



Original Article

Machine Learning Approach to Identify Tomato Bacterial Spot Disease before Symptoms Inclusion Using Hyperspectral Images

Ali Mohammad Shirzadifar^{1*}, Hanieh Abbasi¹, Seyed Mehdi Nassiri¹, Mohammad Amin Zandpour¹, Seyed Mohsen Taghavi², Mohammad Amin Nematollahi¹, Fatemeh Kazemi¹, Sadegh Zarei³

1-Department of Biosystems Engineering, Faculty of Agriculture, Shiraz University, Shiraz, Iran.

2-Department of Plant Protection, Faculty of Agriculture, Shiraz University, Shiraz, Iran.

3-Department of Plant Protection, Faculty of Agriculture, Isfahan University of Technology, Isfahan, Iran.

ARTICLE INFO

ABSTRACT

Keywords:

Bacterial spot disease,
Hyperspectral imaging,
Machine learning,
Support vector machine,
Tomato,

Introduction

Tomato (*Solanum lycopersicum* L.) is a self-pollinating plant from the Solanaceae family, which includes over 3,000 economically significant species. It is one of the most consumed vegetables globally and ranks seventh in importance after crops like maize and rice. Native to the coastal plains of South America, tomatoes were domesticated in Mexico and introduced to Iran in the 19th century. Rich in lycopene, beta-carotene, flavonoids, and vitamin C, tomatoes are renowned for their anti-cancer properties. Globally, tomato production for processing is around 41 million tons, with Asia accounting for about 60% of production. In Iran, about 4.9 million tons were produced in the 2017-2018 agricultural year. Tomato plants are hosts to over 200 pests and plant diseases, with bacterial spot disease being one of the most damaging. The bacterial spot disease can persist in seeds and plant residues for up to 16 months. Detection of bacterial spot disease in early stage is crucial for efficient disease management, however traditional methods like field scouting are inefficient and prone to human error. Recent advancements in spectral imaging, such as hyperspectral imaging, offer a non-destructive, efficient way to detect early stage of plant diseases. This research study explores the use of hyperspectral imagery and machine learning algorithms to detect bacterial spot in tomato leaves before symptoms inclusion, improving early intervention and disease management. The objective of this research study aimed on employing three machine learning algorithms to classify healthy plants versus infected ones using spectral reflectance in the range of 400-800 nm. Before employing machine learning algorithms, the preprocessing methods were employed to remove the noise and improve the classifier algorithm's performance. In order to find the earliest time for identifying bacterial spot disease before symptoms inclusion, the combination of preprocessing methods and classifier algorithms were also evaluated through 7 to 19 days after inoculation.

Materials and Methods

* Corresponding author:

shirzadifar@yahoo.com

In this study, tomato seedlings were sourced from a farm in Bavanat, Fars province, Iran, and transferred to a greenhouse in the Faculty of Agriculture Shiraz University for controlled growth conditions. The plants were inoculated with *Xanthomonas euvesicatoria* pv. *perforans* at the 4-5 leaf

How to cite:

Shirzadifar, MA., Abbasi, H., Nassiri, SM., Zandpour, MA., Taghavi, SM., Nematollahi, MA., Kazemi, F. and Zarei, S. (2025). *Machine Learning Approach to Identify Tomato Bacterial Spot Disease before Symptoms Inclusion Using Hyperspectral Images*. Journal of Agricultural Mechanization, 10 (2):75-89. <https://doi.org/10.22034/jam.2025.66015.1319>.



This is an open-access article under the CC BY NC license
(<https://creativecommons.org/licenses/by-nc/4.0/>)



stage. The research aimed to detect bacterial spot disease before symptoms inclusion, using hyperspectral imaging. Hyperspectral images of 100 tomato samples were taken using hyperspectral camera (Partov Afzar Sanat, Zanjan, Iran). Spectral reflectance of tomato plants was collected in the range of 400–800 nm before inoculation and over the seventh to the nineteenth day after inoculation bacterial spot disease. Data analysis was conducted using Python, where spectral data was preprocessed to remove noise using four methods, including standard normal variate (SNV), multiplicative scatter correction (MSC), first and second derivatives (FD, SD). Principal Component Analysis (PCA) was applied in order to feature reduction. Three machine learning algorithms including random forest (RF), gradient boosting machine (GBM), and support vector machine (SVM) were employed to classify healthy and infected tomato plants. The model's performance was evaluated based on confusion matrix using accuracy, precision, sensitivity, specificity, and F-measure. Accuracy is defined as the ability of the classifier algorithm to detect the healthy and the infected plants correctly. Sensitivity and specificity are defined as the proportion of the healthy or infected plants correctly classified. In the precision calculation, the number of actual predicted infected is divided by the total number of predicted infected plants that were classified as true or false. F-measure defines the harmonic mean of sensitivity and precision where it reaches its best value at 1.0 (perfect precision and sensitivity) and worst value at 0.0. The classification models were validated using a 70-30 split of training and testing data, and the training process was conducted through ten- fold cross validation to ensure reliable results.

