Journal of Soft Computing and Information Technology (JSCIT)
Babol Noshirvani University of Technology, Babol, Iran
Journal Homepage: jscit.nit.ac.ir
Volume 13, Number 1, Spring 2024, pp. 39-51
Received: 05/05/2024, Revised: 09/08/2024, Accepted: 10/03/2024

Early Detection of Plant Diseases Using Image Processing and Machine Learning Techniques

Sekineh Asadi Amiri [1*], Hassan Katal[2]
1*- Dept. Computer Engineering, University of Mazandaran, Babolsar, Iran.
2- Dept. Computer Engineering, University of Mazandaran, Babolsar, Iran.
[1*]s.asadi@umz.ac.ir, [2]h.katal10@umail.umz.ac.ir

Corresponding author's address: Sekineh Asadi Amiri, Faculty of Engineering and Technology, University of Mazandaran, Babolsar, Iran

Abstract- Early detection of plant diseases is crucial for timely prevention of pest infestation. In this paper plant diseases are accurately identified through image processing of leaves and extracting various features such as color, texture, and shape. Generally, plant diseases cause significant changes in the color and appearance of leaves. The color feature can be used to determine the type of plant disease. This paper employs moment and Hu moment features for extracting color characteristics from leaves, and area and perimeter features for detecting shape changes. To enhance the proposed features, additional texture features such as Local Binary Pattern (LBP), co-occurrence matrix, and vein patterns are utilized. Finally, to prevent model complexity and overfitting, the mRMR feature selection method is applied to eliminate irrelevant features. After developing the final model, healthy and diseased samples are classified using Random Forest and Support Vector Machine (SVM) classifiers. The proposed method has been evaluated on the PlantVillage dataset with 87867 samples. The average accuracy of the proposed method in diagnosing diseases in 9 different plant samples using Random Forest and SVM classifiers is 97.63% and 94.98%, respectively, which surpasses existing methods.

Keywords- Plant Disease Detection, Texture Features, Local Binary Pattern, Co-occurrence Matrix, Random Forest Classifier.

مجله علمی رایانش نرم و فناوری اطلاعات
دانشگاه صنعتی نوشیروانی بابل
صفحه مجله: jscit.nit.ac.ir
جلد ۱۳، شماره ۱، بهار ۱۴۰۳ ، صفحه ۳۹-۵۱
دریافت: ۱۴۰۳/۰۲/۱۶، بازنگری: ۱۴۰۳/۰۶/۱۸، پذیرش: ۱۴۰۳/۰۷/۱۲

تشخیص زودرس بیماری گیاهان با استفاده از تکنیک‌های پردازش تصویر و یادگیری ماشین

سکینه اسدی امیری*[۱]، حسن کتال[۲]
*۱- دانشکده مهندسی و فناوری، دانشگاه مازندران، بابلسر، ایران.
۲- دانشکده مهندسی و فناوری، دانشگاه مازندران، بابلسر، ایران.
[1*]s.asadi@umz.ac.ir, [2]h.katal10@umail.umz.ac.ir

* نشانی نویسنده مسئول: سکینه اسدی امیری، بابلسر، دانشگاه مازندران، دانشکده مهندسی و فناوری.

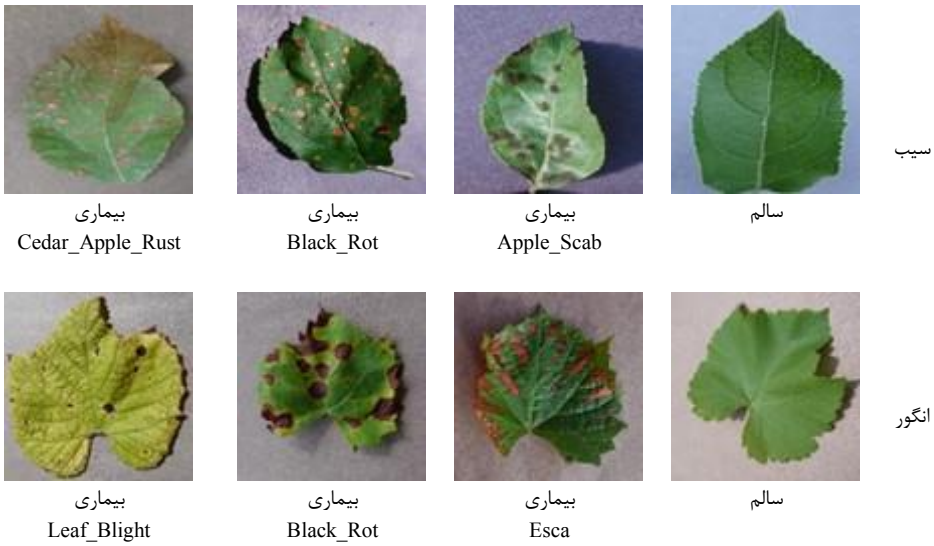
چکیده- تشخیص زود هنگام بیماری گیاهان، امری موثر در جلوگیری به موقع از آفت‌زدگی آنهاست. در این پژوهش، با استفاده از پردازش تصاویر برگ و استخراج ویژگی‌های مختلف مانند رنگ، بافت و شکل، بیماری‌های گیاه با دقت[۱] مناسب تشخیص داده می‌شوند. به طور کلی بیماری گیاهان باعث تغییرات اساسی در رنگ و فرم ظاهری برگ آنان می‌شود. با ویژگی رنگ می‌توان نوع بیماری گیاه را تشخیص داد. در این مقاله از ویژگی‌های ممان و گشتاور برای استخراج ویژگی رنگ از برگ و از ویژگی‌های مساحت و محیط برای تشخیص تغییر فرم آن استفاده شده است. در روش پیشنهادی، به منظور تقویت ویژگی‌های مطرح شده، از ویژگی‌های بافتی از جمله الگوی دودویی محلی، ماتریس هم‌رخداد و الگوی رگبرگ‌ها نیز استفاده شده است. در نهایت برای جلوگیری از پیچیدگی و بیش برازش مدل، از روش انتخاب ویژگی mRMR، برای حذف ویژگی‌های غیر مفید استفاده شده است. پس از ایجاد مدل نهایی، نمونه‌های سالم و بیمار با استفاده از کلاس بندهای جنگل تصادفی و ماشین بردار پشتیبان، دسته بندی شده است. روش پیشنهادی روی مجموعه داده PlantVillage با ۸۷۸۶۷ نمونه ارزیابی شده است. میانگین دقت روش پیشنهادی در تشخیص بیماری ۹ نمونه گیاه مختلف با کلاس بندهای جنگل تصادفی و ماشین بردار پشتیبان به ترتیب ٪ ۹۷/۶۳ و ٪ ۹۴/۹۸ است که بالاتر از روش‌های موجود بوده است.

واژه‌های کلیدی: تشخیص بیماری گیاهان، ویژگی‌های بافت، الگوی دودویی محلی، ماتریس هم‌رخداد، دسته‌بند جنگل تصادفی.

۱- مقدمه

بازده محصولات کشاورزی به عوامل زیادی مانند میزان بارش باران، کیفیت خاک، شرایط آب و هوایی و بیماری گیاهان وابسته است [۱]. یکی از بزرگ‌ترین مشکلات در این زمینه، آفت زدگی گیاهان است که می‌تواند سبب کاهش کیفیت و کمیت محصول گردد. بنابراین شناسایی به موقع بیماری گیاهان آفت زده، به منظور جلوگیری از چنین زیان‌هایی، حیاتی است [۲]. در بسیاری از موارد، نشانه‌های بیماری گیاهان در برگ، ساقه و میوه‌های آن قابل رویت است. در این پژوهش، تمرکز بر برگ گیاهان است. روش‌های سنتی تشخیص بیماری گیاهان که در گذشته مورد استفاده قرار گرفته اند، به دلیل استفاده از تعداد زیادی نیروی کار متخصص و آموزش دیده و همچنین ابزارهای گوناگون، افزون بر دشواری، ناگزیر از صرف هزینه و زمان زیادی بوده است. اما در روش پیشنهادی، با استفاده از فرآیندهای استخراج ویژگی در پردازش تصویر، همچنین دسته‌بندی و تشخیص با مدل‌های یادگیری ماشین، فرآیندی به مراتب آسان‌تر، سریع‌تر، ارزان‌تر و با دقت بالا ارائه شده است.

در گذشته محققان در این زمینه از روش‌های متنوع یادگیری ماشین [۳]، شبکه‌های عصبی [۴] و پردازش تصویر [۵] بهره



شکل ۱: نمونه‌ای از تصاویر برگ گیاهان سالم و بیمار برای دو گیاه سیب و انگور.

می‌بردند. در روش پیشنهادی، با استخراج ویژگی‌های تاثیرگذار در تشخیص بیماری گیاهان، از هر سه دسته ویژگی رنگ، بافت و ساختار برای تشخیص دقیق‌تر بیماری استفاده شده است.

