.

جدول ۴- خلاصه معماری‌ها، منبع پیش‌آموزش و میانگین نتایج ۲۵ اجرای مستقل روی مجموعه آزمون

مدل	نوع معماری	پارامتر (میلیون)	دقت (%)	بازخوانی (%)	امتیاز F1 (%)
Baseline CNN	ساده	۱۷	۸۸/۴۹	۸۵/۱۰	۸۶/۱۵
GoogLeNet	اینسپشن	۶/۸	۹۹/۰۷	۹۹/۰۳	۹۹/۰۴
ResNet-18	باقیمانده	۱۱/۲	۹۹/۳۲	۹۹/۲۶	۹۹/۲۹
ResNet-34	باقیمانده	۲۱/۳	۹۹/۲۷	۹۹/۳۱	۹۹/۲۹
ResNet50	باقیمانده عمق زیاد	۲۳/۵	۹۹/۳۰	۹۹/۳۱	۹۹/۳۰
DenseNet-121	چگال ^۳	۸۰	۹۹/۴۰	۹۹/۳۶	۹۹/۳۸
EfficientNet-B0	مقیاس‌پذیر بهینه	۵/۳	۹۹/۲۵	۹۹/۲۶	۹۹/۲۵

در مقابل، تمامی معماری‌های مبتنی بر یادگیری انتقالی که با وزن‌های از پیش آموزش‌دیده ایمیجنت مقداردهی اولیه شده‌اند، عملکردی بسیار بالا و نزدیک به یکدیگر ارائه داده‌اند؛ به گونه‌ای که مقادیر دقت، بازخوانی و امتیاز F1 برای همه این مدل‌ها بیش از ۹۹٪ گزارش شده است. این نتایج نشان می‌دهد که انتقال دانش پیشین نقش تعیین‌کننده‌ای در بهبود دقت و پایداری مدل‌ها، حتی در شرایط ناهمگونی نسبی داده‌ها، ایفا می‌کند.

در میان مدل‌های بررسی‌شده، DenseNet-121 با ثبت بیشترین امتیاز F1 برابر با ۹۹/۳۸٪ و همچنین بالاترین مقادیر دقت و بازخوانی، بهترین عملکرد کلی را نشان داده است. پس از آن، ResNet50 با امتیاز F1 برابر ۹۹/۳۰٪ عملکردی بسیار نزدیک و رقابتی با DenseNet-121 دارد. مدل‌های ResNet-34 و ResNet-18 نیز با وجود عمق کمتر، نتایجی تقریباً هم‌تراز ارائه داده‌اند که بیانگر کارایی بالای معماری‌های باقیمانده حتی در نسخه‌های کم‌عمق‌تر است.

GoogLeNet با معماری اینسپشن و تعداد پارامتر کمتر نسبت به ResNet50، به امتیاز F1 برابر با ۹۹/۰۴٪ دست یافته است که نشان‌دهنده توان مناسب فیلترهای چندمقیاسه در استخراج الگوهای متنوع بیماری‌هاست. همچنین EfficientNet-B0 با تنها ۵/۳ میلیون پارامتر و امتیاز F1 برابر ۹۹/۲۵٪، عملکردی رقابتی ارائه داده و کارایی بالای راهبرد مقیاس‌پذیری بهینه را در دستیابی به دقت بالا با هزینه محاسباتی کمتر تأیید می‌کند.

تحلیل رفتار یادگیری مدل شبکه پیچشی پایه

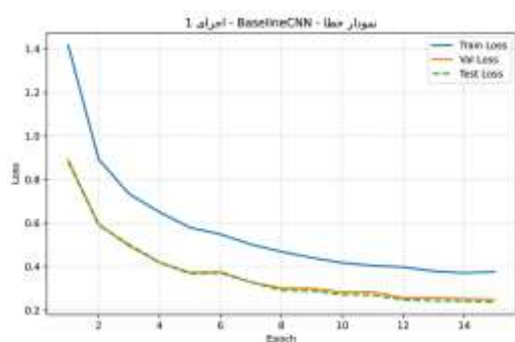
برای بررسی دقیق‌تر رفتار مدل پایه، منحنی‌های یادگیری آن در شکل ۳ ارائه شده است. مطابق این شکل، پس از ۱۵ دوره آموزش، مدل شبکه پیچشی پایه به دقت آزمون حدود ۹۳/۱۹٪ و مقدار خطای حدود ۰/۲۳ دست یافته است. با این حال، روند همگرایی این مدل نسبت به مدل‌های مبتنی بر یادگیری انتقالی کندتر بوده و در بازه زمانی مشابه، به دقت‌های بسیار بالا دست نیافته است.

نکته قابل توجه در شکل ۳، **Error! Reference source not found.** بالاتر بودن دقت اعتبارسنجی و آزمون نسبت به دقت آموزش (حدود ۸۷/۹۸٪ در پایان آموزش) است. این پدیده که تحت تأثیر به کارگیری افزایش داده و روش‌های تنظیم‌کننده در فاز آموزش رخ داده است، نشان می‌دهد که مدل پایه در یادگیری الگوهای پیچیده از داده‌های تغییریافته آموزشی با دشواری مواجه بوده و ظرفیت محدودی برای استخراج ویژگی‌های سطح بالا دارد. اگرچه منحنی‌های اعتبارسنجی و آزمون همبستگی مناسبی نشان می‌دهند و واگرایی شدید مشاهده نمی‌شود؛ اما سرعت همگرایی پایین و سقف عملکرد محدود، ضعف اصلی آموزش از پایه را آشکار می‌سازد.

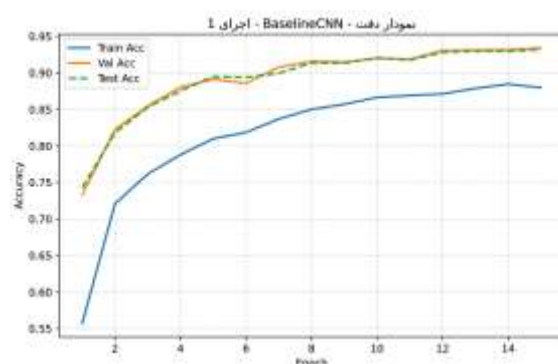
تحلیل رفتار یادگیری بهترین مدل (DenseNet-121)

در مقابل، منحنی‌های یادگیری مدل DenseNet-121 که در شکل ۴ Error! Reference source not found. نمایش داده شده‌اند، رفتاری کاملاً متفاوت را نشان می‌دهند. این مدل به واسطه بهره‌گیری از وزن‌های پیش‌آموزش‌دیده، در همان دوره‌های ابتدایی آموزش (حدود اپوک‌های ۲ تا ۳) از مرز دقت ۹۸٪ عبور کرده و به همگرایی سریع دست یافته است.

مطابق شکل ۴، تعادل بسیار مناسبی میان دقت آموزش، اعتبارسنجی و آزمون برقرار شده و مدل در سطحی نزدیک به ۹۹/۴٪ به پایداری می‌رسد. برخلاف مدل پایه که هنوز در حال کاهش تدریجی خطای آموزش است، DenseNet-121 از همان مراحل اولیه به ناحیه همگرایی پایدار وارد شده است. این رفتار نشان می‌دهد که استفاده از معماری‌های عمیق پیش‌آموزش‌دیده، علاوه بر افزایش دقت نهایی، موجب تسریع فرآیند یادگیری و دستیابی سریع‌تر به عملکرد مطلوب می‌شود.

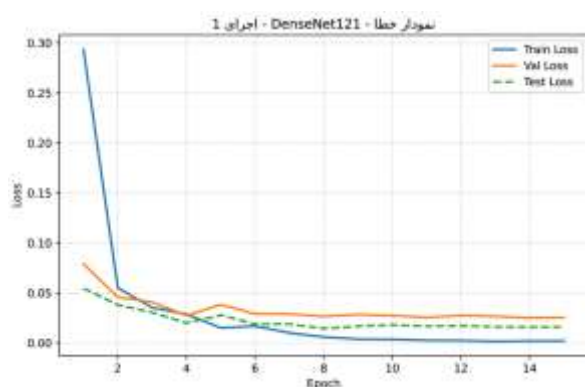


(الف) روند تغییرات تابع هزینه (Loss)

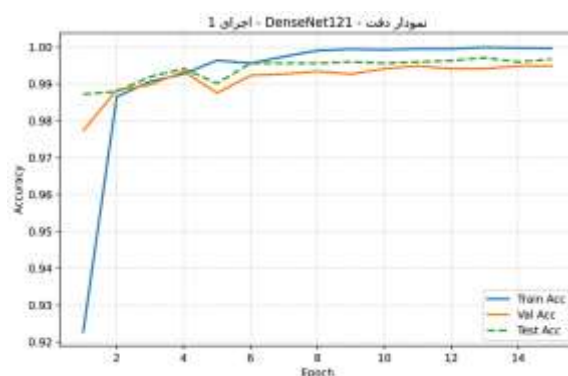


(ب) روند تغییرات میزان صحت (Accuracy) بر روی داده‌های آموزش و اعتبارسنجی.

شکل ۳- منحنی‌های یادگیری مدل شبکه پیچشی پایه در طول فرآیند آموزش



(الف) روند تغییرات تابع هزینه (Loss)



(ب) روند تغییرات میزان صحت (Accuracy) بر روی داده‌های آموزش و اعتبارسنجی.