Results and Discussion

The analysis of spectral reflectance of infected and healthy plants revealed that the bacterial spot disease has a significant effect on the spectral signature infected plants. The most changes on the spectral reflectance of infected plants were happened in range of 740 to 800 nm, which is part of the NIR area and is related to changes from the structure of the leaf tissue. The results showed that FD, SNV, MSC preprocessing methods significantly improved the classification accuracy of healthy and infected plants over the seventh to the nineteenth day after inoculation bacterial spot disease. The FD preprocessing on the 7 and 10 days after inoculation resulted in the highest accuracy (98%), while MSC and SD methods performed best after 14 and 19 days. The RF, SVM, and GBM classification algorithms could classify the infected plants versus healthy plants with 98%, 100% and 100% accuracy respectively. The results indicated that the 7th day after inoculation was the most reliable and earliest time before symptoms inclusion for classifying infected and healthy tomato plants. The highest classification accuracy was achieved with SVM and GBM algorithms and FD preprocessing method on the 7th day (100%), and SVM with MSC on the 19th day (98%).

Conclusion

The results of this research indicated the ability of machine learning algorithms and hyperspectral imagery for classifying healthy plants versus infected ones with bacterial spot disease in tomato plants before symptoms inclusion. Three classification machine learning algorithms including RF, SVM and RGB could classify infected plants on the 7th day after bacterial spot inoculation with more than 97% accuracy. Therefore, spectral reflectance of potato plant leaves in the range of 400 to 900 nm can be a potential way to identify bacterial spot disease in potato plants in early stage of disease for efficient crop management.



کاربرد فناوری یادگیری ماشین در شناسایی بیماری لکه باکتریایی گوجه فرنگی

علی محمد شیرزادی فر^{۱*}، حانیه عباسی^۱، سیدمهدی نصیری^۱، محمدامین زندپور^۱، سیدمحسن تقوی^۲، محمدامین نعمت‌اللهی^۱، فاطمه کاظمی^۱، صادق زارعی^۲

تاریخ دریافت: ۱۴۰۳/۱۲/۰۳ تاریخ بازنگری: ۱۴۰۴/۰۳/۱۲ تاریخ پذیرش: ۱۴۰۴/۰۳/۲۱

۱- بخش مهندسی بیوسیستم - دانشکده کشاورزی - دانشگاه شیراز - شیراز - ایران

۲- بخش گیاهپزشکی - دانشکده کشاورزی - دانشگاه شیراز - شیراز - ایران

۳- بخش گیاهپزشکی - دانشکده کشاورزی - دانشگاه صنعتی اصفهان - اصفهان - ایران

E-mail: shirzadifar@yahoo.com

* نویسنده مسئول

چکیده

یکی از بیماری‌های مهم در گوجه‌فرنگی بیماری لکه باکتریایی است که با تشخیص به موقع و کنترل آن می‌توان موجب جلوگیری از گسترش آن در سطح مزرعه و کاهش خسارات اقتصادی شد. در این پژوهش از روش غیرمخرب تصویربرداری فراطیفی در محدوده ۴۰۰-۸۰۰ نانومتر به منظور تشخیص بیماری لکه باکتریایی گوجه فرنگی حاصل از باکتری *Xanthomonas perforans* در مراحل اولیه بیماری و پیش از ظهور علائم ظاهری استفاده شد. عکس‌های ابر طیفی و امضای طیفی نمونه‌ها در روزهای هفتم، دهم، چهاردهم و نوزدهم پس از مایه زنی جمع‌آوری شد. در این پژوهش از چهار روش پیش پردازش، تصحیح پراکندگی ضربی، متغیر نرمال استاندارد، مشتق اول و مشتق دوم به منظور حذف نوفه-ها و استخراج اطلاعات مفید برای تحلیل و طبقه‌بندی داده‌ها، استفاده شد. سپس طبقه‌بندی گیاه سالم و بیمار و با استفاده از سه الگوریتم جنگل تصادفی (RF)، ماشین بردار پشتیبان (SVM) و تقویت گرادیان (GBM) در فناوری یادگیری ماشین انجام شد. با توجه به نتایج بدست آمده در روز هفتم و دهم روش مشتق اول و در روز چهاردهم و نوزدهم دو روش مشتق دوم و تصحیح پراکندگی ضربی تأثیر بسیاری بر کیفیت طبقه‌بندی گیاهان سالم و بیمار داشت. نتایج حاصل از طبقه‌بندی داده‌ها نشان داد که هر سه الگوریتم SVM، RF و GBM با دقت بیشتر از ۹۰ درصد به منظور جداسازی دو کلاس گیاه سالم از بیمارانه کردند. روش SVM با دقتی بالای ۹۷ درصد در تمام روزهای موردنظر، به عنوان بهترین روش به منظور جداسازی گیاهان سالم از گیاهان بیمار بدون علائم ارائه کرد. همچنین روز هفتم پس از مایه زنی، ترکیب روش SVM با مشتق اول با صحتی معادل ۱۰۰ درصد به عنوان سریع‌ترین زمان به منظور جداسازی گیاهان سالم از گیاهان بیمار بدون علائم انتخاب شد.