معمولا برگ گیاهان آفت زده از نظر رنگ، شکل، همواری سطح و حالت یا تعداد رگبرگ‌ها تغییرات زیادی خواهند داشت [۶]. بنابراین با استفاده از ویژگی‌های رنگ، بافت و شکل می‌توان تا اندازه زیادی بین برگ نمونه‌های آفت زده و سالم تمایز ایجاد کرد. برای این منظور در روش پیشنهادی، ابتدا تصویر با استفاده از فیلتر گاوسی هموار و رفع نویز شده است. در ادامه با استفاده از روش آستانه‌گذاری اوتسو[۲] [۷] تصویر را قطعه‌بندی کرده و ناحیه برگ استخراج شده است. برای دقت بیشتر با عملگرهای هم‌ریختی[۳]، حفره‌های احتمالی حذف شده است. در گام بعد ویژگی‌های رنگ شامل ممان و گشتاور در فضاهای رنگی RGB، HSV و LAB بدست آمده است. برای استخراج ویژگی‌های بافت نیز، از توصیفگرهای الگوی دودویی محلی[۴]، ماتریس هم‌رخداد[۵] و الگوی رگبرگ استفاده شده است. بیماری گیاهان تغییر زیادی در فرم و شکل برگ نسبت به حالت سالم آن ایجاد خواهد کرد [۸]. بنابراین بررسی ویژگی‌های مربوط به شکل و ساختار برگ می‌تواند در تمایز بین برگ نمونه‌های سالم و بیمار یک گیاه نقش به سزایی داشته باشد. از این رو با استخراج خطوط مرزبندی[۶] از تصویر قطعه‌بندی شده، مساحت و محیط برگ بدست آمده است. پس از استخراج ویژگی‌ها، با استفاده از روش انتخاب ویژگی mRMR[۷]، تعدادی از ویژگی‌های غیرمفید حذف شده است. در پایان نیز برای تشخیص نمونه‌های سالم و بیمار از دسته‌بندهای جنگل تصادفی[۸] و ماشین بردار پشتیبان[۹] روی مجموعه داده PlantVillage استفاده شده است.

ادامه مقاله به ترتیب زیر خواهد بود: در بخش ۲ کارهای مرتبط پیشین بررسی شده و در بخش ۳ روش پیشنهادی شرح داده شده است. در بخش ۴ به بررسی مجموعه داده و ارائه نتایج و مقایسه با روش‌های پیشین پرداخته شده است. در بخش پایانی نیز نتیجه گیری و پیشنهاد برای کارهای آینده مطرح شده است.

۲- مروری بر کارهای پیشین

باتوجه به ضرورت و اهمیت تشخیص بیماری گیاهان و افزایش بازده محصولات کشاورزی، محققان کارهای زیادی در این زمینه انجام داده‌اند. هدف از این بخش، مرور رویکردها و روش‌هایی است که می‌توانند در حوزه شناسایی بیماری گیاهان به عنوان منبعی برای مطالعات و تحقیقات آینده قرار بگیرند. تصویر برداری فرا طیفی و دسته‌بندی با روش‌های یادگیری ماشین [۳]، جدا سازی تصویر برگ با روش‌های قطعه‌بندی و سپس دسته‌بندی بوسیله شبکه‌های عصبی [۴]، دسته‌بندی بیماری گیاهان با شبکه‌های عصبی و رویکرد یادگیری انتقالی [۹]، خلاصه‌ای از روش‌هایی است که پیش‌تر در این زمینه مورد استفاده قرار گرفته‌اند. در یکی از پژوهش‌ها، کارایی تصویربرداری فرا طیفی در مساله تشخیص بیماری گیاهان بررسی شده است. در این تحقیق [۳] از طیف‌های مادون قرمز نزدیک و مرئی (VNIR)[۱۰] و همچنین طول موج کوتاه (SWIR)[۱۱] برای بدست آوردن ویژگی‌های شاخص پوشش گیاهی و میانگین بازتاب هر طول موج برای همه پیکسل‌های یک برگ استفاده شده است. محققان در مقاله [۳] از الگوریتم خوشه بندی k-means برای قطعه‌بندی برگ‌ها و سپس مدل ماشین بردار پشتیبان با کرنل تابع پایه شعاعی[۱۲] برای دسته‌بندی بیماری گیاهان استفاده کردند. این روش به دلیل استفاده از دوربین‌های مخصوص تصویر برداری فراطیفی، هزینه زیادی در پی خواهد داشت [۳]. در روشی دیگر، تصویر برگ با

استفاده از روش K-means قطعه‌بندی شد و از شبکه‌های عصبی برای دسته‌بندی گیاهان بیمار و سالم استفاده شده است [۴].

در مرجع [۵]، از ویژگی‌های رنگ و بافت و رویکردهای یادگیری ماشین برای دسته‌بندی بیماری گیاهان استفاده شده است. نویسندگان با کمک کارشناسان آسیب شناسی گیاهی مجموعه داده خود را با استفاده از تصاویری ایجاد کردند که در شرایط کنترل شده و با پس زمینه سفید تهیه شده بودند. در بخش پیش پردازش از روش متعادل سازی رنگ برای بهبود تصاویر استفاده و سپس با تبدیل به فضای رنگ YCbCr، جداسازی بخش‌های بیمار برگ انجام شد. محققان از ویژگی‌های رنگ مانند میانگین و انحراف معیار کانال‌های RGB و YCbCr، و از ویژگی‌های بافت مانند ماتریس هم‌رخداد و همچنین ویژگی‌های گابور در تصویر استفاده کردند. در پایان از دسته‌بندهای حداقل فاصله[۱۳] و همچنین ماشین بردار پشتیبان استفاده نمودند.

P. Kulkarni و همکاران [۸]، با استفاده از تکنیک اوتسو و عملگر هم‌ریختی، فرآیند قطعه‌بندی را روی تصویر برگ انجام داده و ویژگی‌های رنگ و بافت شامل میانگین، واریانس، نرخ رنگ سبز و ماتریس هم‌رخداد (همگنی، کنتراست، انرژی، همبستگی، تضاد) را بدست آورده و در نهایت دسته‌بندی تصاویر نمونه‌های سالم و بیمار برگ را با استفاده از روش جنگل تصادفی انجام داده‌اند.

در یک رویکرد یادگیری عمیق [۹]، برای تشخیص و دسته‌بندی بیماری گیاهان از شبکه عصبی CNN[۱۴] با دو معماری AlexNet و GoogleNet، در دوحالت استفاده شد. ابتدا هر دو معماری تنها روی مجموعه داده PlantVillage آموزش داده شدند. در حالت دیگر، عملکرد آنها با استفاده از روش یادگیری انتقالی و تنظیم وزن برخی لایه‌های مدل از پیش آموزش داده شده (در مجموعه داده ImageNet) ارزیابی شد.

در مرجع [۱۰]، یک سیستم تشخیص پژمردگی میکروبی برای میوه انار با استفاده از ویژگی‌های رنگ، میانگین، انحراف معیار، واریانس، همبستگی، آنتروپی و لبه‌ها ارائه کردند. نویسندگان از روش grab cut برای قطعه‌بندی و از الگوریتم Canny برای بدست آوردن لبه‌های تصویر استفاده کردند. این سیستم می‌تواند میزان آلودگی در میوه را تشخیص دهد. مثلا با بررسی تصویر ۴۵۰ نمونه، ۳۵ عدد دارای آلودگی زیر ٪۲۰، ۷۲ عدد با آلودگی حدود ٪۴۰ و ۳۴۳ عدد نیز دارای آلودگی بالای ٪۸۰ بوده اند.

نویسندگان در پژوهش [۱۱]، با استفاده از تکنیک‌های پردازش تصویر و شبکه عصبی CNN، به شناسایی بیماری گونه سیب در مجموعه داده PlantVillage پرداختند. آنها برای قطعه‌بندی از تکنیک K-means استفاده کرده و همچنین ویژگی‌های رنگ، شکل و بافت را برای جداسازی بیماری‌های مختلف گونه سیب مورد بررسی قرار دادند. از معایب این روش، دقت نه چندان بالای آن بوده است.

در مرجع [۱۹]، بیماری‌های گیاه سیب‌زمینی به صورت اختصاصی مورد بررسی قرار گرفته است. در این مقاله تصاویر نمونه سیب‌زمینی در مجموعه داده PlanVillage و همچنین ۱۷۰۰ تصویر از یک مجموعه داده سفارشی، با استفاده از مدل Densenet مورد ارزیابی قرار گرفته است. در این پژوهش، پنج کلاس مرتبط با بیماری‌های گیاه سیب‌زمینی بررسی شد.

A.I. Khan و همکاران [۲۰]، یک مجموعه داده تخصصی شامل ۹۰۰۰ تصویر RGB با کیفیت بالا از برگ نمونه‌های بیمار گیاه سیب ایجاد کردند. سپس با استفاده از یک سیستم دو مرحله‌ای تشخیص بیماری مبتنی بر یادگیری عمیق، به دسته بندی نمونه‌های سالم و بیمار پرداختند. در مرحله اول، با استفاده از یک شبکه عصبی کانولوشنی مبتنی بر معماری Xception ، تصاویر ورودی را به دسته‌های بیمار و سالم دسته‌بندی کردند. در مرحله دوم با استفاده از مدل‌های Yolov4، Faster R-CNN و EfficientDet نوع بیماری را تشخیص دادند.

نویسندگان در مرجع [۲۱]، مجموعه داده‌ای جدید از تصاویر برچسب گذاری شده برگ گیاه سیب در محیط‌های واقعی ساختند. سپس از روش‌های تشخیص اشیا Faster R-CNN و YOLOv3 برای شناسایی بیماری استفاده کردند. این روش می‌تواند هر ۵ نوع بیماری موجود در برگ گیاه سیب را به عنوان یک شی شناسایی کند. نتایج این پژوهش نشان داده است که Faster R-CNN در شناسایی لکه‌های بیماری کوچکتر و YOLOv3 در سرعت تشخیص در زمان‌های واقعی، نسبت به یکدیگر برتری دارند.