شکل ۴- منحنی‌های یادگیری مدل DenseNet-121 در طول فرآیند آموزش

عملکرد مدل‌ها در طبقه‌بندی بیماری

جدول ۵ میانگین مقادیر ضریب همبستگی متیوز (MCC) و مساحت زیر منحنی ROC (AUC) را برای مجموعه‌های آموزش، اعتبارسنجی و آزمون، حاصل از ۲۵ اجرای مستقل هر مدل، نشان می‌دهد. این شاخص‌ها به منظور ارزیابی پایداری آماری، قابلیت تعمیم و مقاومت مدل‌ها در برابر عدم تعادل کلاس‌ها گزارش شده‌اند. همان گونه که مشاهده می‌شود، مدل‌های مبتنی بر یادگیری انتقالی در مقایسه با شبکه پایه آموزش‌دیده از ابتدا (شبکه پیچشی پایه)، مقادیر به مراتب بالاتری از MCC و AUC را در هر سه مجموعه آموزش، اعتبارسنجی و آزمون به دست آورده‌اند. مقادیر بالای MCC (بیش از ۰/۹۹ در مجموعه آزمون برای اغلب معماری‌ها) نشان می‌دهد که این مدل‌ها

توانسته‌اند تعادلی مناسب میان نرخ مثبت‌های واقعی و منفی‌های واقعی برقرار کنند؛ موضوعی که به‌ویژه در داده‌های دارای عدم تعادل کلاسی، اهمیت اساسی دارد. همچنین، مقادیر AUC نزدیک به یک بیانگر توان بالای مدل‌ها در تفکیک کلاس‌های بیماری، مستقل از آستانه تصمیم‌گیری است. این موضوع نشان می‌دهد که مدل‌ها نه تنها به یک نقطه بهینه خاص وابسته نیستند، بلکه در بازه وسیعی از آستانه‌ها نیز رفتار پایدار و قابل اتکایی دارند.

بحث پایداری و تعمیم‌پذیری

مقایسه مقادیر MCC و AUC بین مجموعه‌های آموزش، اعتبارسنجی و آزمون، نشان‌دهنده شکاف اندک میان این مجموعه‌ها برای مدل‌های یادگیری انتقالی است. این هم‌خوانی آماری بیانگر آن است که مدل‌ها دچار بیش‌برازش نشده و الگوهای معنادار و قابل تعمیمی از داده‌ها استخراج کرده‌اند. هرچند شبکه پیچشی پایه در برخی اجراها مقادیر نسبتاً بالایی از MCC و AUC در مجموعه‌های آزمون و اعتبارسنجی نشان داده است؛ اما تحلیل پایداری بین اجراها بیانگر نوسان بالا و وابستگی شدید به مقداردهی اولیه است که تعمیم‌پذیری این مدل را زیر سؤال می‌برد.

جدول ۵. مقایسه ضرایب MCC و AUC مدل‌های یادگیری عمیق در مجموعه‌های آموزش، اعتبارسنجی و آزمون

مدل	MCC آزمون	AUC آزمون	MCC آموزش	AUC آموزش	MCC اعتبارسنجی	AUC اعتبارسنجی
Baseline CNN	۰/۹۲۰۵	۰/۹۴۷۶	۰/۸۶۲۵	۰/۹۱۱۶	۰/۹۳۰۰	۰/۹۴۸۴
ResNet-18	۰/۹۹۴۶	۰/۹۹۶۴	۰/۹۹۹۴	۰/۹۹۹۵	۰/۹۹۴۴	۰/۹۹۵۹
ResNet-34	۰/۹۹۳۷	۰/۹۹۵۹	۰/۹۹۹۶	۰/۹۹۹۷	۰/۹۹۴۴	۰/۹۹۵۷
ResNet50	۰/۹۹۳۷	۰/۹۹۵۱	۰/۹۹۹۰	۰/۹۹۹۴	۰/۹۹۳۸	۰/۹۹۵۹
EfficientNet-B0	۰/۹۹۴۴	۰/۹۹۷۰	۰/۹۹۸۲	۰/۹۹۸۹	۰/۹۹۴۶	۰/۹۹۶۷
DenseNet-121	۰/۹۹۴۶	۰/۹۹۶۵	۰/۹۹۹۶	۰/۹۹۹۷	۰/۹۹۴۰	۰/۹۹۵۹
GoogLeNet	۰/۹۹۲۹	۰/۹۹۴۹	۰/۹۹۸۱	۰/۹۹۸۸	۰/۹۹۳۶	۰/۹۹۵۱

تحلیل مقایسه‌ای عملکرد، همگرایی و پایداری مدل‌ها

مقایسه عملکرد کلی مدل‌ها بر اساس ۲۵ اجرای مستقل

جدول ۶ خلاصه‌ای از عملکرد آماری معماری‌های مختلف یادگیری عمیق را بر اساس میانگین و انحراف معیار حاصل از ۲۵ اجرای کاملاً مستقل ارائه می‌دهد. انجام ارزیابی‌ها بر روی مجموعه آزمون و گزارش شاخص‌های آموزش، اعتبارسنجی و آزمون، امکان بررسی هم‌زمان دقت، تعمیم‌پذیری و تکرارپذیری نتایج را فراهم کرده است.

مطابق نتایج جدول ۶، تفاوت عملکردی معناداری میان مدل Baseline CNN (آموزش از پایه) و معماری‌های مبتنی بر یادگیری انتقالی مشاهده می‌شود. مدل پایه با میانگین دقت آزمون ۰/۹۸۷ و خطای آزمون ۰/۳۶۱ عملکرد قابل قبولی دارد؛ اما در مقایسه با مدل‌های پیش‌آموزش‌دیده، توانایی محدودتری در استخراج ویژگی‌های پیچیده و تفکیک دقیق کلاس‌های بیماری از خود نشان می‌دهد. در مقابل، تمامی مدل‌های مبتنی بر یادگیری انتقالی شامل DenseNet-121، خانواده ResNet، EfficientNet-B0 و GoogLeNet، به دقت آزمون بالاتر از ۰/۹۹ دست یافته‌اند. این شکاف عملکردی نزدیک به ۱۰٪ در دقت آزمون، نقش تعیین‌کننده استفاده از وزن‌های اولیه ایملیجنت و معماری‌های عمیق‌تر را در مسائل کشاورزی دقیق که با شباهت بصری بالای کلاس‌ها مواجه‌اند، به‌روشنی نشان می‌دهد.

برتری معماری DenseNet-121

در میان مدل‌های ارزیابی‌شده، DenseNet-121 بهترین عملکرد کلی را ثبت کرده است. این مدل با میانگین دقت آزمون ۰/۹۹۶ و کمترین میانگین خطای آزمون برابر با ۰/۰۱۷ در صدر معماری‌های مورد بررسی قرار می‌گیرد.

برتری DenseNet-121 را می‌توان به ساختار اتصال متراکم آن نسبت داد؛ ساختاری که در آن هر لایه به تمامی لایه‌های بعدی متصل بوده و موجب بهبود جریان اطلاعات و گرادینان در طول شبکه می‌شود. این ویژگی ضمن کاهش خطر محو شدن گرادینان، امکان

استفاده مجدد از ویژگی‌ها را فراهم کرده و به یادگیری الگوهای ظریف بیماری‌های گیاهی با کارایی بالا منجر شده است؛ نتیجه‌ای که در کاهش همزمان خطا و افزایش دقت آزمون به‌خوبی منعکس شده است.

جدول ۶- ارزیابی عملکرد و پایداری معماری‌های یادگیری عمیق بر اساس معیارهای دقت و خطا (نتایج حاصل از ۲۵ تکرار)

مدل	شاخص آماری	دقت آزمون	دقت اعتبارسنجی	دقت آموزش	خطای آزمون
Baseline CNN	میانگین	۰/۸۹۷	۰/۸۹۷	۰/۸۳۱	۰/۳۶۱
	انحراف معیار	۰/۰۰۸۳	۰/۰۰۷۸	۰/۰۰۵۲	۰/۰۱۵۰
DenseNet-121	میانگین	۰/۹۹۶	۰/۹۹۶	۰/۹۹۹	۰/۰۱۷
	انحراف معیار	۰/۰۰۱۱	۰/۰۰۱۰	۰/۰۰۰۶	۰/۰۰۴۲
EfficientNet-B0	میانگین	۰/۹۹۵	۰/۹۹۵	۰/۹۹۴	۰/۰۱۸
	انحراف معیار	۰/۰۰۱۴	۰/۰۰۱۰	۰/۰۰۱۹	۰/۰۰۳۷
GoogLeNet	میانگین	۰/۹۹۳	۰/۹۹۴	۰/۹۹۶	۰/۰۲۴
	انحراف معیار	۰/۰۰۱۳	۰/۰۰۱۴	۰/۰۰۰۸	۰/۰۰۴۲
ResNet-18	میانگین	۰/۹۹۵	۰/۹۹۵	۰/۹۹۹	۰/۰۱۹
	انحراف معیار	۰/۰۰۱۰	۰/۰۰۱۰	۰/۰۰۰۵	۰/۰۰۳۹
ResNet-34	میانگین	۰/۹۹۵	۰/۹۹۶	۰/۹۹۸	۰/۰۱۷
	انحراف معیار	۰/۰۰۱۴	۰/۰۰۱۲	۰/۰۰۰۹	۰/۰۰۴۴
ResNet50	میانگین	۰/۹۹۵	۰/۹۹۶	۰/۹۹۸	۰/۰۱۸
	انحراف معیار	۰/۰۰۱۳	۰/۰۰۱۳	۰/۰۰۱۱	۰/۰۰۴۸

تحلیل پایداری و قابلیت اطمینان آماری مدل‌ها

یکی از مهم‌ترین دستاوردهای این پژوهش که در ستون انحراف معیار (Std) جدول ۶ قابل مشاهده است، تفاوت چشمگیر در پایداری عملکرد میان مدل پایه و معماری‌های مبتنی بر یادگیری انتقالی است.

مدل Baseline CNN دارای انحراف معیار دقت آزمون برابر با ۰/۰۰۸۳ است که نشان‌دهنده نوسان محسوس عملکرد آن در اجراهای مختلف و وابستگی به مقداردهی اولیه وزن‌هاست. در مقابل، DenseNet-121 با انحراف معیار بسیار پایین ۰/۰۰۱۱، پایداری آماری بالایی را در تمامی ۲۵ اجرای مستقل نشان می‌دهد. این بدان معناست که این مدل در شرایط یکسان آموزشی، همواره به دقت بالا و همگرایی یکنواخت دست یافته است.

سایر معماری‌های مبتنی بر یادگیری انتقالی نیز، از جمله ResNet-18، ResNet-34، ResNet50 و EfficientNet-B0، با انحراف معیارهایی در بازه حدود ۰/۰۰۱۰ تا ۰/۰۰۱۴، پایداری بالایی از خود نشان داده‌اند. با این حال، DenseNet-121 به دلیل ترکیب همزمان دقت بالاتر و کمترین واریانس عملکرد، از نظر قابلیت اطمینان آماری در جایگاه برتر قرار می‌گیرد.