کلمات کلیدی: بیماری لکه باکتریایی، تصویربرداری فراطیفی، گوجه فرنگی، یادگیری ماشین، ماشین بردار پشتیبان

۱- مقدمه

گوجه‌فرنگی با نام علمی *solanum lycopersicum L* گیاهی خودگرده افشان از خانواده *solanaceae* است. این خانواده بیش از ۳۰۰۰ گونه از مهم‌ترین گیاهان اقتصادی را در خود جای داده است و گوجه‌فرنگی از پر مصرف‌ترین سبزیجات جهان در این خانواده است که در طی ۲۰ سال گذشته به عنوان هفتمین گیاه مهم پس از ذرت، برنج، گندم، سیب‌زمینی، سویا و کاساوا قرار گرفته است (Bergougnoux, 2014). این محصول از دشت‌های ساحل غرب آمریکای جنوبی منشأ گرفته و از اکوادور تا شیلی گسترش یافته است و پس از آن در طی زمان در سراسر جهان منتشر شده است. اهلی کردن گوجه‌فرنگی در مکزیک صورت گرفت (Razdan & mattoo, 2005) و برای اولین بار در اوایل قرن نوزدهم در خاورمیانه و در زمان ناصرالدین شاه قاجار وارد ایران شد. این گیاه سرشار از لیکوپن، بتاکاروتن، فلاونوئید و ویتامین، دارای خاصیت ضد سرطانی می‌باشد که به دلیل اهمیت فراوان آن در تغذیه بشر، کیفیت و میزان تولید آن در سراسر جهان افزایش یافته است (Gerszberg et al., 2015). طبق آخرین آمار جهانی تولید گوجه‌فرنگی مخصوص فرآوری، در سراسر جهان حدود ۴۱ میلیون تن و برای تازه خوری حدود ۱۷۰ میلیون تن بوده است که آسیا با حدود ۶۰ درصد تولید گوجه‌فرنگی پیشرو در این امر است (FAO, 2015). همچنین با توجه به آخرین آمار ارائه شده توسط مرکز آمار ایران، میزان تولید گوجه‌فرنگی در کل کشور در طول سال زراعی ۱۴۰۱-۲۰۲ معادل ۵۱۰۰۰۰ تن در سطح زیر کشت ۱۰۵۹۴۳ هکتار بوده است که این آمار بیانگر میزان اهمیت این محصول نه تنها در ایران بلکه در سراسر جهان می‌باشد.

گوجه‌فرنگی میزان مناسبی برای بیش از ۲۰۰ نوع آفات (حشرات و حیوانات کوچک) و بیماری‌های گیاهی (باکتری‌ها، قارچ‌ها، ویروس‌ها، ویروئیدها و اتومایست‌ها) به شمار می‌رود (Bergougnoux, 2014) که مهمترین عامل در کاهش تولید محصولات کشاورزی محسوب می‌شوند. از مهم‌ترین گونه‌های بیماری باکتریایی که باعث خسارت و ضرر زیادی در گوجه‌فرنگی می‌شود و در سال‌های اخیر بسیار شیوع پیدا کرده است بیماری لکه باکتریایی گوجه‌فرنگی است. عامل بیماری لکه باکتریایی در زمان غیاب گیاه می‌تواند تا ۱۶ ماه در بذر باقی بماند و زمستان‌گذرانی در ریشه و ساقه گوجه‌فرنگی مرده صورت می‌گیرد. منبع اولیه این آلودگی بذر و بقایای گیاهی آلوده و همچنین علف‌های هرز میزبان هستند (Madden et al., 2007). معمولاً باکتری‌ها توسط روزنه‌های برگ‌ها، زخم‌های ایجاد شده روی برگ‌ها توسط باد، شن و کارگران نفوذ می‌کنند و در بافت تکثیر می‌نمایند. بطور کلی با آلوده شدن گوشه‌ای از مزرعه بقیه قسمت‌ها نیز به وسیله باد و تراوش آب همراه با باکتری از سطح برگ، آلوده خواهند شد. در نهایت عواملی مانند باد و بارندگی شدید، شب‌بنم سنگین و طولانی،

دمای ۱۹ تا ۲۸ درجه سانتی‌گراد، تراکم بالای آب و رطوبت در سطح برگ و استفاده بیش از حد از کود ازته باعث شیوع و تکثیر بیشتر این بیماری می‌گردد (Miller & Bollen, 1996; Zehr & Shepard, 1994). برگ، شاخه و میوه گیاه قسمت‌هایی از میزبان هستند که مورد حمله باکتری‌ها قرار می‌گیرند (Shepard & Zehr, 1997; Moffett, 1994).

تشخیص و کنترل زود هنگام بیماری برای جلوگیری از پیشرفت آن در مزرعه و تاثیر بر روی کیفیت محصول امری ضروری است. از متداول‌ترین روش‌های سنتی تشخیص بیماری‌های گیاهی گشت زنی در مزرعه و مشاهده بصری توسط افراد خبره می‌باشد که این روش نه تنها به دلیل عدم امکان بررسی روزانه تک تک گیاهان و خطای انسانی بلکه به دلیل هزینه کارگری بالا روش مناسبی نیست. همچنین این روش‌ها نمی‌توانند مناطق وسیعی از مزارع را پوشش دهند و اطلاعات اولیه ضروری را برای فرایند تصمیم‌گیری ارائه دهند (Miller et al., 2009). فرایند تصمیم‌گیری به منظور تشخیص زود هنگام و حتی قبل از بروز علائم بیماری بر روی گیاه امری ضروری است (Mahlein et al., 2012). از این رو فناوری‌های نوین که قابلیت تشخیص علایم بیماری در مراحل اولیه آلودگی گیاه را دارند و باعث توسعه راه‌حل‌هایی جهت نظارت بر سلامت گیاهان و جلوگیری از مصرف بی‌رویه نهاده‌های شیمیایی برای حفاظت از محصولات در برابر بیماری می‌شوند، بسیار مورد توجه قرار گرفته‌اند. یکی از روش‌های نوین و غیرمخرب در تشخیص گیاه بیمار عکس‌برداری با استفاده از دوربین فراطیفی است. از این روش به منظور شناسایی، تشخیص، مدل‌سازی و طبقه‌بندی بیماری‌های گیاهی استفاده می‌شود (Moghadam et al., 2017). تصویربرداری فراطیفی نور بازتابی از گیاهان که به صورت صدها نوار باریک در سراسر طیف الکترومغناطیسی به عنوان یک ابر مکعب است، را اندازه‌گیری می‌کند. این فناوری پتانسیل زیادی در تشخیص زودهنگام بیماری از طریق تصاویر فراطیفی گرفته شده، دارد و می‌تواند پاسخ‌های بیوشیمیایی و فیزیولوژیکی مانند رنگدانه برگ، ترکیبات و محتوای آب با انتساب پیکسلی^۱ علائم خاص بیماری و بافت سالم را تعیین کند (Xuan et al., 2022).