در مرجع [۲۲]، پژوهشگران یک روش سه مرحله‌ای را توسعه دادند که با استفاده از ۵ مدل از پیش آموزش داده شده به نام‌های ResNet50، AlexNet، GoogleNet، VGG19 و EfficientNet به شناسایی بیماری در تصاویر برگ سه نمونه گیاه فلفل دلمه‌ای، سیب زمینی و گوجه فرنگی می‌پردازد. بدین ترتیب که در مرحله نخست، نوع محصول مشخص می‌شود. سپس سالم یا بیمار بودن آن تشخیص داده می‌شود و در پایان نیز نوع بیماری گیاه اعلام می‌شود.

در پژوهش [۲۳]، نویسندگان مجموعه داده اختصاصی خود را با ۴۰۰ تصویر برگ گیاه گوجه فرنگی ساختند. روش آنها بر پایه پردازش تصویر و تکنیک‌های یادگیری ماشین استوار بود. پژوهشگران ابتدا مراحل پیش پردازش تصویر یعنی برش ناحیه

شامل تصویر برگ، تغییر اندازه و بهبود کیفیت تصویر را انجام دادند. سپس تصویر را به بخش‌های کوچکتر تقسیم کرده و با استفاده از ماتریس هم‌رخداد، ۱۳ ویژگی مربوط به بافت را از آن بدست آوردند. در پایان نمونه‌ها را با روش SVM، به ۴ کلاس شامل ۳ بیماری و یک گروه سالم دسته‌بندی کردند.

A. Haridasan و همکاران [۲۴]، عملکرد تکنیک‌های یادگیری ماشین و یادگیری عمیق را روی مجموعه داده اختصاصی تصاویر برگ نمونه‌های گیاه برنج، که از مزارع مختلف جمع‌آوری شده بود بررسی کردند. محققان ابتدا تصاویر را به صورت دستی تهیه و آنها را برچسب گذاری کردند. سپس مراحل پیش‌پردازش، قطعه‌بندی تصویر و استخراج ویژگی‌ها را انجام دادند. در نهایت دسته‌بندی نمونه‌ها را با مدل‌های SVM و CNN انجام دادند.

در مرجع [۲۵]، نویسندگان از رویکردهای یادگیری عمیق و XAI[۱۵] برای دقت بیشتر و تفسیرپذیری مدل استفاده کرده‌اند. آنها ابتدا تصاویر ورودی را با تغییر اندازه و همچنین نرمالسازی پیکسل‌ها، پیش‌پردازش کردند. سپس از دو شبکه CNN و MobileNet برای دسته بندی نمونه‌ها استفاده کردند. در نهایت برای بررسی تفسیرپذیری مدل از تکنیک Grad-CAM[۱۶] بهره بردند. نویسندگان با ارائه یک نمایش تصویری از نحوه تشخیص الگوها و تصمیم‌گیری های مدل نشان دادند چه پیکسل‌هایی در ارزیابی مدل‌ها مهمتر هستند.

در مرجع [۲۶]، روش‌های یادگیری ماشین و یادگیری عمیق برای دسته‌بندی تصاویری از گیاهان سیب، گندم، سویا، ذرت و برنج بررسی شد. نویسندگان ابتدا تصاویر گیاهان را با استفاده از یک پهباد مجهز به دوربین نوری در ارتفاع پایین جمع‌آوری کردند. سپس مراحل پیش‌پردازش شامل تغییر اندازه تصاویر، حذف نویز و افزایش کنتراست را انجام دادند. در ادامه ویژگی‌های بافت را با استفاده از ماتریس هم‌رخداد استخراج کرده و در پایان دسته‌بندی نمونه‌ها را با تکنیک‌های جنگل تصادفی، ماشین بردار پشتیبان، Naïve Bayes و همچنین مدل‌های CNN و LSTM[۱۷] انجام دادند.

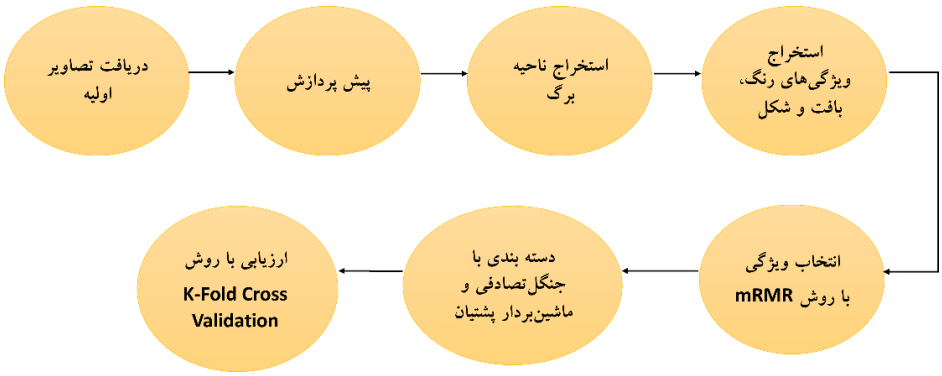
محققان در مرجع [۲۷] ابتدا مجموعه داده اختصاصی خود را با استفاده از ۴۰۰۰ تصویر برگ نمونه‌های بیمار گیاه چای از ۴ باغ در بنگلادش ایجاد کردند. سپس با استفاده از روش داده‌افزایی تعداد نمونه‌ها را افزایش دادند. در پایان، از نسخه‌ای از مدل YOLOv7 به نام YOLO-T برای دسته‌بندی تصاویر گیاه چای استفاده کردند.

۳- روش پیشنهادی

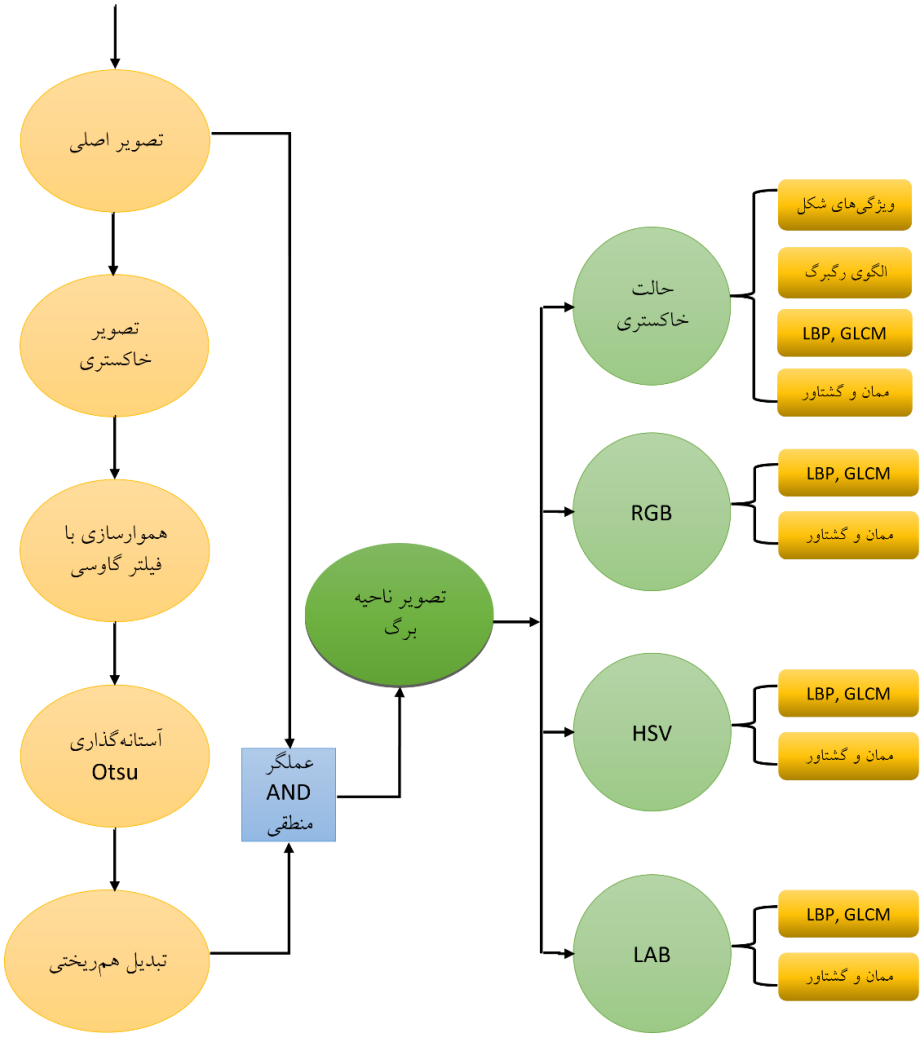
در این مقاله روش جدیدی برای تشخیص زودهنگام بیماری گیاهان، با استفاده از تصاویر برگ آنها ارائه شده است. اثر بیماری و آفت زدگی گیاهان، اغلب در برگ آنها نمایان است و نشانه‌های برجسته‌ای از خود بروز می‌دهد. در روش پیشنهادی، پس از قطعه‌بندی ناحیه برگ و استخراج ویژگی‌های مبتنی بر رنگ، بافت، شکل کلی و ظاهر برگ‌ها ، اطلاعات مفیدی را از بیماری گیاهان استخراج نموده و در اختیار مدل یادگیری ماشین قرار دادیم. برخی از این ویژگی‌ها ممکن است افزونه یا نویز بوده و تاثیر چندانی در بهبود عملکرد مدل نداشته باشند. از این رو با استفاده از روش انتخاب ویژگی mRMR، تعدادی از ویژگی‌های غیرمفید حذف شده است. در نهایت از دو دسته‌بند جنگل تصادفی و ماشین بردار پشتیبان استفاده شده و دقت بالاتری نسبت به دیگر روش‌های پیشین به دست آمد. نمودار گردش کار روش پیشنهادی و مراحل استخراج ویژگی به ترتیب در شکل‌های (۲) و (۳) نشان داده شده و در ادامه، جزییات روش پیشنهادی شرح داده شده است.