تحلیل ماتریس‌های اغتشاش و ارزیابی خطاهای متقابل

براساس ۵۰ ماتریس‌های هم‌ریختگی مدل‌ها نشان‌دهنده عملکرد کاملاً متفاوت دو رویکرد «آموزش از ابتدا» و «یادگیری انتقالی» در طبقه‌بندی بیماری‌های برگ گوجه‌فرنگی است.

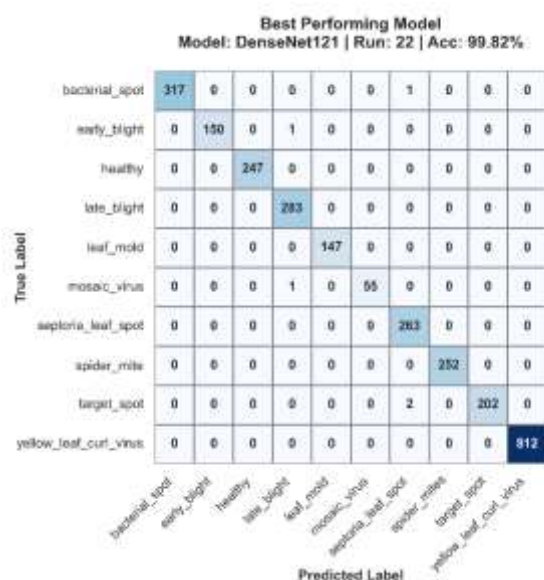
عملکرد مدل شبکه پیچشی پایه (بدترین مدل)

شبکه پیچشی پایه در طبقه‌بندی بیماری‌ها عملکرد نسبتاً ضعیفی داشته و انواع خطاهای پراکنده در اغلب کلاس‌ها رخ داده است (شکل الف). برای نمونه: در کلاس «لکه باکتریایی»، از مجموع ۳۱۵ نمونه، تنها ۲۹۰ نمونه به درستی شناسایی شده‌اند و ۱۵ نمونه با «پیچیدگی برگ زرد» اشتباه گرفته شده‌اند. همچنین ۷ مورد به بیماری‌های دیگر اشتباه تخصیص یافته است. کلاس «سوختگی زودرس» با ۱۵۳ نمونه، تنها ۸۳ نمونه به درستی دسته‌بندی شده‌اند و خطاهای قابل توجهی مثلاً ۱۸ نمونه به «لکه باکتریایی»، ۲۹ نمونه به «سوختگی دیررس»، و ۱۳ نمونه به «لکه هدف» ثبت شده است. حتی کلاس «سالم» که انتظار می‌رود به آسانی تشخیص داده شود، ۲۳۷ نمونه صحیح طبقه‌بندی شده و چند مورد اشتباه (۲ نمونه به «لکه هدف»، ۱ نمونه به «سوختگی دیررس») دیده می‌شود. علاوه بر این، کلاس‌های «ویروس موزاییک» و «کنه تارنکبوتی» نیز خطاهای متقابل قابل ملاحظه‌ای دارند.

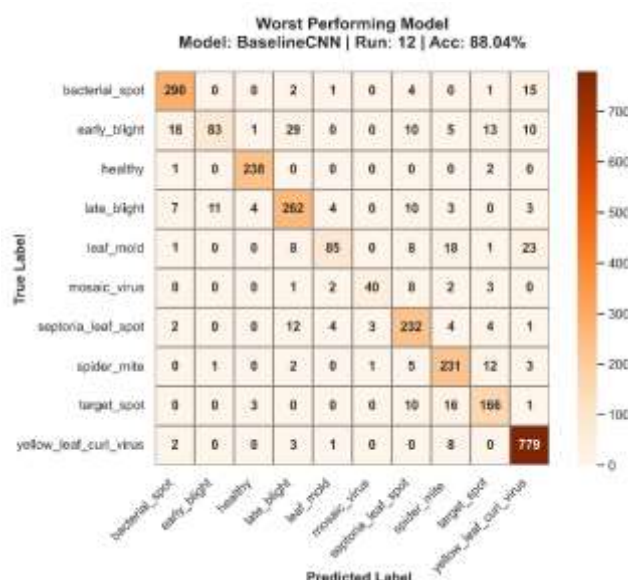
این الگوی پراکندگی خطاها نشان می‌دهد که مدل شبکه پایه: در تفکیک بیماری‌های با شباهت ظاهری و بافتی بالا مانند لکه‌های باکتریایی و سوختگی‌ها ضعف قابل توجهی دارد. حتی در کلاس‌های ساده‌تر، مانند برگ سالم، خطا دارد که نشان‌دهنده ناتوانی در استخراج ویژگی‌های متمایز است. این مدل دقت پایین و همچنین پراکندگی ناپایا در نتیجه‌های مستقل را نشان می‌دهد که با حساسیت بالای وزن‌دهی اولیه و مدل‌سازی غیر بهینه مرتبط است.

عملکرد مدل DenseNet-121 (بهترین مدل)

در نقطه مقابل، مدل DenseNet-121 با بهره‌گیری از ساختار اتصال چگال و وزن‌های پیش‌آموزش‌دیده، عملکرد بسیار برجسته‌ای نشان داده است (شکل ب): کلیه نمونه‌های «لکه باکتریایی» (۳۱۸ مورد)، «سوختگی زودرس» (۱۵۱ نمونه) و «برگ سالم» (۲۴۷ نمونه) به طور کامل و بدون خطا طبقه‌بندی شده‌اند. بیماری‌های «سوختگی دیررس» (۲۸۳ نمونه) و «لکه برگ کپک‌دار» (۱۴۷ نمونه) نیز با دقت کامل پیش‌بینی شده‌اند. تنها خطای اندکی در کلاس «لکه هدف» دیده شد که در آن ۲ مورد به «کنه تارنکبوتی» اختصاص یافته است. این میزان خطا در مقایسه با سایر کلاس‌ها بسیار ناچیز بوده و قابل چشم‌پوشی است. بیماری‌های ویروسی مانند «پیچیدگی برگ زرد» و «ویروس موزاییک گوجه‌فرنگی» نیز به طور کامل و با دقت نزدیک صد درصد شناسایی شدند.



ب- DenseNet-121



الف- شبکه پیچشی پایه

شکل ۵- ماتریس اغتشاش

مقایسه کمی و کیفی و نتیجه‌گیری

جدول ۷ مقایسه دقت طبقه‌بندی کلاس به کلاس بین شبکه پیچشی پایه و DenseNet-121 بر اساس ماتریس هم‌ریختگی مجموعه آزمون ارائه شده است. اختلاف قابل توجه میان ماتریس‌های دو مدل به وضوح اهمیت یادگیری انتقالی و استفاده از معماری‌های عمیق را در شناسایی دقیق بیماری‌های گیاهی آشکار می‌سازد. شبکه پیچشی پایه با دقت‌های پایین‌تر و انتشار خطاهای گسترده به وضوح نشان داد که توانایی تمایز بیماری‌هایی که ویژگی‌های ظاهری مشابه دارند (مانند لکه‌های باکتریایی و سوختگی‌ها) را ندارد. اما DenseNet-121 با دقت تقریباً صد درصد در همه کلاس‌ها، از قابلیت استخراج ویژگی‌های بسیار دقیق بهره‌مند است. این نتایج ضمن تأیید برتری یادگیری انتقالی، نشان‌دهنده اهمیت معماری‌های پیچیده مانند DenseNet با اتصالات چگال در دستیابی به پایداری و دقت بالا در طبقه‌بندی بیماری‌های برگ گوجه‌فرنگی است.

تحلیل مقایسه‌ای دقت آموزش، اعتبارسنجی و آزمون

به منظور ارزیابی هم‌زمان عملکرد، تعمیم‌پذیری و پایداری آماری مدل‌های یادگیری عمیق، نتایج حاصل از ۲۵ اجرای مستقل برای هر معماری مورد تحلیل قرار گرفت. شکل ۶ خلاصه‌ای جامع از رفتار مدل‌ها را از سه منظر میانگین عملکرد، توزیع دقت و توزیع خطا ارائه می‌دهد.

مطابق ۰ (الف)، مدل شبکه پیچشی پایه ضعیف‌ترین عملکرد را در میان تمامی مدل‌ها نشان می‌دهد. این مدل اگرچه به دقت نسبتاً قابل قبول در مرحله آموزش دست یافته است؛ اما در مجموعه‌های اعتبارسنجی و آزمون دچار افت عملکرد شده که بیانگر شکاف تعمیم و ناپایداری یادگیری است. میله‌های خطای بزرگ‌تر (انحراف معیار بالاتر) برای این مدل نیز نشان می‌دهد که نتایج آن به شدت به مقداردهی اولیه وزن‌ها و تصادفی‌بودن فرآیند آموزش وابسته است.

در مقابل، تمامی مدل‌های مبتنی بر یادگیری انتقالی دقتی بسیار بالا و نزدیک به یک را در هر سه مجموعه آموزش، اعتبارسنجی و آزمون ثبت کرده‌اند. هم‌پوشانی قابل توجه دقت‌ها در این سه مجموعه و انحراف معیار بسیار ناچیز نشان می‌دهد که این معماری‌ها از توان تعمیم‌پذیری بالا و رفتار پایدار برخوردارند. این موضوع به‌طور مستقیم برتری استفاده از ویژگی‌های ازپیش‌آموزش‌دیده ایمیج‌نت را در مسائل کشاورزی تأیید می‌کند.

جدول ۷- مقایسه دقت طبقه‌بندی کلاس به کلاس بین شبکه پیچشی پایه و DenseNet-121 بر اساس ماتریس هم‌ریختگی مجموعه آزمون

معیار	شبکه پیچشی پایه	DenseNet-121
دقت طبقه‌بندی لکه باکتریایی	$91.7\% \approx 315 / 340$	$100\% = 318 / 318$
دقت طبقه‌بندی سوختگی زودرس	$54.2\% \approx 153 / 282$	$100\% = 151 / 151$
دقت طبقه‌بندی برگ سالم	$98.7\% \approx 240 / 243$	$100\% = 247 / 247$
دقت طبقه‌بندی لکه هدف	$79.4\% \approx 209 / 263$	$99.0\% \approx 204 / 206$
دقت طبقه‌بندی پیچیدگی برگ زرد	$98.1\% \approx 795 / 810$	$100\% = 812 / 812$

تحلیل پایداری دقت در ۲۵ اجرای مستقل

شکل ۶ (ب) توزیع دقت مجموعه آزمون را برای هر مدل در قالب نمودار جعبه ای نمایش می‌دهد. شبکه پیچشی پایه نه تنها کمترین میانه دقت (حدود ۰/۹) را دارد، بلکه گستره بین چارکی (IQR) بزرگ‌تر و نقاط پرت بیشتری را نیز نشان می‌دهد. این پراکندگی بالا بیانگر بی‌ثباتی شدید مدل در اجراهای مختلف است؛ مسئله‌ای که یکی از دلایل اصلی غیرقابل اعتماد بودن این مدل در کاربردهای عملی محسوب می‌شود.