در این پژوهش امکان استفاده از الگوریتم‌های یادگیری ماشین برای تشخیص گیاه آلوده به بیماری لکه باکتریایی گوجه‌فرنگی قبل از بروز علائم ظاهری اولیه، واسنجی شد. عکس‌های ابر طیفی و امضای طیفی نمونه‌های سالم و بیمار در روزهای هفتم، دهم، چهاردهم و نوزدهم پس از مایه زنی جمع‌آوری شد. پس از پیش‌پردازش داده‌های طیفی به منظور حذف نوفه-ها طبقه‌بندی گیاه سالم و بیمار و با استفاده از سه الگوریتم در فناوری یادگیری ماشین انجام شد.

پژوهش از چهار روش پیش پردازش، تصحیح پراکندگی ضربی^۲، متغیر نرمال استاندارد^۳، مشتق اول^۴ و مشتق دوم^۵ به منظور حذف نوفه از داده‌های طیفی استفاده شد (Shirzadifar et al., 2018).

تصحیح پراکندگی ضربی (MSC)، به منظور تصحیح اثرات جمعی، ضربی یا ترکیبی از هردو، انحراف ناشی از اثرات فیزیکی در داده‌های طیفی مانند پخش غیر یکنواخت، ضریب شکست نور و اندازه نمونه و همچنین تصحیح جابه جایی خط پایه مورد استفاده قرار می گیرد (Nicolai et al., 2007). این روش به یک طیف مرجع و یک طیف متوسط نیاز دارد که در آن انحرافات طول موج وابسته از داده‌های طیفی باقی‌مانده جدا شده‌اند. در واقع MSC با انجام عملیات تبدیل خطی از هر طیف توسط دو مرحله تخمین ضرایب تصحیح و محاسبه طیف‌های اصلاح شده به هدف خود دست پیدا می کند که روابط ۱ و ۲ این مراحل را به خوبی نمایش میدهد.

$$X_{org} = b_0 + b_{ref1}X + e \quad (1)$$

$$X_{corr} = \frac{X_{org} - b_0}{b_{ref1}} = X_{ref1} + \frac{e}{b_{ref1}} \quad (2)$$

در روابط بالا X_{org} یک طیف نمونه اصلی اندازه‌گیری شده توسط ابزار طیف‌سنج، X میانگین طیف، b_0 و b_{ref1} مولفه‌های اسکالر، X_{ref1} یک طیف مرجع که برای پیش پردازش کل مجموعه داده استفاده می‌شود، e بخش مدل نشده X_{org} و X_{corr} یا طیف تصحیح شده، هستند.

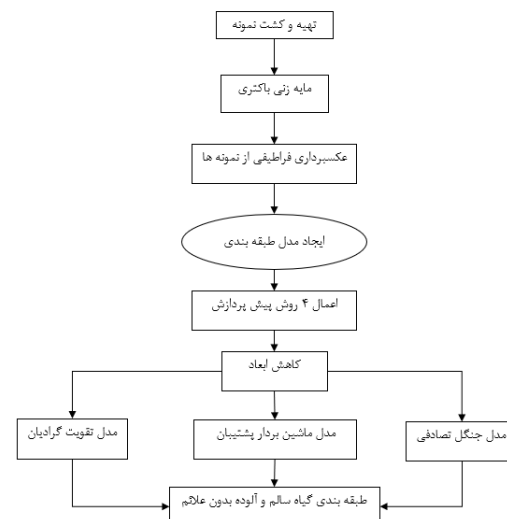
بطور کلی اندازه ذرات، پراکندگی و چند خطی بودن، عوامل اصلی بازدارنده برای تفسیر درست طیف‌های بازتابی هستند. روش توزیع نرمال استاندارد (SNV) با کم کردن میانگین طیف از هر طیف و سپس تقسیم آن بر انحراف استاندارد نمونه طیف، به تصحیح اثرات تغییر فاصله بین آشکارساز و نمونه و همچنین اندازه نمونه پرداخته است (Fu et al., 2007) (۳) رابطه (۳).

$$X_{corr} = \frac{X_{org} - a_0}{a_1} \quad (3)$$

a_0 مقدار متوسط طیف نمونه و a_1 انحراف استاندارد طیف نمونه است. مشتق گیری (FD و SD) یکی از روش‌های پیش پردازش برای افزایش قدرت تفکیک طیف و حذف اثرات افزایشی و ضربی در طیف است که بعد از هموارسازی طیف‌ها استفاده می شود. مشتقات طیفی با به دست آوردن اختلاف بین دو نقطه متوالی به حل هم پوشانی اوج و حذف رانش خطی و ثلث خطی بین نمونه‌ها می پردازد (رابطه ۴).

$$X_{i.fsd} = X_i - X_{i-1} \quad (4)$$

در رابطه ۴، $X_{i.fsd}$ نشان دهنده مشتق مرتبه اول در طول موج i است.