۳-۱- استخراج ناحیه برگ

ابتدا با استفاده از فیلتر گاوسی نویزهای احتمالی تصویر برگ ورودی حذف و هموار شده است. سپس با روش آستانه گذاری اوتسو، عمل

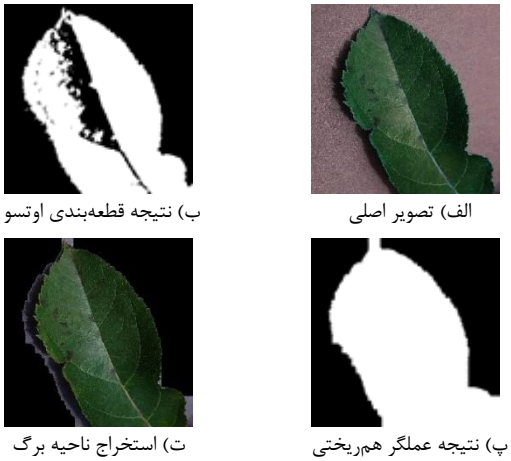


شکل ۲: گردش کار روش پیشنهادی برای تشخیص زودرس بیماری گیاهان.



شکل ۳: مراحل پیش پردازش، استخراج ناحیه برگ و استخراج ویژگی روش پیشنهادی.

قطعه‌بندی را انجام داده و ناحیه برگ از پس زمینه جدا شده است. در روش اوتسو قطعه‌بندی تصویر بر اساس بیشینه کردن واریانس بین دو قطعه صورت می‌گیرد. در این روش مقداری برای آستانه انتخاب می‌شود، که دو ناحیه با اختلاف واریانس زیاد بدست آید. در ادامه با استفاده از عملگر هم‌ریختی، حفره‌های احتمالی بدست آمده از مرحله قطعه‌بندی با استفاده از روش بستن[۱۸]، حذف شده است تا با این کار دقت قطعه‌بندی افزایش یابد. سپس با انجام عملگر منطقی بین تصویر دودویی بدست آمده در این مرحله و تصویر ورودی اولیه، تصویر رنگی ناحیه برگ از پس زمینه جدا شده است. در شکل (۴)، مراحل استخراج ناحیه برگ شامل قطعه‌بندی به روش اوتسو، خروجی عملگر هم‌ریختی و ناحیه برگ استخراج شده برای نمونه‌ای از گیاه سیب نشان داده شده است.



شکل ۴: مراحل استخراج ناحیه برگ گیاه سیب.

۳-۲- ویژگی‌های رنگ

با توجه به اهمیت بالقوه رنگ برگ در تشخیص بیماری‌های گیاهی، روش پیشنهادی از پردازش تصویر و استخراج ویژگی‌ها در فضاهای رنگی متنوع استفاده کرده است. در حالی که در بسیاری از تحقیقات قبلی تنها تصویر خاکستری برگ مورد بررسی قرار گرفته است،

استفاده از فضاهای رنگ متفاوت، امکان تشخیص دقیق‌تری از بیماری‌های گیاهی را فراهم می‌سازد. در این راستا برای بدست آوردن ویژگی‌های رنگ، از سه فضای رنگ RGB، HSV و LAB استفاده شده است. فضای رنگ RGB شامل کانال‌های قرمز، سبز و آبی است. فضای رنگ HSV شامل مولفه‌های H (طول موج)، S (غلظت) و V (شدت روشنایی رنگ) است [۷]. فضای رنگ LAB نیز شامل سه کانال L (روشنایی) و A و B (اطلاعات رنگ) است [۱۲]. برای استخراج ویژگی‌های رنگ، از معیارهای ممان[۱۹] و گشتاور در ۹ کانال رنگ یاد شده استفاده شده است. ویژگی‌های گشتاور شامل گشتاور اول(جرم) و گشتاور دوم(اینرسی) است و ویژگی‌های ممان نیز شامل میانگین، واریانس، چولگی و کشیدگی هستند که به ترتیب در روابط (۱) تا (۱۵) نشان داده شده‌اند. این ویژگی‌ها می‌توانند اطلاعات مفیدی درباره چگونگی توزیع سطح روشنایی پیکسل‌ها در تصویر ارائه نمایند.

$$\mathrm{M_{00}} = \sum_{\mathrm{i=0}}^{\mathrm{h-1}} \sum_{\mathrm{j=0}}^{\mathrm{w-1}} \mathrm{I(i,j)} \qquad (۱)$$

$$M_{10} = \sum_{i=0}^{h-1} \sum_{j=0}^{w-1} i.I(i,j) \qquad (۲)$$

$$M_{01} = \sum_{i=0}^{h-1} \sum_{j=0}^{w-1} j.I(i,j) \qquad (۳)$$

$$M_{20} = \sum_{i=0}^{h-1} \sum_{j=0}^{w-1} (i - M_{10})^2.I(i,j) \qquad (۴)$$

$$M_{02} = \sum_{i=0}^{h-1} \sum_{j=0}^{w-1} (j - M_{01})^2.I(i,j) \qquad (۵)$$

$$M_{11} = \sum_{i=0}^{h-1} \sum_{j=0}^{w-1} (i - M_{10}).(j - M_{01}).I(i,j) \qquad (۶)$$

$$In_{11} = \frac{M_{10}}{M_{00}} \qquad (۷)$$

$$In_{12} = \frac{M_{01}}{M_{00}} \qquad (۸)$$

$$In_{21} = \frac{M_{20}}{M_{00} - {In_{11}}^2} \qquad (۹)$$

$$In_{22} = \frac{M_{02}}{M_{00} - {In_{12}}^2} \qquad (۱۰)$$

$$In_{23} = \frac{M_{11}}{M_{00} - (In_{11} \times In_{12})} \qquad (۱۱)$$

$$\mu = \frac{1}{n} \sum_{i=0}^{h-1} \sum_{j=0}^{w-1} I(i,j) \qquad (۱۲)$$

$$\sigma^2 = \frac{1}{n} \sum_{i=0}^{h-1} \sum_{j=0}^{w-1} (I(i,j) - \mu)^2 \qquad (۱۳)$$

$$\gamma = \frac{1}{n} \sum_{i=0}^{h-1} \sum_{j=0}^{w-1} \left(\frac{I(i,j) - \mu}{\sigma}\right)^3 \qquad (۱۴)$$

$$\beta = \frac{1}{n} \sum_{i=0}^{h-1} \sum_{j=0}^{w-1} \left(\frac{I(i,j) - \mu}{\sigma}\right)^4 \qquad (۱۵)$$

در این روابط، $I(i,j)$ نشـان دهنده تصویر ورودی در مختصات (i,j) است. h و w به ترتیب طول و عرض تصویر را نشـان می‌دهند. M_{00} به معنی ممان صفرم یا مجموع مقادیر همه پیکسل‌های تصویر است. ممان M_{10} بیانگر توزیع افقی مقادیر پیکسل‌ها در تصویر است. به همین ترتیب M_{01} توزیع عمودی مقادیر پیکسل‌ها در تصویر را نشان می‌دهد. M_{20} ، M_{02} و M_{11} از ممان‌های دوم اینرسی بوده و به ترتیب توزیع جرم تصویر را در راستای عمودی، افقی و قطری نشان می‌دهند. In_{11} و In_{12} نشان دهنده موقعیت مرکز جرم تصویر (مختصات طول و عرض مرکز) بوده و In_{21} ، In_{22} و In_{23} نمایانگر انحراف از مرکز تصویر هستند. μ وσ^2 به ترتیب بیانگر میانگین و انحراف معیار شدت رنگ پیکسل‌های تصویر هستند و معیارهای γ و β نیز نشان دهنده ویژگی‌های چولگی و کشیدگی هستند.

۳-۳- ویژگی‌های بافت

با توجه به تغییراتی که بیماری‌های گیاهی در فرم و الگوی برگ ایجاد می‌کنند، در روش پیشنهادی برای کسب اطلاعات دقیق‌تر از بافت تصویر، علاوه بر استفاده از ماتریس هم‌رخداد، از الگوی دودویی محلی و الگوی رگبرگ‌ها نیز استفاده شده است تا جزئیات بیشتری را از تصاویر برگ گیاهان استخراج کند.

۳-۳-۱- الگوی دودویی محلی

الگوی دودویی محلی یک توصیفگر بافت و تکنیک مناسبی برای بدست آوردن الگوهای موجود در تصویر است. با استفاده از رابطه (۱۶)، می‌توان تصویر الگوی دودویی محلی ناحیه برگ را بدست آورد.

$$\begin{bmatrix} x_0 & x_1 & x_2 \\ x_7 & T & x_3 \\ x_6 & x_5 & x_4 \end{bmatrix}$$

$$x_i = \begin{cases} 1, x_i \geq T \\ 0, x_i < T \end{cases} \gg\gg T = LBP = \sum_{i=0}^{7} 2^i x_i \qquad (۱۶)$$

در این رابطه، T پیکسل مرکزی تصویر و x_i ها، ۸ همسایگی با شعاع یک از مرکز هستند. مقدار همه همسایه‌ها با پیکسل مرکزی مقایسه می‌شود. اگر بزرگتر یا مساوی با پیکسل مرکزی باشد، مقدار آن پیکسل برابر با یک و اگر کوچکتر از پیکسل مرکزی باشد مقدار آن برابر صفر خواهد شد. در نهایت مقادیر همسایه‌ها را به صورت یک دنباله دودویی نوشته و معادل دهدهی آن به عنوان الگوی مورد نظر در پیکسل مرکزی ثبت می‌شود. الگوی دودویی محلی تصویری به اندازه تصویر ورودی است. برای بدست آوردن بردار ویژگی مناسب، هیستوگرام الگوی دودویی محلی استخراج شد که برداری به طول ۲۵۶ است. برای کاهش ابعاد ویژگی، این مقادیر چندی سازی[۲۰] و به ۱۶ دسته تبدیل شده است. در شکل (۵)، تصویر توصیفگر الگوی دودویی محلی و هیستوگرام آن برای دو نمونه برگ سالم و بیمار در گیاه ذرت نمایش داده شده است. همانطورکه در این شکل مشخص است، هیستوگرام الگوی دودویی محلی ویژگی مناسبی برای تمایز دو نمونه برگ سالم و بیمار است.