در مقابل، مدل‌های یادگیری انتقالی توزیعی بسیار فشرده دارند و میانه دقت آن‌ها تقریباً برابر با یک است. و EfficientNet-B0 به‌ویژه کمترین پراکندگی را نشان داده‌اند که بیانگر پایداری آماری بسیار بالا در کنار دقت چشمگیر است. این رفتار یکنواخت در ۲۵ اجرا، نشان می‌دهد که عملکرد این مدل‌ها مستقل از مقداردهی اولیه بوده و از قابلیت اعتماد بالایی برخوردارند.

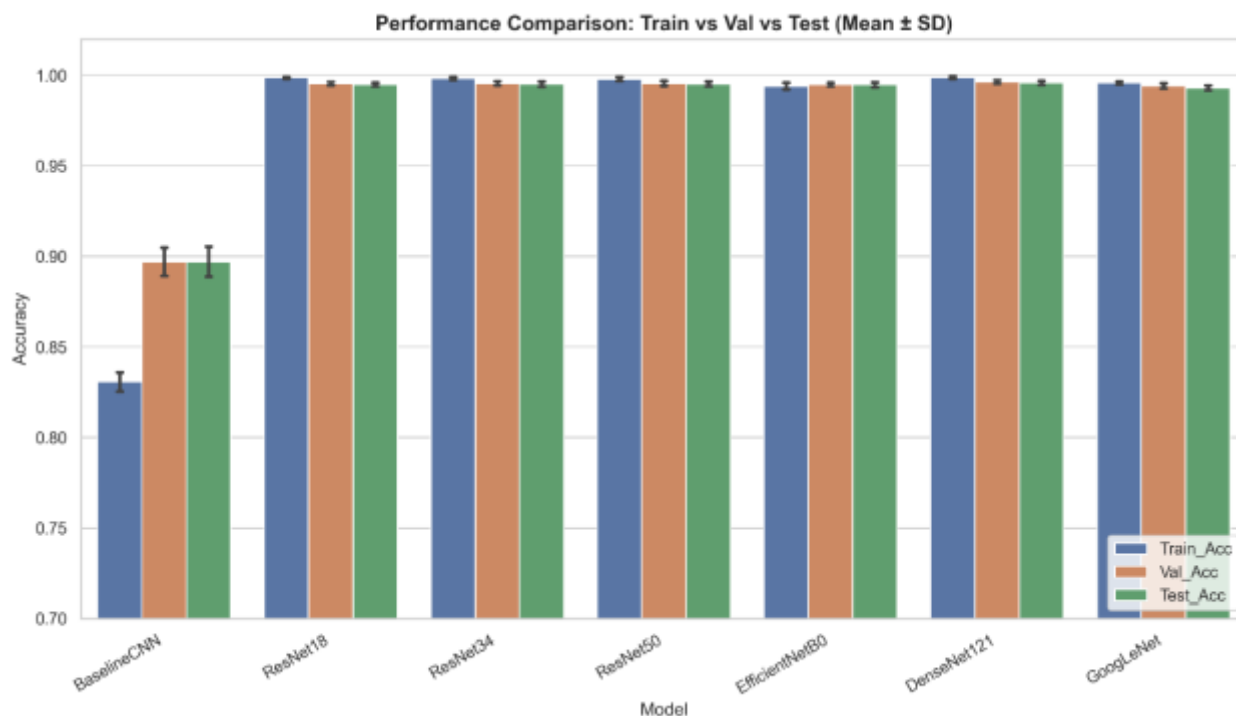
تحلیل توزیع خطای مجموعه آزمون

مطابق شکل ۶ (ج)، شبکه پیچشی پایه بالاترین مقدار خطای آزمون را با میانه‌ای در حدود ۰/۳۵ ثبت کرده و پراکندگی قابل توجهی در مقادیر خطا دارد. این موضوع به‌وضوح نشان می‌دهد که مدل نه تنها از نظر دقت، بلکه از نظر پایداری همگرایی و کیفیت یادگیری نیز عملکرد ضعیفی دارد. در مقابل، تمامی مدل‌های یادگیری انتقالی دارای خطای آزمون بسیار پایین (حدود ۰/۰۲ تا ۰/۰۳) و توزیعی فشرده هستند. DenseNet-121 و GoogLeNet کمترین میانه خطا را نشان می‌دهند که بیانگر یادگیری مؤثرتر نمایش‌های ویژگی و بهینه‌سازی پایدارتر است. همگرایی یکنواخت این مدل‌ها در اجراهای مستقل، نقش کلیدی معماری‌های عمیق و انتقال دانش را در کاهش خطا و افزایش اطمینان نتایج تأیید می‌کند.

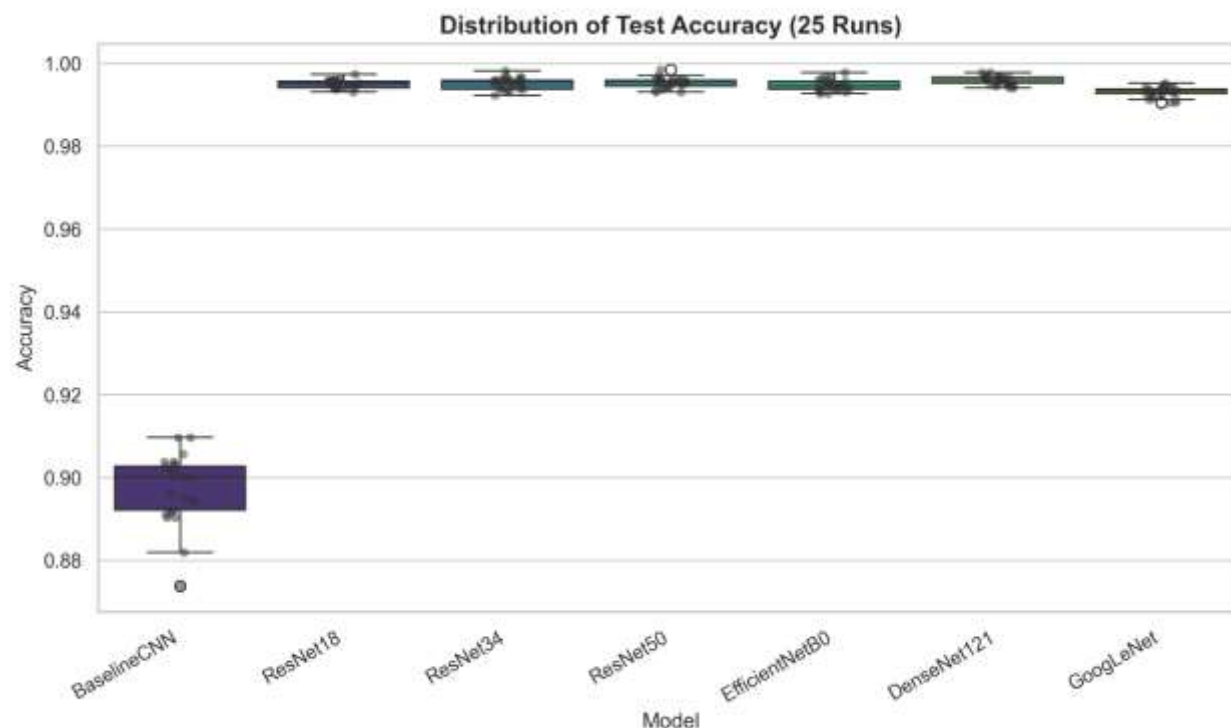
تحلیل واریانس و مقایسه آماری عملکرد مدل‌ها

رای ارزیابی آماری پایداری عملکرد مدل‌های مختلف در خطای اعتبارسنجی، میانگین حاصل از ۲۵ اجرای مستقل برای هر مدل، در قالب طرح آماری طرح بلوک کامل تصادفی (RCBD) مورد تجزیه و تحلیل قرار گرفت. نتایج تحلیل واریانس نشان داد که اثر تکرارها بر خطای اعتبارسنجی از نظر آماری معنادار است ($F = 5.022, df = 24, p < 0.001$) که نشان‌دهنده وجود نوسان بین اجراهای مستقل مدل‌ها است. مهم‌تر از آن، اثر خطای اعتبارسنجی مدل‌ها (تیمار)، که بیانگر تفاوت بسیار معنی‌دار حداقل دو مدل می‌باشد