شکل ۱- جریان کار تحقیق

Fig. 1. Research workflow

۲- مواد و روش‌ها

با مراجعه و سرکشی به گلخانه‌های کشت گوجه‌فرنگی استان فارس، نشا با کیفیت موردنظر، از یکی از مزارع گوجه‌فرنگی بولنت تهیه شد و به گلخانه گیاهپزشکی دانشکده کشاورزی در منطقه باجگاه در نزدیکی شیراز انتقال داده شد. تعداد ۱۰۰ نشا در گلدان مناسب کشت شد و در شرایط زیستی و رشد یکسان و در ناحیه‌ای به دور از بقیه گیاهان به منظور جلوگیری از بروز هرگونه آلودگی نگهداری شدند (شکل ۲). همچنین گیاهان گوجه‌فرنگی در مرحله چهار تا پنج برگی با استفاده از باکتری *X. euvesicatoria* pv. *perforans* مایه‌زنی شدند (شکل ۳).

۲-۱- تهیه تصاویر فراطیفی و پیش پردازش آن‌ها

از آنجایی که هدف از این پژوهش تشخیص بیماری لکه باکتریایی پیش از بروز علائم ظاهری روی برگ بود، در روزهای هفتم، دهم، چهاردهم و نوزدهم، پیش از ظهور علائم به ترتیب از ۷۹، ۹۴، ۸۰ و ۶۰ نمونه آلوده به باکتری و ۹۵، ۹۶، ۸۴ و ۹۰ نمونه سالم در شرایط کاملاً یکسان توسط دوربین فراطیفی ساخت شرکت پرتوافزار صنعت ایران (زنجان) (شکل ۴) تصویربرداری فراطیفی انجام گرفت. پس از عکسبرداری فراطیفی و ثبت و ذخیره داده‌ها، با استفاده از نرم‌افزار پایتون ۳/۸ به آنالیز و تحلیل داده‌ها پرداخته شد.

پس از جمع‌آوری داده‌های طیفی، به منظور رسیدن به یک مدل واسنجی درست و قابل اعتماد، حذف نوفه‌ها و استخراج اطلاعات مفید برای تحلیل و طبقه‌بندی داده‌ها، پیش‌پردازش داده‌ها انجام شد. در این



شکل ۲- نشا گوجه فرنگی

Fig. 2. Tomato seedlings



الف) برگ گیاه سالم ب) برگ گیاه آلوده در روز هفتم ج) برگ گیاه آلوده در روز دهم د) برگ گیاه آلوده در روز چهاردهم ه) برگ گیاه آلوده در روز نوزدهم

شکل ۳- برگ گیاه گوجه فرنگی الف) برگ گیاه سالم ب) برگ گیاه آلوده در روز هفتم ج) برگ گیاه آلوده در روز دهم د) برگ گیاه آلوده در روز چهاردهم ه) برگ گیاه آلوده در روز نوزدهم

Fig. 3. Tomato plant leaf a) Healthy plant leaf b) Infected plant leaf on the seventh day c) Infected plant leaf on the day d) Infected plant leaf on the fourteenth day e) Infected plant leaf on the nineteenth day



شکل ۴- تصویر دوربین فرایضی شرکت پرتوافزار صنعت

Fig. 4. The image of the hyperspectral camera of Parto-Afzar-Sanat company

داده‌هایی با ویژگی‌های زیاد می‌باشد. در این روش ابتدا با در نظر گرفتن میانگین دامنه افقی و عمودی تمام نقاط داده، خاستگاه جدید داده‌ها را می‌یابد. بعد از آن PCA یک محور جدید به منظور به حداکثر رساندن واریانس تمام نقاط داده تنظیم می‌کند که این محور جدید در واقع بیشترین فواصل ترکیبی را از نقاط داده دارد. سپس محور بعدی را به صورت متعامد با محور اول تنظیم می‌کند و همچنین واریانس باقیمانده را به حداکثر می‌رساند. در نهایت بعد از نمایش تمام

۲-۲- واکاوی مؤلفه‌های اصلی (PCA) ^۶

در این پژوهش برای تفکیک گیاه سالم از گیاه بیمار از روش واکاوی مؤلفه‌های اصلی (PCA) که یکی از مشهورترین روش‌های کاهش ابعاد است استفاده شد. واکاوی مؤلفه‌های اصلی روشی است که در عین حالی که اطلاعات را تا حد امکان حفظ می‌کند، داده‌های با ابعاد بالاتر را نیز به ابعاد پایین‌تر تبدیل می‌کند. این روش بسیار مناسب

موجود شروع می شود و سپس خطاهای موجود در مدل پایه را شناسایی می کند. پس از شناسایی خطا، یک مدل ثانویه ساخته می شود و در ادامه مدل سوم در این فرآیند معرفی می شود. به این ترتیب، این روند معرفی مدل های بیشتر ادامه می یابد تا به مجموعه داده های آموزشی کاملی دست یابیم که بر اساس آن مدل به درستی پیش بینی می کند.