۳-۳-۲- ماتریس هم‌رخداد

در این مقاله افزون بر هیستوگرام الگوی دودویی محلی، از ماتریس هم‌رخداد برای استخراج بافت تصویر استفاده شده است. این ماتریس، مربعی بوده و ابعاد آن متناظر با تعداد سطوح خاکستری تصویر است. بنابراین اگر تصویر ۸ بیتی باشد، تعداد سطوح خاکستری ۲۵۶ بوده و این ماتریس دارای ۲۵۶ سطر و۲۵۶ ستون خواهد بود. هر خانه از ماتریس هم‌رخداد، نشان دهنده تعداد دفعات رخداد دو شدت رنگ همسایه، در یکی از چهار جهت ۰، ۴۵، ۹۰ و ۱۳۵ درجه در تصویر است که می‌تواند جزییات خوبی را از نوع بافت تصویر بیان کند. در روش پیشنهادی پس از استخراج ماتریس هم‌رخداد، چهار ویژگی کنتراست، همگنی، انرژی و همبستگی استخراج شده که به ترتیب در روابط (۱۷) تا (۲۰) نشان داده شده‌اند:

$$\sum_{ij=0}^{M-1} P_{ij}(i-j)^2 \qquad (۱۷)$$

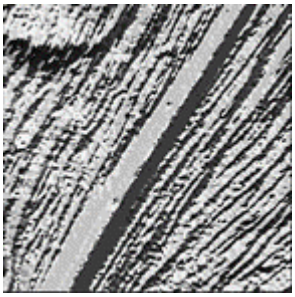
$$\sum_{ij=0}^{M-1} \frac{P_{ij}}{1+(i-j)^2} \qquad (۱۸)$$

$$\sum_{i=0}^{M-1}\sum_{j=0}^{M-1} {P_{ij}}^2 \qquad (۱۹)$$

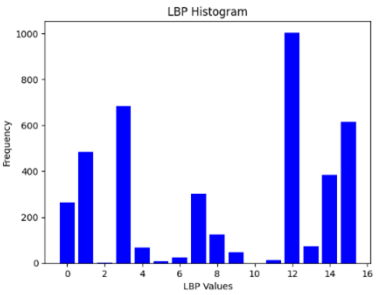
$$\frac{\sum_{i=0}^{M-1}\sum_{j=0}^{M-1}(i-\bar{x})(j-\bar{y})Pij}{\sigma_x \sigma_y} \qquad (۲۰)$$



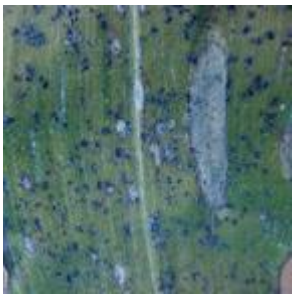
الف) تصویر اصلی نمونه سالم



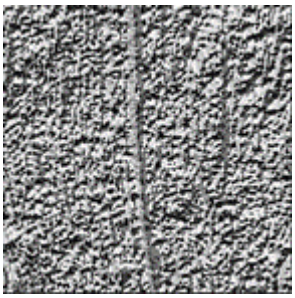
ب) تصویر الگوی دودویی محلی



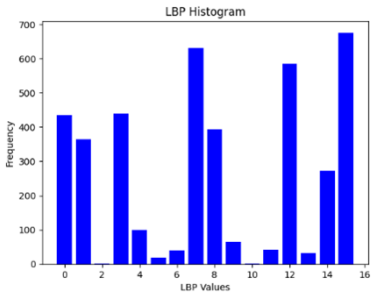
پ) هیستوگرام الگوی دودویی محلی



ت) تصویر اصلی نمونه بیمار



ث) تصویر الگوی دودویی محلی



ج) هیستوگرام الگوی دودویی محلی

شکل ۵: تصویر الگوی دودویی محلی برای دو نمونه سالم و بیمار گیاه ذرت.

در این روابط، i و j سطر و ستون‌های ماتریس هم‌رخداد، P_{ij} مقدار ماتریس هم‌رخداد برای جفت مرتب (i,j) و $\bar{x}$، $\bar{y}$، σ_x و σ_y به ترتیب میانگین و واریانس مجموع مقادیر سطری و ستونی هستند. M نیز نمایانگر تعداد سطر و ستون ماتریس هم‌رخداد است که در روش پیشنهادی برابر با ۲۵۶ در نظر گرفته شده است.

۳-۳-۳- الگوی رگبرگ

همانطور که اشاره شد، الگوی رگبرگ‌ها می‌تواند اطلاعات خوبی درباره وضعیت برگ از جهت فرم و سطح آن ارائه کند. روش کار بدین ترتیب است که ابتدا با استفاده از الگوریتم لبه یابی Canny، لبه‌های موجود در تصویر ناحیه برگ و سپس خطوط مرزبندی موجود در آن بدست آمده است. خطوط مرزبندی، عامل اتصال نقاط با شدت رنگ یکسان در امتداد یک شی هستند که در اینجا می‌توانند الگوی رگبرگ‌ها را توصیف کنند. با استفاده از مستطیل‌هایی که به دور هر یک از این خطوط مرزبندی در نظر گرفته شد، مقادیر مساحت و نسبت عرض به ارتفاع هر یک از مستطیل‌ها محاسبه شد. در پایان تعداد رگبرگ‌ها، میانگین مساحت و همچنین میانگین نسبت عرض به ارتفاع همه مستطیل‌ها استخراج شد. در شکل (۶)، مراحل استخراج ویژگی‌های رگبرگ نمایش داده شده است. در شکل (۶)، ابتدا تصویر ورودی برگ نشان داده شده است. سپس نتیجه اعمال تکنیک لبه‌یابیCanny و در پایان نیز، مستطیل‌هایی به دور هر یک از خطوط مرزبندی نمایش داده شده است.

۳-۴- ویژگی‌های شکل

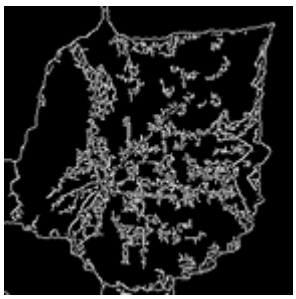
استفاده از ویژگی‌های شکل می‌تواند در تشخیص و جداسازی نمونه‌های بیمار از سالم، کمک کننده و مفید باشد. برای این منظور، با استفاده از تصویری که با روش هم‌ریختی، در مرحله استخراج ناحیه برگ بدست آمده بود، خطوط مرزبندی ناحیه برگ و همچنین مساحت و محیط این خطوط محاسبه شد. بدین ترتیب ویژگی‌های مساحت و محیط برگ نیز بدست آمد.

۳-۵- انتخاب ویژگی‌ها

در پایان مراحل استخراج ویژگی، تعداد ۳۰۵ ویژگی از هر تصویر برگ ورودی بدست آمده که جزییات زیادی را جهت بررسی و دسته‌بندی نمونه‌ها در فرآیند یادگیری ماشین در اختیار ما قرار می‌دهد. اما در عین حال ممکن است تعدادی از این ویژگی‌ها افزونه بوده و کارایی چندانی نداشته باشند. انتخاب ویژگی یکی از مهم‌ترین بخش‌ها در فرآیند یادگیری ماشین بوده و هدف آن کاهش ابعاد داده است. یعنی زیر مجموعه‌ای از ویژگی‌ها را انتخاب کنیم که بیشترین تاثیر را در عملکرد مدل یادگیری ماشین داشته باشند. بنابراین باید ویژگی‌های غیرمرتبط، تکراری و نویزی حذف شوند. انتخاب ویژگی درست می‌تواند کارایی، دقت و تفسیرپذیری مدل را تا اندازه مناسبی افزایش داده و در سوی مقابل، پیچیدگی محاسباتی آن را کاهش دهد [۱۳] و همچنین از بیش برازش مدل جلوگیری کند. یکی از روش‌های انتخاب ویژگی کارا و موثر، روش mRMR است. این روش به دلیل استفاده از معیارهای ساده و سریع برای محاسبه ارتباط بین ویژگی‌ها، سرعت بالایی در انتخاب ویژگی دارد. بدین ترتیب شناسایی ویژگی‌های مهم در مجموعه داده‌های بزرگ، به شکل سریع و کارآمد ممکن خواهد بود. روش mRMR از دسته روش‌های انتخاب ویژگی بر پایه فیلتر است. در روش‌های مبتنی بر فیلتر، به هر یک از ویژگی‌ها براساس معیارهایی مانند ضریب همبستگی و آنتروپی، امتیازی داده می‌شود. سپس با در نظر گرفتن یک مقدار آستانه مناسب، می‌توان ویژگی‌های مفید را که دارای امتیازی بیشتر از مقدار آستانه هستند، انتخاب کرده و ویژگی‌های کم اهمیت و دارای امتیازی پایین‌تر از مقدار آستانه را حذف کرد. در روش mRMR ابتدا مطابق رابطه (۲۱)، برای هر یک از ویژگی‌ها یک رتبه در نظر گرفته می‌شود.