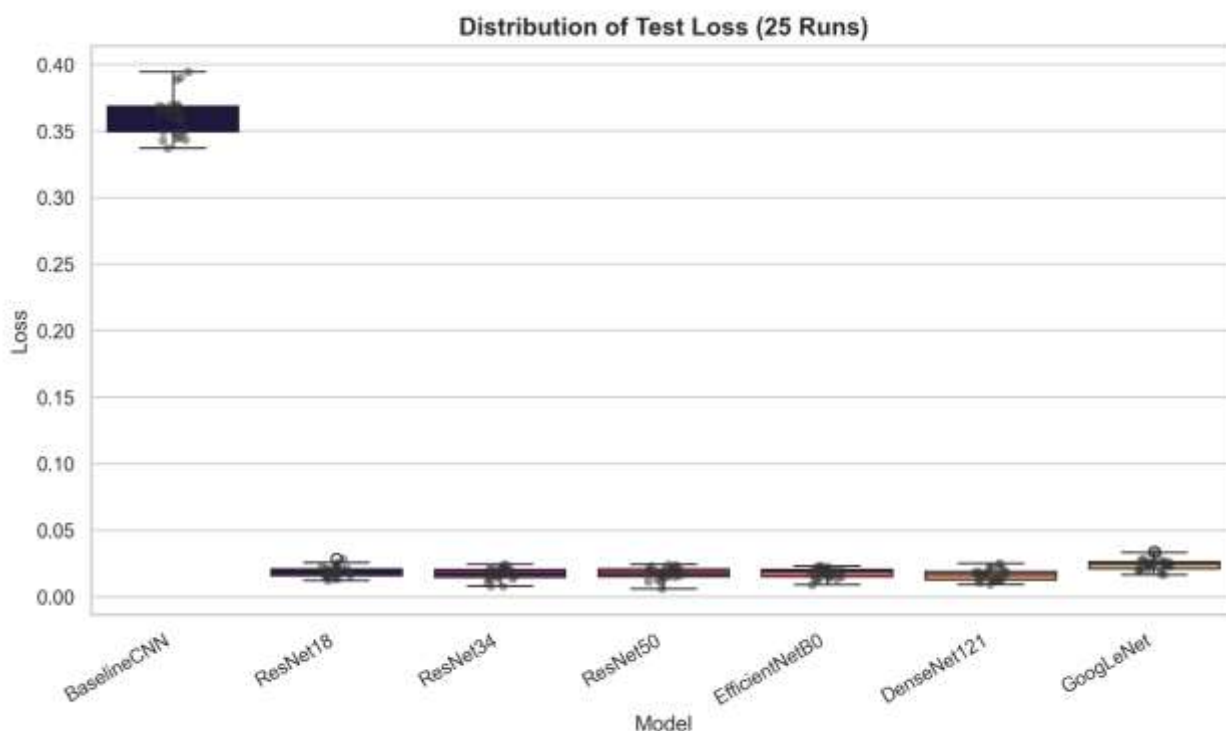
($F = 6106.244, df = 6, p < 0.001$). بر این اساس، مقایسه زوجی میانگین تیمارها با استفاده از آزمون تعقیبی (LSD) انجام شد تا مدل‌های برتر از نظر آماری شناسایی شوند؛ نتایج این مقایسه نشان می‌دهد خطای اعتبارسنجی مدل Baseline CNN به طور بسیار معنی‌دار از مدل‌های دیگر بیشتر است.



الف- مقایسه میانگین دقت آموزش، اعتبارسنجی و آزمون (میانگین $\pm$ انحراف معیار)



ب- توزیع دقت مجموعه آزمون مدل‌ها در ۲۵ اجرای مستقل



ج- توزیع خطای مجموعه آزمون مدل‌ها در ۲۵ اجرای مستقل
 شکل ۶- تحلیل مقایسه‌ای عملکرد و پایداری آماری مدل‌های یادگیری عمیق در مجموعه آزمون بر اساس ۲۵ اجرای مستقل

بحث

نتایج این پژوهش نشان می‌دهد که به کارگیری معماری‌های یادگیری انتقالی، در مقایسه با شبکه‌ی پیچشی آموزش‌دیده از ابتدا، نه تنها موجب بهبود دقت طبقه‌بندی شده است، بلکه از منظر پایداری آماری و قابلیت اعتماد در اجراهای مستقل نیز مزیت معناداری ایجاد کرده است. ارزیابی عملکرد مدل‌ها در چندین اجرای مستقل، یکی از جنبه‌های متمایز و کمتر گزارش‌شده در مطالعات مشابه است و امکان تحلیل عمیق‌تری از رفتار واقعی مدل‌ها فراتر از یک اجرای منفرد فراهم می‌کند.

در این چارچوب، مدل ResNet50 با دستیابی به دقت اعتبارسنجی ۸۶٪ و امتیاز F1 برابر با ۸۷/۶٪، توازن قابل قبول میان دقت و ثبات عملکرد نشان داد. در مقابل، شبکه‌ی پیچشی پایه که بدون استفاده از وزن‌های پیش‌آمورخته آموزش داده شد، با دقت ۴۹٪ عملکردی به مراتب ضعیف‌تر داشت. این شکاف قابل توجه، نشان‌دهنده‌ی آن است که آموزش شبکه‌های عمیق از ابتدا، در شرایط محدودیت داده و شباهت بالایی الگوهای بصری بیماری‌ها، با چالش‌های جدی در همگرایی و تعمیم‌پذیری مواجه است. این یافته، به‌ویژه زمانی اهمیت می‌یابد که رفتار مدل‌ها در اجراهای مستقل مورد بررسی قرار می‌گیرد؛ جایی که شبکه‌ی پایه نوسانات محسوسی در عملکرد نشان می‌دهد، در حالی که مدل‌های انتقالی از توزیع فشرده‌تری برخوردارند.

یکی از نوآوری‌های اصلی این پژوهش، تحلیل پایداری مدل‌ها در چندین اجرای مستقل است. برخلاف بسیاری از مطالعات که تنها یک بار آموزش و ارزیابی مدل را گزارش می‌کنند، در این تحقیق، توزیع دقت در اجراهای تکرارشونده نشان داد که مدل‌های مبتنی بر یادگیری انتقالی، به‌ویژه DenseNet-121، از انحراف معیار بسیار پایینی برخوردارند (۰/۰۰۱۱). این مقدار اندک بیانگر آن است که عملکرد مدل به مقداردهی اولیه وزن‌ها یا نوسانات تصادفی فرآیند آموزش وابستگی ناچیزی دارد. در مقابل، شبکه‌ی پیچشی پایه با انحراف معیار بالاتر (۰/۰۰۸۳) رفتار ناپایدارتری از خود نشان داد. این تفاوت، اهمیت یادگیری انتقالی را نه فقط از منظر دقت متوسط، بلکه از حیث قابلیت اعتماد عملی برجسته می‌کند؛ ویژگی‌ای که در کاربردهای واقعی کشاورزی، نقشی تعیین‌کننده دارد.

تحلیل ماتریس‌های اغتشاش نیز به‌طور روشنی تفاوت کیفی میان مدل‌ها را آشکار می‌سازد. در شبکه‌ی پیچشی پایه، خطاهای متقاطع قابل توجهی میان کلاس‌هایی با شباهت بافتی و رنگی بالا مشاهده می‌شود؛ به‌ویژه در بیماری‌هایی نظیر «سوختگی زودرس» و «لکه باکتریایی» که الگوهای ظاهری همپوشان دارند. این خطاها نشان می‌دهد که مدل پایه در استخراج ویژگی‌های عمیق و تمایزبخش ناتوان است. در مقابل، ماتریس‌های اغتشاش DenseNet-121 تقریباً به‌صورت قطری و بدون پراکندگی خطا هستند و بیانگر تفکیک

موفق تمامی کلاس‌ها می‌باشند. این رفتار کیفی، مکمل نتایج کمی بوده و نشان می‌دهد که بهبود عملکرد صرفاً محدود به افزایش دقت کلی نیست، بلکه شامل بهبود تفکیک کلاس‌های دشوار نیز می‌شود.

پس از تبیین نتایج تحقیق حاضر، تطبیق آن‌ها با مطالعات پیشین نشان می‌دهد که یافته‌ها در امتداد روندهای گزارش شده در ادبیات قرار دارند؛ اما با افزوده‌ای مهم در حوزه‌ی پایداری. Zhao et al. (۲۰۲۵) در مقاله‌ی مروری خود تأکید می‌کنند که مدل‌های CNN آموزش دیده از ابتدا، به دلیل شباهت بالای علائم بیماری‌ها، مستعد خطاهای متقاطع هستند و یادگیری انتقالی راهکاری کلیدی برای غلبه بر این محدودیت است. نتایج پژوهش حاضر، به‌ویژه در تحلیل ماتریس‌های اغتشاش، مصداق تجربی این ادعا را نشان می‌دهد.

مطالعه‌ی Nagamani و Sarojadevi (۲۰۲۲) حداکثر دقت ۹۶/۷٪ را برای R-CNN گزارش کرده است؛ اما تحلیل پایداری میان اجراها در آن مطالعه ارائه نشده است. در مقابل، تحقیق حاضر نشان می‌دهد که DenseNet-121 نه تنها به دقت بسیار بالا نزدیک به ۱۰۰٪ دست یافته، بلکه این عملکرد در اجراهای مستقل نیز به‌طور یکنواخت تکرار شده است. این تفاوت، نشان می‌دهد که ارزیابی پایداری می‌تواند تصویری واقع‌بینانه‌تر از قابلیت عملی مدل‌ها ارائه دهد.

Ali et al. (۲۰۲۴) بر اهمیت مدل‌های سبک و انتقالی برای سامانه‌های هوشمند کشاورزی تأکید کرده‌اند و نشان داده‌اند که ResNet و MobileNet در محیط‌های واقعی عملکرد باثبات‌تری دارند. یافته‌های این پژوهش نیز با این دیدگاه همسو است؛ اما با افزودن تحلیل اجرای‌های مستقل، نشان می‌دهد که حتی در میان مدل‌های انتقالی، تفاوت‌های معناداری از نظر ثبات وجود دارد.

در حوزه‌ی معماری‌های Transformer، Shehu و همکاران (۲۰۲۵) گزارش کردند که اگرچه ViT در دیتاست دهکده گیاه به دقت ۹۹/۱۷٪ دست یافته است؛ اما عملکرد آن در داده‌های میدانی به بازه‌ی ۷۷٪-۹۲٪ کاهش یافته است. این نوسان، اهمیت پایداری و تعمیم‌پذیری را برجسته می‌کند. نتایج تحقیق حاضر نشان می‌دهد که CNNهای انتقالی، به‌ویژه DenseNet-121، در کنار دقت بالا، از ثبات بیشتری برخوردارند؛ موضوعی که Zhao و همکاران (۲۰۲۵) نیز آن را به‌عنوان مزیت CNNها نسبت به Transformerها در کاربردهای واقعی مطرح کرده‌اند.

همچنین، نتایج این پژوهش با یافته‌های Ahmadi (۲۰۲۵) همخوانی دارد؛ جایی که EfficientNet پیش‌آموزش دیده، پس از حدود ۲۰ دوره آموزش به همگرایی پایدار رسیده و نوسانات عملکرد آن محدود گزارش شده است. با این حال، پژوهش حاضر با بررسی توزیع عملکرد در اجراهای مستقل و تحلیل ماتریس‌های اغتشاش، گامی فراتر از گزارش همگرایی صرف برداشته و تصویری جامع‌تر از رفتار مدل ارائه می‌دهد.