مجموعه آموزشی در این پژوهش به منظور ساخت مدل طبقه بندی به طور تصادفی گونه های بیمار بدون علائم و سالم در نظر گرفته شد. در هر سه روش ۷۰ درصد از نمونه ها به صورت تصادفی برای ساخت مدل و ۳۰ درصد آنها برای ارزیابی مدل طبقه بندی انتخاب شدند. محاسبه شد.

۲-۴- ارزیابی مدل های طبقه بندی

عملکرد طبقه بندی بر اساس دو کلاس بیمار و سالم با استفاده از ماتریس اغتشاش ارزیابی شد. ماتریس اغتشاش روشی برای خلاصه کردن عملکرد یک الگوریتم طبقه بندی است. به منظور ارزیابی عملکرد مدل، مقدار صحت (رابطه ۵)، حساسیت (رابطه ۶)، دقت (رابطه ۷)، اختصاصیت (رابطه ۸) و معیار F (رابطه ۹) محاسبه شدند. معیارهای صحت، دقت، حساسیت و اختصاصیت همگی در بازه ۰ و ۱ قرار دارند و هر چه به ۱ نزدیک تر باشند، بیانگر ارزیابی بهتر در مدل طبقه بندی هستند.

$$\text{Accuracy} = (TP + TN) / (TP + FP + FN + TN) \quad (5)$$

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

$$\text{Precision} = TP / (TP + FP) \quad (7)$$

$$\text{Specificity (TNR)} = TN / (TN + FP) \quad (8)$$

$$F \text{ measure} = \frac{(\beta^2 + 1) \times \text{sensitivity} \times \text{precision}}{\text{sensitivity} + (\beta^2 \times \text{precision})} \quad (9)$$

TP، بیانگر نمونه بیمار که به درستی بیمار تشخیص داده شده است. TN، بیانگر نمونه سالم که به درستی سالم تشخیص داده شده است. FP، بیانگر نمونه سالم که به غلط بیمار تشخیص داده شده است. FN، بیانگر نمونه بیمار که به غلط سالم تشخیص داده شده است.

نقاط داده بر روی یک صفحه جدید، PCA روابط موجود در داده ها را بین مشاهدات و متغیرها مورد بررسی قرار می دهد. در واقع پراکندگی بیشتر یک ویژگی در مجموعه جدید بیانگر اهمیت آن نسبت به سایر ویژگی ها است. (صفاری، ۱۳۹۸).

۲-۳- طبقه بندی

پس از انجام پیش پردازش داده ها به منظور طبقه بندی صحیح داده ها، سه الگوریتم طبقه بندی جنگل تصادفی^۱ (RF)، تقویت گرادیان^۲ و ماشین بردار پشتیبان^۳ (SVM) از نوع غیرخطی، روی داده ها پیاده سازی شدند و در نهایت مدلی که بهترین نتیجه را برای جداسازی گیاه سالم از بیمار ارائه کرد، انتخاب شد.

جنگل تصادفی (RF) یک روش طبقه بندی کننده نظارت شده و غیرپارامتری است که به عنوان یک روش دقیق یادگیری ماشین استفاده می شود. جنگل تصادفی متشکل از چندین درخت تصمیم است که هر کدام از این درخت ها به صورت تصادفی از زیرمجموعه ای از داده ها، به عنوان داده آموزشی استفاده می کند. در نهایت نتیجه پیش بینی شده نهایی با اکثریت آرای مجموعه درخت های تصمیم مشخص می شود. از مهمترین مزیت های الگوریتم جنگل تصادفی جلوگیری از برازش بیش از حد در اکثر مواقع با تولید درختان کوچکتر با استفاده از زیرمجموعه های تصادفی از ویژگی است (Shirzadifar et al., 2020).

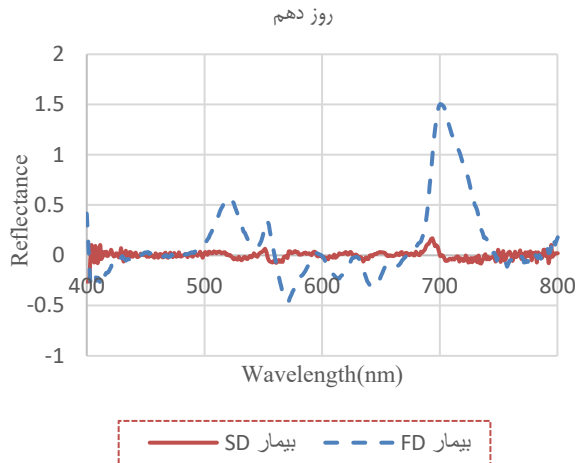
ماشین بردار پشتیبان (SVM) یک الگوریتم نظارت شده یادگیری ماشین است که وظایف طبقه بندی را با ساختن یک تقسیم کننده انجام می دهد که داده ها را در دو دسته جدا می کند. تقسیم کننده بهینه تقسیم کننده ای است که در فاصله مساوی از مرزهای هر گروه باشد. در واقع هدف از الگوریتم ماشین بردار پشتیبان، یافتن یک ابر صفحه در یک فضای N بعدی (N تعداد ویژگی ها) است که به طور مشخص نقاط داده را طبقه بندی می کند. این روش به دو صورت خطی و غیرخطی انجام می شود. روش غیرخطی این الگوریتم با استفاده از جداسازی گروه ها توسط یک ابر صفحه خاص باعث به حداکثر رساندن مقدار حاشیه می شود (Rumpf et al., 2010). در این پژوهش از روش غیرخطی این الگوریتم استفاده شده است.