الف) تصویر ورودی



ب) خروجی روش لبه یابی Canny



پ) استخراج محدوده رگبرگ‌ها

شکل ۶: مراحل استخراج ویژگی‌های رگبرگ.

$$mRMR = max_S \left[\frac{1}{|S|} \sum_{i \epsilon S} I(c,i) - \frac{1}{|S|^2} \sum_{i,j \epsilon S} I(i,j) \right] \quad (۲۱)$$

در این رابطه، S مجموعه ویژگی‌ها و $I(i,j)$ معیار اطلاعات متقابل[۲۱] بین ویژگی‌های i و j است. متغیر c نشان دهنده ی برچسب یا کلاس نمونه‌ها است. هدف این است که مجموعه ویژگی‌ها به گونه‌ای انتخاب شود که جمله اول بیشینه و جمله دوم کمینه گردد تا بدین ترتیب مقدار نهایی mRMR نیز افزایش یابد.

رتبه بندی در روش mRMR بر پایه معیار اطلاعات متقابل بین ویژگی‌ها صورت می‌گیرد. بدین ترتیب که هرچه مقدار اطلاعات متقابل بین دو ویژگی بیشتر باشد، وابستگی دو ویژگی بیشتر است و هرچه مقدار این معیار کمتر باشد، دو ویژگی استقلال بیشتری دارند. در این روش به دنبال ویژگی‌هایی هستیم که کمترین وابستگی را با دیگر ویژگی‌ها و بیشترین وابستگی را با برچسب داشته باشند.

در روش پیشنهادی با استفاده از تکنیک انتخاب ویژگی mRMR و مقدار آستانه ۰/۲، از بین ۳۰۵ ویژگی استخراج شده تنها تعدادی ویژگی موثر انتخاب و سایر ویژگی‌های کم اهمیت حذف شده اند. مقدار آستانه ۰/۲، به صورت تجربی و پس از چندین آزمایش بدست آمده است. در جدول (۱)، تعداد ویژگی‌های حذف شده بوسیله روش mRMR برای ۹ نمونه گیاه نشان داده شده است. همانطور که در این جدول مشاهده می‌گردد، برای اغلب گیاهان، تعداد زیادی از ویژگی‌ها حذف شده است.

جدول ۱: تعداد ویژگی‌های حذف شده با استفاده از روش mRMR با مقدار آستانه ۰/۲.

نام گیاه	تعداد ویژگی حذف شده	تعداد ویژگی باقی مانده
سیب	۲۵۵	۵۰
ذرت	۷۵	۲۳۰
انگور	۱۸۷	۱۱۸
سیب زمینی	۲۰۰	۱۰۵
گوجه	۱۹۷	۱۰۸
گیلاس	۱۹۲	۱۱۳
هلو	۲۴۲	۶۳
فلفل	۳۰۱	۴
توت فرنگی	۲۶۲	۴۳

۳-۶- دسته‌بندها

در این بخش از دو دسته‌بند جنگل تصادفی و ماشین بردار پشتیبان استفاده شده است. جنگل تصادفی از چند درخت تصمیم استفاده می‌کند که هریک روی بخشی خاص از مجموعه داده، آموزش می‌بینند. در پایان نتیجه دسته‌بندی که بیشترین تکرار یا رای را بدست بیاورد، به عنوان خروجی الگوریتم در نظر گرفته می‌شود [۱۴]. ماشین بردار پشتیبان یکی از محبوب‌ترین الگوریتم‌های یادگیری ماشین بوده که در مسائل دسته‌بندی و رگرسیون مورد استفاده قرار می‌گیرد. از برتری‌های آن می‌توان به عملکرد خوب در مجموعه داده‌های کوچک و همچنین مقاومت در برابر داده‌های پرت[۲۲] اشاره کرد [۱۵]. در روش پیشنهادی، ارزیابی مدل‌ها با استفاده از روش K-Fold Cross Validation با K=۵ انجام شده است. در این تکنیک، مجموعه داده به ۵ بخش تقسیم شده و فرآیند آموزش مدل ۵ بار تکرار می‌گردد که در هر بار ۴ بخش به عنوان مجموعه آموزش و یک بخش برای مجموعه آزمایش در نظر گرفته شده و دقت مدل محاسبه می‌شود. در پایان، میانگین دقت ۵ مرحله به عنوان ارزیابی نهایی مدل ثبت می‌گردد.

۴- نتایج

در این مقاله برای تشخیص زودرس بیماری گیاهان، آزمایشات روی مجموعه داده PlantVillage انجام شده است. برای ارزیابی عملکرد مدل از معیار دقت و همچنین ماتریس سردرگمی[۲۳] و نمودار ROC[۲۴] استفاده شده است. در پایان نتایج روش پیشنهادی با دیگر پژوهش‌های انجام شده در این زمینه مقایسه شد، که نشان دهنده برتری دقت روش پیشنهادی است.

۴-۱- مجموعه داده

مجموعه داده PlantVillage [۱۶] یکی از بزرگترین و رایج‌ترین مجموعه‌های داده در زمینه بیماری‌های برگ گیاهان است که تاکنون کارهای تحقیقات زیادی در ارتباط با آن انجام شده است. این مجموعه داده بوسیله روش داده افزایی[۲۵] روی نسخه اصلی آن [۱۷]، بازسازی شده و دارای ۸۷۸۶۷ تصویر از برگ نمونه‌های سالم و بیمار مربوط به ۹ نوع گیاه مختلف سیب، ذرت، انگور، سیب زمینی، گوجه، گیلاس، هلو، فلفل و توت فرنگی است. در شکل (۷)، چند نمونه تصاویر برگ سالم و بیمار از این مجموعه داده نشان داده شده است. همچنین در جدول (۲)، گزارشی از مشخصات نمونه‌های مجموعه داده ارائه شده است.



شکل ۷: نمونه تصاویر گیاهان سالم و بیمار در مجموعه داده PlantVillage.

جدول ۲: مشخصات نمونه‌های موجود در مجموعه داده.

نام گیاه	تعداد نمونه	تعداد کلاس
سیب	۹۷۱۴	۴
ذرت	۹۱۴۵	۴
انگور	۹۰۲۷	۴
سیب زمینی	۵۷۰۲	۳
گوجه	۱۸۳۴۵	۱۰
گیلاس	۳۵۰۹	۲
هلو	۳۵۶۶	۲
فلفل	۳۹۰۱	۲
توت فرنگی	۳۵۹۸	۲

۴-۲- معیارهای ارزیابی

نمودار ROC ابزار مفیدی برای ارزیابی عملکرد مدل‌های یادگیری ماشین به شمار می‌رود. در این نمودار محور افقی مربوط به نرخ مثبت اشتباه[۲۶] یا FPR بوده و بیانگر نسبت تعداد نمونه‌های مثبتی که به اشتباه تشخیص داده شده‌اند، به کل نمونه‌های منفی است. محور عمودی بیانگر نرخ مثبت درست[۲۷] یا TPR بوده و بیانگر نسبت تعداد نمونه‌های مثبتی که به درستی تشخیص داده شده‌اند، به کل نمونه‌های مثبت است. با توجه به این دو نرخ، می‌توان در نمودار ROC یک منحنی رسم کرد که نشان می‌دهد با تغییر آستانه، چگونه نرخ تشخیص نادرست و درست تغییر می‌کنند. هر نقطه در نمودار ROC به یک آستانه مشخص در تصمیم‌گیری مدل برمی‌گردد. این نقاط وابسته به مقادیر TPR و FPR محاسبه شده برای هر آستانه خاص هستند. خط یکنواخت از (۰،۰) به (۱،۱) در نمودار، نشان‌دهنده عملکرد تصادفی مدل است. همچنین مساحت زیر نمودار ROC (معیار AUC[۲۸]) نیز می‌تواند به عنوان یک معیار جامع از عملکرد مدل استفاده شود، که مقدار بین ۰ و ۱ دارد. مقدار AUC بزرگتر از ۰/۵ نشان می‌دهد که مدل به طور متوسط بهتر از یک دسته‌بند تصادفی تصمیم می‌گیرد. ماتریس سردرگمی به عنوان ابزار مفیدی برای بررسی جزیی‌تر مدل مورد استفاده قرار می‌گیرد.

۴-۳- نتایج ارزیابی

روش پیشنهادی روی ۹ نمونه گیاه مختلف، با استفاده از زبان برنامه‌نویسی پایتون در محیط Google Colab پیاده‌سازی شده و دسته‌بندی با استفاده از الگوریتم‌های جنگل تصادفی و ماشین بردار پشتیبان صورت گرفته است. مقایسه نتایج ارزیابی روش پیشنهادی با مرجع [۸]، روی مجموعه داده PlantVillage در جدول‌های (۳) و (۴) نشان داده شده است. جدول (۳)، نتایج ارزیابی روش پیشنهادی در فضای رنگ RGB را نشان می‌دهد، در حالی که در جدول (۴) نتایج پردازش در چند فضای رنگی نشان داده شده است. همانطور که از نتایج مشاهده می‌گردد، استخراج ویژگی‌ها در چند فضای رنگی، منجر به بهبود دقت مدل می‌شود و روش پیشنهادی نسبت به مرجع [۸] بهتر عمل کرده است.