در مجموع، نوآوری اصلی این تحقیق در ترکیب ارزیابی کمی، تحلیل کیفی ماتریس‌های اغتشاش و بررسی پایداری آماری در اجراهای مستقل نهفته است. این رویکرد نشان می‌دهد که انتخاب مدل برای کاربردهای واقعی کشاورزی نباید صرفاً بر اساس بیشینه دقت یک اجرای منفرد انجام شود، بلکه باید ثبات، قابلیت اعتماد و رفتار کلاس به کلاس مدل نیز مدنظر قرار گیرد. بر این اساس، نتایج پژوهش حاضر ضمن هم‌راستایی با ادبیات موجود، سهمی افزوده در جهت ارائه‌ی چارچوبی قابل اتکا برای انتخاب مدل‌های تشخیص بیماری برگ گوجه‌فرنگی فراهم می‌کند.

مهم‌ترین محدودیت این پژوهش، استفاده از مجموعه داده دهکده گیاه است که تصاویر آن در شرایط کنترل شده و با پس‌زمینه یکنواخت ثبت شده‌اند. بنابراین انتظار می‌رود عملکرد مدل‌ها در شرایط واقعی مزرعه، شامل تغییرات نور، سایه‌ها، باد، لرزش دوربین و پس‌زمینه‌های پیچیده — پایین‌تر از مقادیر گزارش شده باشد. با وجود این محدودیت، نتایج به‌دست آمده می‌توانند به‌عنوان یک خط‌مبنای پایدار و قابل اتکا برای مطالعات آتی مورد استفاده قرار گیرند. ارزیابی مدل‌ها بر روی داده‌های میدانی و توسعه راهکارهای سازگار با محیط واقعی می‌تواند در ادامه پژوهش دنبال شود.

در کنار تحلیل دقت و پایداری آماری مدل‌ها، لازم است به این نکته نیز اشاره شود که اگرچه DenseNet-121 در میان مدل‌های بررسی شده پایدارترین رفتار را در ۲۵ اجرای مستقل نشان داد؛ اما به دلیل ساختار اتصالات متراکم، در مرحله استنتاج نسبت به مدل‌های سبک‌تر حافظه بیشتری مصرف می‌کند. البته پس از اتمام فرآیند آموزش، مرحله استنتاج همچنان بسیار سریع انجام می‌شود؛ با این حال، برای کاربردهای بلادرنگ و استقرار در دستگاه‌های کم‌منبع در محیط مزرعه، معماری‌های سبک‌تری مانند EfficientNet-B0 یا ResNet-18 می‌توانند گزینه‌های مناسب‌تری باشند. بنابراین هدف اصلی این پژوهش ارائه یک مدل نهایی برای استقرار میدانی نبوده، بلکه تمرکز بر ارزیابی دقیق پایداری و ارائه یک خط‌مبنای آماری قابل اتکا برای پژوهش‌های آینده بوده است. توسعه مدل‌های سبک‌تر و بهینه‌سازی شده برای پیاده‌سازی واقعی در شرایط مزرعه می‌تواند به عنوان مسیر آتی پژوهش دنبال شود.

نتیجه گیری

در این پژوهش، مجموعه‌ای متنوع از معماری‌های یادگیری عمیق شامل یک شبکه‌ی پیچشی آموزش‌دیده از ابتدا و چندین مدل شاخص مبتنی بر یادگیری انتقالی از خانواده‌های GoogLeNet، ResNet، EfficientNet و DenseNet بر روی مجموعه‌داده مرجع به صورت جامع مورد ارزیابی قرار گرفت. برخلاف بسیاری از مطالعات پیشین که صرفاً بر گزارش دقت نهایی یک اجرای منفرد تکیه دارند، تمرکز اصلی این تحقیق بر رفتار واقعی، پایداری آماری و قابلیت اعتماد مدل‌ها در اجراهای تکرارشونده بوده است.

نتایج نشان داد که شبکه‌ی پیچشی پایه، علی‌رغم دستیابی به دقت متوسط قابل قبول، از نوسان عملکرد بالاتری برخوردار است و در تفکیک کلاس‌های دارای شباهت بافتی و رنگی، خطاهای متقاطع قابل توجهی از خود نشان می‌دهد. این ضعف، هم در تحلیل منحنی‌های یادگیری و هم در ماتریس‌های اغتشاش به وضوح مشاهده شد و بیانگر محدودیت ذاتی آموزش از ابتدا در استخراج ویژگی‌های سطح بالا و تعمیم‌پذیر است. در مقابل، تمامی مدل‌های مبتنی بر یادگیری انتقالی عملکردی به مراتب پایدارتر و قابل اتکاتر ارائه کردند.

در میان مدل‌های بررسی شده، DenseNet-121 به طور معناداری به عنوان گزینه‌ی برتر شناسایی شد. این مدل نه تنها به بالاترین دقت آزمون و اعتبارسنجی (۰/۹۹۶) دست یافت، بلکه با کمترین انحراف معیار (۰/۰۰۱۱) در ۲۵ اجرای کاملاً مستقل، بالاترین سطح پایداری آماری را نیز نشان داد. فشردگی توزیع دقت DenseNet-121 در اجراهای تکراری بیانگر آن است که عملکرد این مدل به مقارن‌دهی اولیه تصادفی یا نوسانات فرآیند آموزش وابستگی ناچیزی دارد؛ ویژگی‌ای که برای کاربردهای عملی و سامانه‌های تصمیم‌یار کشاورزی اهمیت حیاتی دارد.

تحلیل ماتریس‌های اغتشاش نیز نشان داد که DenseNet-121 در تفکیک کلاس‌های دشوار، که در شبکه‌ی پایه بیشترین میزان اشتباه را داشتند، عملکردی تقریباً بدون خطا ارائه می‌دهد. این موضوع تأیید می‌کند که برتری مدل پیشنهادی صرفاً به افزایش دقت کلی محدود نمی‌شود، بلکه شامل بهبود کیفی در تشخیص کلاس‌به‌کلاس و کاهش خطاهای بحرانی نیز هست. سایر مدل‌های انتقالی نظیر ResNet50 و EfficientNet-B0 نیز دقت‌های بالایی نزدیک به DenseNet-121 کسب کردند؛ اما از نظر واریانس عملکرد، در رتبه‌های بعدی قرار گرفتند.

در مجموع، نتایج این پژوهش نشان می‌دهد که انتخاب مدل در مسائل تشخیص بیماری برگ گیاهان نباید صرفاً بر اساس بیشینه دقت یک اجرای منفرد انجام شود. پایداری آماری، رفتار یادگیری در اجراهای مستقل و الگوی خطاهای کلاس‌به‌کلاس معیارهایی کلیدی هستند که تصویر واقع‌بینانه‌تری از قابلیت اعتماد مدل ارائه می‌دهند. از این منظر، DenseNet-121 به عنوان مدلی که هم دقت بسیار بالا و هم ثبات آماری کم‌نظیری دارد، گزینه‌ای مناسب برای توسعه سامانه‌های هوشمند تشخیص بیماری برگ گوجه‌فرنگی محسوب می‌شود. نوآوری اصلی این تحقیق در ارائه‌ی یک چارچوب ارزیابی فراتر از مقایسه‌ی دقت، و تمرکز نظام‌مند بر تحلیل پایداری آماری مبتنی بر ۲۵ اجرای مستقل به همراه ارزیابی کیفی ماتریس‌های اغتشاش نهفته است. این رویکرد نشان می‌دهد که چگونه می‌توان با استفاده از تحلیل‌های تکرارشونده و چندبعدی، مدل‌هایی با قابلیت اعتماد بالاتر برای کاربردهای واقعی کشاورزی انتخاب کرد؛ موضوعی که در بخش عمده‌ای از ادبیات موجود مغفول مانده و می‌تواند مسیر پژوهش‌های آینده در این حوزه را به طور معناداری جهت‌دهی کند. با اینکه DenseNet-121 از نظر پایداری آماری در میان مدل‌های بررسی شده برتری دارد، ولیکن به دلیل مصرف حافظه بیشتر، برای کاربردهای بلادرنگ در محیط‌های کم‌منبع مناسب‌ترین گزینه نیست و می‌توان مدل EfficientNet-B0 را پیشنهاد نمود.

پیشنهادهای

بر اساس یافته‌های این تحقیق و با توجه به شکاف‌های شناسایی شده، پیشنهادهای زیر برای مطالعات آتی ارائه می‌شود:

ادغام مکانیزم‌های توجه: به منظور بهبود دقت در تشخیص بیماری‌های با علائم مشابه (مانند لکه باکتریایی و سوختگی زودرس)، پیشنهاد می‌شود مکانیزم‌های توجه در معماری مدل‌ها ادغام شوند تا تمرکز مدل به طور خودکار بر روی نواحی بیمار و مرتبط برگ معطوف گردد.

استفاده از معماری‌های پیشرفته‌تر: بررسی قابلیت‌های مدل‌های ترنسفورمر بینایی و معماری‌های هیبریدی جدید مانند ConvNeXt برای این مسئله خاص، با هدف دستیابی به دقت‌های بالاتر و تحلیل بهتر ویژگی‌های سطح بالا پیشنهاد می‌شود.

توسعه مدل برای داده‌های واقعی مزرعه: آموزش و ارزیابی مدل بر روی مجموعه‌داده‌هایی که در شرایط واقعی مزرعه و با

حضور عوامل مخدوش‌کننده مانند نورپردازی متغیر، سایه، خاک و تنوع پس‌زمینه گرفته شده‌اند، برای افزایش قابلیت اجرا در دنیای واقعی ضروری است.