الگوریتم تقویت گرادیان (GBM) از روش های یادگیری ماشین برای طبقه بندی است. این روش یکی از الگوریتم های مدل سازی مجموعه یادگیری ماشین است که برای ساخت طبقه بندی کننده های قوی از طبقه بندی کننده های ضعیف مختلف استفاده می شود (Bahad et al., 2020). با ساختن یک مدل اولیه از مجموعه داده های آموزشی

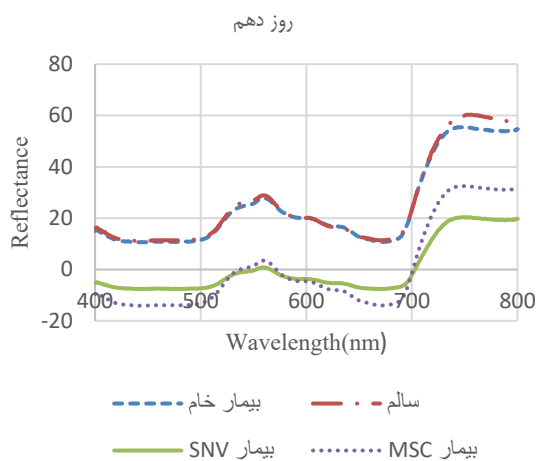
۳- نتیجه گیری

۳-۱- مقایسه طیف‌ها بر روی نمودار

میانگین طیف‌های برداشت شده در بازه ۴۰۰-۸۰۰ نانومتر از دسته سالم، بیمار خام و بیمار با اعمال پیش پردازش‌های مختلف در روزهای هفتم، دهم، چهاردهم و نوزدهم در شکل‌های ۵ تا ۸ نشان داده شده است. با توجه به شکل تفاوت ظاهری طیف نمونه‌ها به خوبی نمایش داده شده است.



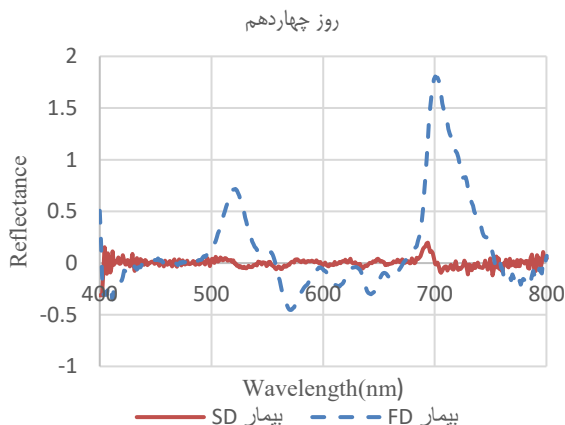
الف



ب

شکل ۶- مقایسه میانگین امضا طیفی روز دهم الف) بیمار با پیش-پردازش FD و SD ب) گروه سالم و بیمار خام و بیمار با پیش-پردازش MSC و SNV

Fig. 6. Comparison of the average spectral signature on the 10th day a) patient with FD and SD pre-processing b) healthy group and raw patient and patient with SNV and MSC pre-processing



ب

شکل ۵- تصویر مقایسه میانگین امضا طیفی روز هفتم الف) بیمار با پیش-پردازش FD و SD ب) گروه سالم و بیمار خام و بیمار با پیش-پردازش MSC و SNV

Fig. 5. Comparison image of the average spectral signature on the seventh day a) patient with FD and SD pre-processing b) healthy group and raw patient and patient with SNV and MSC pre-processing

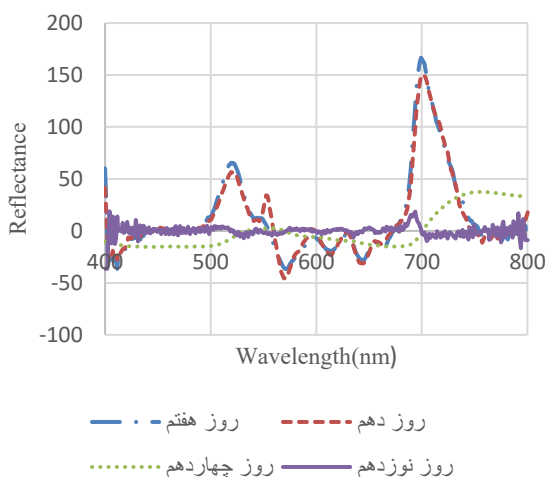
ب

شکل ۸- تصویر مقایسه میانگین امضا طیفی روز نوزدهم الف) بیمار با پیش پردازش FD و SD ب) گروه سالم و بیمار خام و بیمار با پیش پردازش SNV و MSC

Fig. 8. Comparison image of the average spectral signature on the 19th day a) Patient with FD and SD pre-processing b) Healthy group and raw patient and patient with SNV and MSC pre-processing

همچنین در شکل ۹ به مقایسه میانگین طیف های برداشته شده در روزهای موردنظر با اعمال بهترین پیش پردازش مناسب آن روز نشان داده شده است.