همچنین در روش پیشنهادی، دسته‌بند جنگل تصادفی، عملکرد بهتری نسبت به روش ماشین بردار پشتیبان داشته است. مرجع [۸] مجموعه گیاهان گیلاس، هلو، فلفل و توت فرنگی را مورد بررسی قرار نداده و نتایج را تنها روی ۵ نوع گیاه گزارش کرده است.

جدول ۳: نتایج ارزیابی روش پیشنهادی در فضای رنگ RGB، با معیار دقت روی ۹ نمونه گیاه.

نام گیاه	روش پیشنهادی (با استفاده از ماشین بردار پشتیبان)	روش پیشنهادی (با استفاده از جنگل تصادفی)
سیب	٪ ۵۵٫۳۸	٪ ۷۸٫۴۲
ذرت	٪ ۸۲٫۷۱	٪ ۹۲٫۳۶
انگور	٪ ۷۹٫۷۹	٪ ۹۴٫۷۴
سیب زمینی	٪ ۷۴٫۹۲	٪ ۹۶٫۹۴
گوجه	٪ ۴۶٫۲۵	٪ ۷۸٫۷۱
گیلاس	٪ ۹۶٫۰۱	٪ ۹۸٫۹۷
هلو	٪ ۹۱٫۴۵	٪ ۹۴٫۷۲
فلفل	٪ ۷۳٫۸۹	٪ ۹۳٫۲۹
توت فرنگی	٪ ۷۵٫۱۶	٪ ۹۳٫۱۵

جدول ۴: نتایج ارزیابی روش پیشنهادی در فضاهای رنگی چندگانه، با معیار دقت روی ۹ نمونه گیاه.

نام گیاه	روش پیشنهادی (با استفاده از ماشین بردار پشتیبان)	روش پیشنهادی (با استفاده از جنگل تصادفی)	مرجع [۸]
سیب	٪ ۹۱٫۷۵	٪ ۹۶٫۲۶	٪ ۹۱
ذرت	٪ ۹۵٫۷۶	٪ ۹۶٫۴۴	٪ ۹۴
انگور	٪ ۹۹٫۶۴	٪ ۹۹٫۷۲	٪ ۹۵
سیب زمینی	٪ ۹۷٫۱۲	٪ ۹۹٫۲۹	٪ ۹۸
گوجه	٪ ۸۳	٪ ۹۱٫۴۴	٪ ۸۷
گیلاس	٪ ۹۹٫۷۱	٪ ۹۹٫۶۸	---
هلو	٪ ۹۸٫۷۹	٪ ۹۹٫۲۹	---
فلفل	٪ ۸۴٫۴۱	٪ ۹۶٫۵۶	---
توت فرنگی	٪ ۱۰۰	٪ ۱۰۰	---

در شکل (۸) خلاصه‌ای از عملکرد روش پیشنهادی روی چند نمونه گیاه مختلف با تعداد ۲، ۳، ۴ و ۹ کلاس، در قالب نمودار ROC و ماتریس سردرگمی نشان داده شده است. همان گونه که از ماتریس سردرگمی روش پیشنهادی قابل مشاهده است، مقادیر قطر اصلی نسبت به دیگر مقادیر هم ردیف بسیار بزرگتر بوده و این نشان می‌دهد که مدل به خوبی عمل کرده و نمونه‌ها تا اندازه زیادی در کلاس درست خود دسته‌بندی شده‌اند. همچنین نمودار ROC به طرف بالا و راست متمایل بوده که این نشان دهنده سطح زیرنمودار بیشتر، مقادیر TPR بالاتر و درنتیجه عملکرد خوب مدل است.

از آنجا که در پژوهش‌های پیشین، همه گونه‌های گیاهی در مجموعه داده PlantVillage مورد بررسی قرار نگرفته‌اند، میانگین دقت روش پیشنهادی در دو حالت (با ۵ گیاه و ۹ گیاه) محاسبه شده است. برای ۵ گیاه سیب، ذرت، انگور، سیب زمینی و گوجه میانگین دقت ٪ ۹۶/۶۳ به دست آمد. سپس با افزودن نمونه‌های گیلاس، هلو، فلفل و توت فرنگی، میانگین دقت در مجموع به ٪ ۹۷/۶۳ رسیده است. در مرجع [۱۱] میانگین دقت مدل روی نمونه سیب و همچنین در مرجع [۱۸] میانگین دقت برای نمونه‌های سیب زمینی و گوجه از مجموعه داده PlantVillage محاسبه شده است. در جدول (۵)، روش پیشنهادی با دیگر کارهای انجام شده در این زمینه مقایسه و برتری آن نشان داده شده است.

جدول ۵: مقایسه میانگین دقت روش پیشنهادی با روش‌های پیشین روی مجموعه داده PlantVillage.

مرجع	مرجع [۱۱]	مرجع [۱۸]	مرجع [۸]	روش پیشنهادی	
			ارزیابی با ۵ گیاه	ارزیابی با ۵ گیاه	ارزیابی با ۹ گیاه
دقت	٪ ۸۲٫۸۲	٪ ۸۸٫۸۰	٪ ۹۳	٪ ۹۶٫۶۳	٪ ۹۷٫۶۳

۵- نتیجه گیری

در این مقاله روشی برای تشخیص زودهنگام بیماری گیاهان ارائه شده است. با بررسی شرایط ظاهری و شکل برگ گیاهان بیمار، این نتیجه حاصل شده است که ویژگی‌های رنگ، بافت و شکل که ارتباط مستقیمی با ساختار ظاهری برگ‌ها دارند، می‌توانند در فرآیند دسته‌بندی و تشخیص نمونه‌های سالم از بیمار کمک به‌سزایی نمایند. بنابراین در روش پیشنهادی، با استفاده از تکنیک‌های پردازش تصویر، این ویژگی‌ها از تصویر برگ گیاهان استخراج شده و پس از انتخاب ویژگی، آن دسته که بیشترین تاثیر را در فرآیند دسته‌بندی ایفا می‌نمایند، مورد استفاده قرار گرفتند. روش پیشنهادی با دقتی بالا، می‌تواند بیماری ۹ گیاه سیب، ذرت، انگور، سیب زمینی، گوجه، گیلاس، هلو، فلفل و توت فرنگی را با میانگین دقت ٪ ۹۷٫۶۳ با دسته‌بند جنگل تصادفی و ٪ ۹۴٫۹۸ با دسته‌بند ماشین بردار پشتیبان تشخیص داده و در جلوگیری از کاهش کیفیت و کمیت محصول، کارآمد باشد. این روش در مقایسه با تکنیک‌های پیشین، دارای دقت بالاتری است. در آینده انتظار می‌رود این سیستم بتواند در محیط‌های واقعی و تصاویری که با پس زمینه‌های ناهموار و چالشی یا شدت نور متفاوت گرفته می‌شوند، کارایی خود را نشان دهد. همچنین می‌توان روش پیشنهادی را برای شناسایی بیماری گیاهان با استفاده از تصویر بخش‌های دیگر آنها، مانند ساقه یا میوه توسعه داد. علاوه بر این، استفاده از رویکردهای یادگیری عمیق برای تشخیص بیماری‌های گیاهی از تصاویر، می‌تواند به عنوان یک پیشنهاد برای ادامه کار مورد بررسی قرار گیرد.