استقرار مدل‌های تشخیص بیماری بر روی سخت‌افزارهای لبه: از جمله تلفن‌های هوشمند، پهپادها یا بردهای محاسباتی کم‌هزینه مورد استفاده در ماشین‌های کشاورزی، می‌تواند امکان پایش بلادرنگ بیماری‌ها در شرایط واقعی مزرعه را فراهم کند. با این حال، محیط‌های واقعی مزرعه با چالش‌هایی نظیر تغییرات شدید نورپردازی، حرکت برگ‌ها در اثر باد، لرزش دوربین در پهپادها یا ماشین‌های کشاورزی، و وجود پس‌زمینه‌های پیچیده شامل خاک، سایه‌ها و سایر گیاهان همراه است؛ شرایطی که اغلب در مجموعه داده‌های آزمایشگاهی مانند PlantVillage به طور کامل بازنمایی نمی‌شوند. از این رو، علاوه بر ارزیابی مدل‌ها در چنین شرایطی، بهینه‌سازی مدل‌های منتخب برای استقرار روی سخت‌افزارهای لبه نیز ضروری است. در این راستا می‌توان از روش‌هایی نظیر کمی‌سازی مدل و هرس شبکه برای کاهش حجم و مصرف حافظه مدل‌های عمیق مانند DenseNet استفاده کرد تا امکان اجرای کارآمد آن‌ها بر روی دستگاه‌های با منابع محدود فراهم شود.

تبدیل به سامانه تشخیص عینی: توسعه مدل از یک سامانه طبقه‌بندی صرف به یک سامانه تشخیص عینی (مانند YOLO یا R-CNN) که توانایی شناسایی و مکان‌یابی همزمان چندین ضایعه و بیماری را بر روی یک برگ دارد، می‌تواند اطلاعات دقیق‌تری برای تصمیم‌گیری فراهم کند.

ملاحظات اخلاقی

حامی مالی

این پژوهش از هیچ‌گونه کمک مالی خاصی از سوی سازمان‌های حمایت‌کننده در بخش‌های دولتی، تجاری یا غیرانتفاعی بهره‌مند نشده است.

بیانیه دسترسی به داده‌ها

داده‌هایی پژوهش حاضر از طریق درخواست از نویسندگان قابل دسترسی است.

سپاسگزاری

از معاونت محترم پژوهشی دانشگاه تهران به خاطر حمایت معنوی در اجرای پژوهش حاضر سپاسگزاری می‌شود. از داوران محترم به خاطر ارائه نظرهای ساختاری و علمی سپاسگزاری می‌شود.

پیروی از اصول اخلاق پژوهش

نویسندگان اصول اخلاقی را در انجام و انتشار این پژوهش علمی رعایت نموده‌اند و این موضوع مورد تأیید همه آنهاست.

تعارض منافع

بنا بر اظهار نویسندگان این مقاله تعارض منافع ندارد.

منابع

- اسدی، حمید. (۱۴۰۴). شناسایی و طبقه‌بندی بیماری‌های برگ گوجه‌فرنگی با کمک شبکه عصبی کانولوشنی عمیق. مقالات همایش توسعه پایدار در کشاورزی، علوم دام و صنایع غذایی با رویکرد ارتقاء کیفیت، سلامت و امنیت غذایی. تهران.
- افشاری‌پور، مجتبی و شمسی، محسن (۱۴۰۱). مروری بر توسعه تکنیک‌های یادگیری عمیق در پردازش تصاویر برای تشخیص بیماری‌های محصولات کشاورزی. چهاردهمین کنگره ملی مهندسی مکانیک بیوسیستم و مکانیزاسیون ایران. <https://civilica.com/doc/1535909>
- آق‌آتابای، حبیب، شیخ‌زاده، محمد جواد و ترشیزی، مهدی (۱۳۹۵). تشخیص بیماری‌های برگ گیاه گوجه‌فرنگی از روی تصویر با استفاده از یادگیری عمیق. همایش ملی پایش و پیش‌آگاهی در گیاه پزشکی. <https://sid.ir/paper/876674/fa>
- پوردربانی، راضیه، و سبزی، سجاد (۱۴۰۲). تشخیص بیماری‌های رایج گل‌کلم با استفاده از پردازش تصویر و یادگیری عمیق. نشریه مطالعات علوم



محیط زیست ۸(۳)، ۷۰۸۷-۷۰۹۲.

- پیمان، سید حسین، بخشی پور یارتگاهی، عادل و جعفری، عبدالعباس (۱۳۹۵). بررسی امکان به کارگیری پردازش تصویر جهت تشخیص بیماری های برگ برنج. ماشین های کشاورزی ۶(۱)، ۶۹-۷۹.
- حسینی، هادی، محمدزamani، داود و ارباب، عباس (۱۳۹۶). سامانه تشخیص بیماری قارچی سفیدک پودری و آنتراکنوز برگ خیار با تکنیک پردازش تصویر و شبکه عصبی مصنوعی. گیاه پزشکی ۴۰(۴)، ۱۵-۲۸.
- روستایی، م. و نوروزی، م. (۱۴۰۳). تحلیل روش های یادگیری انتقالی در بهینه سازی تشخیص بیماری های درخت سیب. علوم و فناوری اطلاعات کشاورزی، ۷(۱-۱۳)، ۲۳-۳۳. <https://doi.org/10.22092/jaist.2024.364053.1105>
- سلیمانی، اعظم، رضوی، سید محمد، روحانی، حمید، خانی مقدم و عصمت مهدی (۱۳۹۵). تشخیص بیماری سفیدک پودری هلو با استفاده از پردازش تصویر و شبکه عصبی. کنگره گیاه پزشکی ایران.
- شیشه چی، سمن (۱۴۰۳). بررسی روش های تشخیص بیماری گیاهی با استفاده از یادگیری عمیق. اولین کنفرانس ملی محیط زیست، آب و انرژی پاک.
- قاسمی و روحانی، زهره، محتسبی، سیدسعید، قاسمی، هادی و عمرانی، الهام (۱۳۹۷). توسعه یک سامانه ترکیبی جدید به منظور تشخیص بیماری های برگ درخت سیب. مهندسی بیوسیستم ایران (علوم کشاورزی ایران)، ۴۹(۲)، ۲۱۵-۲۲۵. SID. <https://sid.ir/paper/144198/fa>
- کریمی، علی و نادری فر، وحیده (۱۴۰۱). روشی جدید برای تبدیل طبقه بندی بیماری های گیاهی با استفاده از موجک. سیستم های اطلاعاتی هوشمند ۴۱(۴)، ۴۹-۴۱.
- نادری بنی، امین، باقرپور، حسین و امیری پریان، جعفر (۱۴۰۳). توسعه و بهینه سازی الگوریتم ویژه یادگیری عمیق در تشخیص بیماری های مختلف برگی درخت به ماشین های کشاورزی ۱۴(۴)، ۴۴۵-۴۵۸.
- نجف آبادی ها، محسن، محمدزamani، داود و غلامی پرشکوهی، محمد (۱۴۰۲). تشخیص سه نوع بیماری برگ انگور بر پایه پردازش تصویر با استفاده از الگوریتم بهینه ساز پروانه و ماشین بردار پشتیبان. تحقیقات مهندسی سامانه ها و مکانیزاسیون کشاورزی ۲۴(۸۷)، ۳۹-۵۴.
- نصیری، سجاد، و خجسته نژاد، مصطفی (۱۴۰۱). روش مبتنی بر پردازش تصویر به منظور تشخیص خودکار بیماری برگ انگور. مهندسی بیوسیستم ایران ۵۳(۱)، ۶۱-۷۶.

REFERENCES

- Assadi, H. (2025). Identification and classification of tomato leaf diseases using deep convolutional neural networks. In Proceedings of the Conference on Sustainable Development in Agriculture, Animal Science, and Food Industries with an Emphasis on Quality Improvement, Health, and Food Security. Tehran, Iran. (In Persian).
- Afsharipour, M., & Shamsi, M. (2022). A review of deep learning technique developments in image processing for crop disease detection. 14th National Congress of Biosystems Mechanical Engineering and Mechanization of Iran. <https://civilica.com/doc/1535909>. (In Persian).
- Agarwal, M., Singh, A., Arjaria, S., Sinha, A., & Gupta, S. (2020). ToLeD: Tomato leaf disease detection using convolution neural network. *Procedia Computer Science*, 167, 293-301.
- Agh-Atabai, H. H., Sheikhzadeh, M. J., & Torshizi, M. (2016). Detection of tomato leaf diseases from images using deep learning. National Conference on Monitoring and Forecasting in Plant Protection. <https://sid.ir/paper/876674/fa>. (In Persian).
- Ahmadi, I. (۲۰۲۵). Detection and Classification of Some Diseases of Tomato Crops Using. Transfer Learning. *Journal of Agricultural Machinery*, ۱۵(۳)، ۳۱۹-۳۳۵. <https://doi.org/10.22067/jam.2024.88500.1258>
- Ali, A. M., Słowik, A., Hezam, I. M., & Abdel-Basset, M. (2024). Sustainable smart system for vegetables plant disease detection: Four vegetable case studies. *Computers and Electronics in Agriculture*, 227(Part 2), 109672. <https://doi.org/10.1016/j.compag.2024.109672>
- Atila, Ü., Uçar, M., Akyol, K., & Uçar, E. (2021). Plant leaf disease classification using EfficientNet deep learning model. *Ecological Informatics*, 61, 101182. <https://doi.org/10.1016/j.ecoinf.2020.101182>
- Azadshahraki, F., K. Sharifi, B. Jamshidi, R. Karimzadeh, and H. Naderi. (2022). Diagnosis of Early Blight Disease in Tomato Plant based on Visible/Near-Infrared Spectroscopy and Principal Components Analysis- Artificial Neural Network Prior to Visual Disease Symptoms. *Journal of Agricultural Machinery* 12 (1): 81-94. DOI: 10.22067/jam.2021.67436.1001
- Chungcharoen, T., Donis-Gonzalez, I., Phetpan, K., Udompetaiikul, V., Sirisomboon, P., & Suwalak, R. (2022). Machine learning-based prediction of nutritional status in oil palm leaves using proximal multispectral images. *Computers and Electronics in Agriculture*, 198,