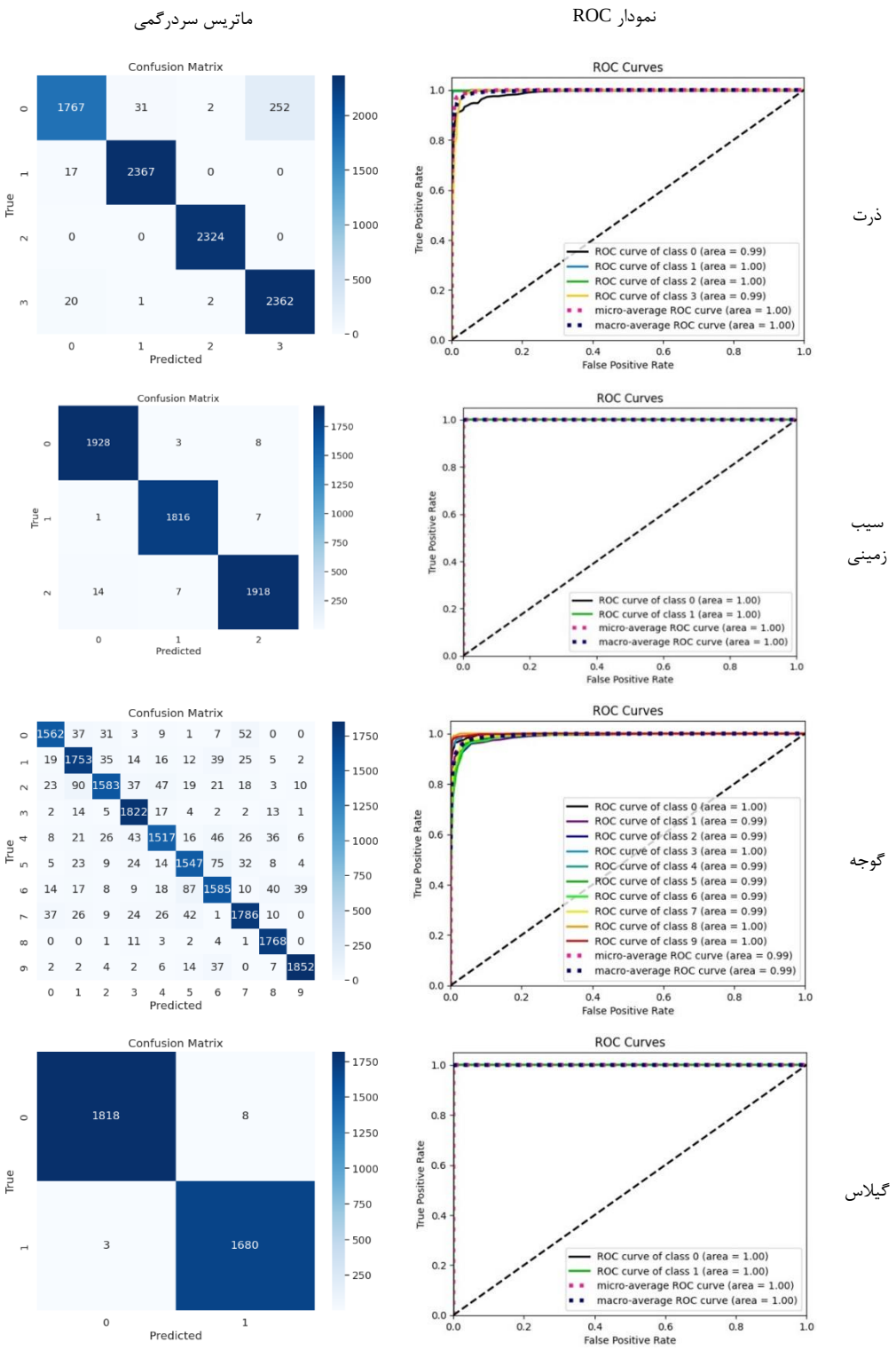
مراجع

[1] V. Singh, N. Sharma, S. Singh, "A review of imaging techniques for plant disease detection" *Artificial Intelligence in Agriculture*, vol. 4, pp. 229-242, 2020, doi: 10.1016/j.aiia.2020.10.002.

[2] A. Khakimov, I. Salakhutdinov, A. Omolikov, S. Utaganov, "Traditional and current-prospective methods of agricultural plant diseases detection: A review" in *Conf. Earth and Environmental Science*, 2022, doi: 10.1088/1755-1315/951/1/012002.

[3] P. Moghadam, D. Ward, E. Goan, S. Jayawardena, P. Sikka, E. Hernandez, " Plant Disease Detection Using Hyperspectral Imaging" in *Conf Digital Image Computing: Techniques and Applications (DICTA)*, pp.1-8, 2017, doi: 10.1109/DICTA.2017.8227476

[4] D. Albashish, M. Braik, S. Bani-Ahmad, "Detection and Classification of Leaf Diseases using K-means-based Segmentation and Neural-networks-based Classification"



شکل ۸: عملکرد روش پیشنهادی روی ۴ نمونه گیاه ذرت، سیب زمینی، گوجه و گیلاس.

Information Technology Journal, vol.10, 2011, doi: 10.3923/itj.2011.267.275.

[5] S. Madiwalar, M. Wyawahare, "Plant disease identification: A comparative study" in *Conf 2017 International Conference on Data Management, Analytics and Innovation (ICDMAI)*, Pune, India, 2017, pp.13-18, doi: 10.1109/ICDMAI.2017.8073478.

[6] M. Li, et al., "High-Performance Plant Pest and Disease Detection Based on Model Ensemble with Inception Module and Cluster Algorithm", *Plants (Basel)*, vol.12, 2023, doi: 10.3390/plants12010200.

[۷] حمید حسن‌پور، سکینه اسدی امیری، "*مفاهیم جامع پردازش تصویر دیجیتال به‌همراه پیاده‌سازی الگوریتم‌ها با متلب*"، انتشارات دانشگاه صنعتی شاهرود، ۱۳۹۵

[8] P. Kulkarni, A. Karwande, T. Kolhe, S. Kamble, A. Joshi, M. Wyawahare, "Plant Disease Detection Using Image Processing and Machine Learning," Nov 2021, doi:10.48550/arXiv.2106.10698.

[9] S. Mohanty, D. Hughes, M. Salathe, "Using Deep Learning for Image-Based Plant Disease Detection," *Frontiers in Plant Science*, vol.7, Sep 2016, doi: 10.3389/fpls.2016.01419.

[10] D.M. Sharath, A. Akhilesh, S.A. Kumar, M.G. Rohan, C. Prathap, "Image based Plant Disease Detection in Pomegranate Plant for Bacterial Blight," in Conf International Conference on Communication and Signal Processing, Chennai, India, 2019, doi: 10.1109/ICCSP.2019.8698007

[11] V. Suresh, M. Krishnan, M. Hemavarthini, D. Jayanthan, "Plant Disease Detection using Image Processing," International Journal of Engineering Research, vol.9 Mar 2020, doi: 10.17577/IJERTV9IS030114.

[12] N. Ibraheem, M. Hasan, R.Z. Khan, P. Mishra, "Understanding Color Models: A Review," *ARPN Journal of Science and Technology*, vol.2, Jan 2012,

[13] J. Miao, L. Niu, "A Survey on Feature Selection," *Procedia Computer Science*. Vol.91, pp.919-926, Jul 2016, doi: 10.1016/j.procs.2016.07.111.

[14] J. Ali, R. Khan, N. Ahmad, I. Maqsood, "Random Forests and Decision Trees," *International Journal of Computer Science Issues(IJCSI)*, vol.9, Sep 2012.

[15] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, A. Lopez, "A comprehensive survey on support vector machine classification: Applications, challenges and trends," *Neurocomputing*, vol.408, pp. 189-215, Sep 2020, doi: 10.1016/j.neucom.2019.10.118.

[16] S. BHATTARAI, (2018), New Plant Diseases Dataset. PlantVillage, [online] Available: https://www.kaggle.com/datasets/vipoooool/new-plant-diseases-dataset?resource=download.

[17] S. Mohanty, (2016), PlantVillage Dataset, Available:https://github.com/spMohanty/PlantVillage-Dataset.

[18] G. Shrestha, Deepsikha, M. Das, N. Dey, "Plant Disease Detection Using CNN," in *Conf IEEE Applied Signal Processing Conference (ASPCON)*, Kolkata, India, 2020, pp.109-113, doi:10.1109/ASPCON49795.2020.9276722

[19] R. Mahum, H. Munir, Z. Mughal, "A novel framework for potato leaf disease detection using an efficient deep learning model," Human and Ecological Risk Assessment: An International Journal, vol.29, pp.303-326, Feb 2023, doi: 10.1080/10807039.2022.2064814

[20] A.I. Khan, S.M.K Quadri, S. Banday, J.L. Shah " Deep diagnosis: A real-time apple leaf disease detection system based on deep learning," Computers and Electronics in Agriculture Journal, vol.198, July 2022, doi: https://doi.org/10.1016/j.compag.2022.107093

[21] X.Gong, S.Zhang " An Analysis of Plant Diseases Identification Based on Deep Learning Methods," *The Plant Pathology Journal*, vol.39(4), 319-334, Aug 2023, doi: https://doi.org/10.5423/PPJ.OA.02.2023.0034

[22] M.Jung, J.S.Song, A.Shin, B.Choi, S.Go " Construction of deep learning-based disease detection model in plants," *Scientific Reports Journal*, vol.13, May 2023, doi: https://doi.org/10.5423/PPJ.OA.02.2023.0034

[23] S.U.Rahman, F.Alam, N.Ahmad, S.Arshad " Image processing based system for the detection, identification and treatment of tomato leaf diseases," *Multimedia Tools and Applications Journal*, vol.82, pp.9431-9445, Sep 2022, doi: https://doi.org/10.1007/s11042-022-13715-0

[24] A.Haridasan, J.Thomas, E.D.Raj " Deep learning system for paddy plant disease detection and classification," *Environmental Monitoring and Assessment Journal*, vol.195, May 2022, doi: https://doi.org/10.1007/s10661-022-10656-x

[25] M.M.Khalid, O.Karan " Deep Learning for Plant Disease Detection," *International Journal of Mathematics, Statistics, and Computer Science*, vol.2, pp.75-84, Nov 2023, doi: https://doi.org/10.59543/ijmscs.v2i.834

[26] I.Ahmed, P.K.Yadav " Plant disease detection using machine learning approaches," *Expert Systems Journal*, vol.40, Oct 2022, doi: https://doi.org/10.1111/exsy.13136

[27] M.Soeb, M.Jubayer, T.Tarin, et al " Tea leaf disease detection and identification based on YOLOv7 (YOLO-T)," *Scientific Reports Journal*, vol.13, Apr 2023, doi: https://doi.org/10.1038/s41598-023-33270-4

پاورقی‌ها:

[1] Accuracy
[2] Otsu
[3] Morphological
[4] LBP
[5] Gray Level Co-Occurrence Matrix (GLCM)
[6] Contour
[7] minimum Redundancy - Maximum Relevance
[8] Random Forrest
[9] Support Vector Machine
[10] Visible and Near-Infrared Range
[11] Short-Wave Infrared Range
[12] RBF: Radial Basis Function
[13] Minimum Distance Classifier
[14] Convolutional Neural Network
[15] Explainable Artificial Intelligence
[16] Gradient-weighted Class Activation Mapping
[17] Long Short-Term Memory
[18] Closing
[19] Moment
[20] Quantize
[21] Mutual Information
[22] Outlier
[23] Confusion Matrix
[24] Receiver Operating Characteristic
[25] Data Augmentation
[26] False Positive Rate
[27] True Positive Rate
[28] Area Under the ROC Curve