107019. <https://doi.org/10.1016/j.compag.2022.107019>
- Das, A., Pathan, F., Kabir, M. M., Jim, J. R., Mridha, M. F., & Ouishy, M. R. (2025a). XLTLDisNet: A novel and lightweight approach to identify tomato leaf diseases with transparency. *Heliyon*, 11, e42575.
- Das, A., Pathan, F., Jim, J. R. M., Kabir, M. M., & Mridha, M. F. (2025b). Deep learning-based classification, detection, and segmentation of tomato leaf diseases: A state-of-the-art review. *Artificial Intelligence in Agriculture*, 15, 192–220. <https://doi.org/10.1016/j.aiaa.2025.02.006>
- Durmus, H., Gunes, E. O., & Kirci, M. (2017). Disease detection on the leaves of the tomato plants by using deep learning. In *2017 6th International Conference on Agro-Geoinformatics* (pp. 1-5). IEEE.
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861–874. <https://doi.org/10.1016/j.patrec.2005.10.010>
- Ferentinos, K. P. (2018). Deep learning models for plant disease detection and diagnosis. *Computers and Electronics in Agriculture*, 145, 311–318. <https://doi.org/10.1016/j.compag.2018.01.009>
- Fuentes, A., Yoon, S., Kim, S. C., & Park, D. S. (2022). A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition. *Sensors*, 17(8), 17.
- Ghanbari, F., Gonbadi, H., & Saberi, M. (2022). Application of deep neural networks in tomato leaf disease detection. *Iranian Journal of Agricultural Science and Technology*, 53(1), 28–42. (In Persian)
- Ghasemi Varjani, Z., Mohtasebi, S. S., Ghasemi, H., & Omrani, E. (2018). Development of a new hybrid system for apple leaf disease detection. *Iranian Biosystems Engineering (Iranian Journal of Agricultural Sciences)*, 49(2), 215–225. <https://sid.ir/paper/144198/fa>. (In Persian).
- Ghosh, S. K., & Ghosh, A. (2022). ENResNet: A novel residual neural network for chest X-ray enhancement based COVID-19 detection. *Biomedical Signal Processing and Control*, 72, 103286. <https://doi.org/10.1016/j.bspc.2021.103286>
- Gilligan, C. A. (2008). Sustainable agriculture and plant diseases: an epidemiological perspective. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1492), 741–759.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 770–778). <https://doi.org/10.1109/CVPR.2016.90>
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778. <https://doi.org/10.1109/CVPR.2016.90>
- He, Z., & Tong, M. (2025). LT-YOLO: A lightweight network for detecting tomato leaf diseases. *Computers, Materials and Continua*, 82(3), 4301–4317. <https://doi.org/10.32604/cmc.2025.060550>
- Hong, H., Lin, J., & Huang, F. (2020). Tomato disease detection and classification by deep learning. In *2020 International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE)* (pp. 25-29). IEEE.
- Hosseini, H., Mohammadzamani, D., & Arbab, A. (2017). System for detecting powdery mildew and anthracnose fungal diseases of cucumber leaves using image processing and artificial neural networks. *Plant Protection*, 40(4), 15–28. (In Persian).
- Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. (2017). Densely connected convolutional networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 4700–4708. <https://doi.org/10.1109/CVPR.2017.243>
- Karimi, A., & Naderifar, V. (2022). A novel method for transforming plant disease classification using wavelets. *Intelligent Information Systems*, 1(4), 41–49. (In Persian).
- Karthik, R., Hariharan, M., Anand, S., Mathikshara, P., Johnson, A., & Menaka, R. (2020). Attention embedded residual CNN for disease detection in tomato leaves. *Applied Soft Computing*, 86, 105933.
- Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. *ICLR*.
- Kumar, N. S., Sony, J., Premkumar, A., Meenakshi, R., & Nair, J. J. (2024). Transfer learning-based object detection models for improved diagnosis of tomato leaf disease. *Procedia Computer Science*, 235, 3025–3034. <https://doi.org/10.1016/j.procs.2024.04.286>
- Matthews, B. W. (1975). Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica et Biophysica Acta (BBA) – Protein Structure*, 405(2), 442–451. [https://doi.org/10.1016/0005-2795\(75\)90109-9](https://doi.org/10.1016/0005-2795(75)90109-9)
- Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for image-based plant disease detection. *Frontiers in Plant Science*, 7, 1419. <https://doi.org/10.3389/fpls.2016.01419>
- Naderibani, A., Bagherpour, H., & Amiri-Piryan, J. (2024). Development and optimization of a specialized



- deep learning algorithm for detecting various leaf diseases of quince trees. *Agricultural Machinery*, 14(4), 445–458. (In Persian).
- Nagamani, H. S., & Sarojadevi, H. (2022). Tomato leaf disease detection using deep learning techniques. *International Journal of Advanced Computer Science and Applications*, 13(1), 305–311.
- Najafabadiha, M., Mohammadzamani, D., & Gholami-Parshkouhi, M. (2023). Detection of three grape leaf diseases based on image processing using moth-flame optimizer and support vector machine. *Research in Agricultural Systems Engineering and Mechanization*, 24(87), 39–54. (In Persian).
- Nasiri, S., & Khojasteh-Nezhad, M. (2022). Image-processing-based method for automatic grape leaf disease detection. *Iranian Biosystems Engineering*, 53(1), 61–76. (In Persian).
- Odusami, M., Misra, S., Abayomi-Alli, O., Adigun, M., Crawford, B., & Soto, J. D. (2021). Analysis of features of Alzheimer's disease: Detection of early stage from functional brain changes in magnetic resonance images using a finetuned ResNet-18 network. *Diagnostics*, 11(6), Article 1071. <https://doi.org/10.3390/diagnostics11061071>
- Paymode, A. S., & Malode, V. B. (2022). Transfer learning for multi-crop leaf disease image classification using convolutional neural network VGG. *Artificial Intelligence in Agriculture*, 6, 23–33.
- Peyman, S. H., Bakhshipour Ziaratgahi, A., & Jafari, A. (2016). Feasibility of using image processing for rice leaf disease detection. *Agricultural Machinery*, 6(1), 69–79. (In Persian).
- Pourderbani, R., & Sabzi, S. (2023). Detection of common cauliflower diseases using image processing and deep learning. *Journal of Environmental Sciences Studies*, 8(3), 7087–7092. (In Persian).
- PyTorch. (2023). *ResNet: Torchvision main documentation*. <https://pytorch.org/vision/main/models/resnet.html>
- Qabulio, M., Memon, M.S., Iqbal, S., Kumar, P., & Tsetse, A. (2024). Effective tomato leaf disease identification model using MobileNetV3Small. *International Journal of Information Systems and Computer Technologies*, 3(1), 57–72.
- Quan, S., Wang, J., Jia, Z., Yang, M., & Xu, Q. (2023). MS-Net: A novel lightweight and precise model for plant disease identification. *Frontiers in Plant Science*, 14, 1276728. <https://doi.org/10.3389/fpls.2023.1276728>
- Reimers, N., & Gurevych, I. (2017). Reporting score distributions makes a difference: Performance study of LSTM networks for sequence tagging. *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 338–348. <https://aclanthology.org/D17-1035/>
- Rezaei, M., Diepeveen, D., Laga, H., Jones, M. G. K., & Sohel, F. (2024). Plant disease recognition in a low data scenario using few-shot learning. *Computers and Electronics in Agriculture*, 219, 108812. <https://doi.org/10.1016/j.compag.2024.108812>
- Rostaei, M., & Norouzi, M. (2024). Analysis of transfer learning methods in optimizing apple tree disease detection. *Journal of Agricultural Information Science and Technology*, 7(13–1), 23–33. <https://doi.org/10.22092/jaist.2024.364053.1105>. (In Persian)
- Ruhad, F. M., Fahim, M., Hossain, M. S., Monir, M. F., Islam, A., & Amin, M. A. (2025). Beyond classification: Benchmarking object detection models for efficient tomato leaf disease identification on a real-world dataset. *Smart Agricultural Technology*, 12, 101336. <https://doi.org/10.1016/j.atech.2025.101336>
- Shanthi, D. L., Vinutha, K., Ashwini, N., & Vashistha, S. (2024). Tomato Leaf Disease Detection Using CNN. *Procedia Computer Science*, 235, 2975–2984.
- Shehu, H. A., Ackley, A., Marvellous, M., & Eteng, O. E. (2025). Early detection of tomato leaf diseases using transformers and transfer learning. *European Journal of Agronomy*, 168, 127625. <https://doi.org/10.1016/j.eja.2025.127625>
- Shishechi, S. (2024). A review of plant disease detection methods using deep learning. 1st National Conference on Environment, Water, and Clean Energy. (In Persian).
- Soleimani Dizicheh, E., Razavi, S. M., Rouhani, H., & Khani Moghaddam, A. M. (2016). Detection of peach powdery mildew disease using image processing and neural networks. *Iranian Plant Protection Congress*. (In Persian).
- Sun, H., Nicholas, I. T., Fu, R., & Kang, D.-K. (2024). YOLO-FMDI: A Lightweight YOLOv8 Focusing on a Multi-Scale Feature Diffusion Interaction Neck for Tomato Pest and Disease Detection. *Electronics*, 13(15), 2974.
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2015). Going deeper with convolutions. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1–9. <https://doi.org/10.1109/CVPR.2015.7298594>
- Tan, M., and Le, Q. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks.

- Proceedings of the 36th International Conference on Machine Learning (ICML)*, 6105–6114.
<https://doi.org/10.48550/arXiv.1905.11946>
- Too, E. C., Yujian, L., Njuki, S., & Yingchun, L. (2019). A comparative study of fine-tuning deep learning models for plant disease identification. *Computers and Electronics in Agriculture*, 161, 272-279.
- Zhang, Y., Huang, S., Zhou, G., Hu, Y., & Li, L. (2023). Identification of tomato leaf diseases based on multi-channel automatic orientation recurrent attention network. *Computers and Electronics in Agriculture*, 205, 107605. <https://doi.org/10.1016/j.compag.2022.107605>
- Zhang, Y., Wang, L., & Zhao, X. (2020). ResNet-based tomato leaf disease classification. *IEEE Access*, 8, 196076-196087.
- Zhao, J., Xu, L., Ma, Z., Li, J., Wang, X., Liu, Y., & Du, X. (2025). A review of plant leaf disease identification by deep learning algorithms. *Frontiers in Plant Science*, 16, 1637241. <https://doi.org/10.3389/fpls.2025.1637241>
- Zhao, X., Liu, Q., & Zhang, Y. (2022). Customized EfficientNet models for tomato leaf disease identification. *Expert Systems with Applications*, 200, 117084.
- Zhao, Z. Q., Zheng, P., Xu, S. T., & Wu, X. (2019). Object detection with deep learning: A review. *IEEE Transactions on Neural Networks and Learning Systems*, 30(11), 3212–3232. <https://doi.org/10.1109/TNNLS.2018.2876865>.