امضا طیفی هر چهار روز با اعمال بهترین پیش پردازش



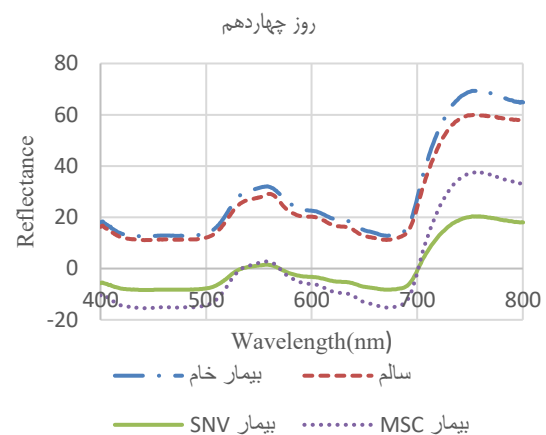
شکل ۹- مقایسه میانگین طیف های برداشته شده در روزهای موردنظر با اعمال بهترین پیش پردازش مناسب آن روز

Fig. 9. Comparison of the average spectra taken on the desired days by applying the best suitable pre-processing for that day

اعمال روش های پیش پردازش تأثیر بسیاری بر کیفیت طبقه بندی دسته های مختلف از یکدیگر داشت. با توجه به نتایج بدست آمده در روز هفتم و دهم روش FD و در روز چهاردهم و نوزدهم روش SD و MSC عملکرد بسیار خوبی داشتند و نتایج خوبی را در طبقه بندی ارائه کردند. طبقه بندی بر روی بازه ی طول موج ۴۰۰ تا ۸۰۰ نانومتر انجام شد. در پژوهش های پیشین نیز پیش پردازش مشتق به منظوراز بین بردن نوفه ها و افزایش دقت طبقه بندی مورد استفاده قرار گرفته است. برای مثال در سال ۲۰۱۸ سیدلیسکا و همکاران مشتق دوم طیف اصلی را برای تشخیص عفونت در توت فرنگی و برای ساخت مدل های طبقه بندی نظارت شده مورد استفاده قرار دادند.

طیف های الکترومغناطیس شامل گستره ای از تمام تشعشعات الکترومغناطیسی است که بر اساس تفاوت رفتار در انتشار، انتقال و جذب امواج مربوطه از بخش های مختلفی مانند نور مرئی یا اشعه

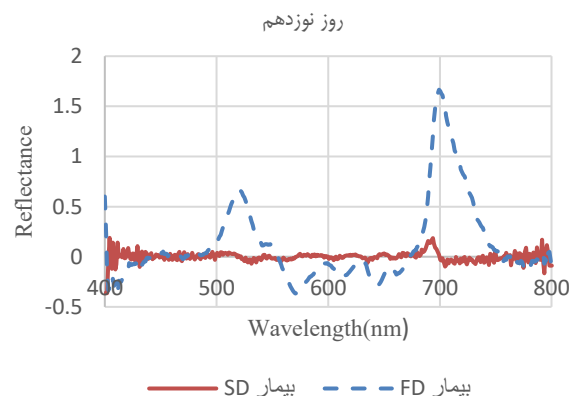
الف



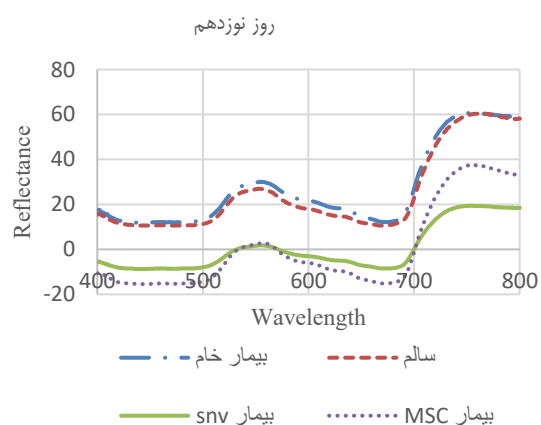
ب

شکل ۷- مقایسه میانگین امضا طیفی روز چهاردهم الف) بیمار با پیش پردازش FD و SD ب) گروه سالم و بیمار خام و بیمار با پیش پردازش SNV و MSC

Fig. 7. Comparison of the average spectral signature on the fourteenth day a) patient with FD and SD pre-processing b) healthy group and raw patient and patient with SNV and MSC pre-processing

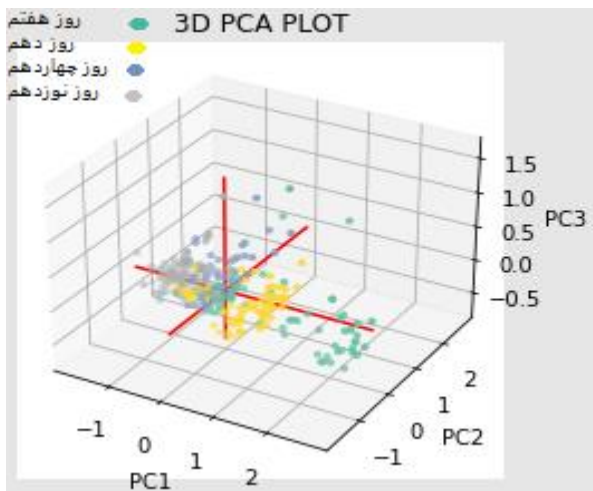


الف



شکل ۱۰- نمودار دو بعدی PCA امتیازهای PCA در روزهای مختلف پس از مایه زنی

Figure 10- Two-dimensional scatter plot of PCA scores on different days after inoculation



شکل ۱۱- نمودار سه بعدی PCA امتیازهای PCA در روزهای مختلف پس از مایه زنی

Figure 11- 3D scatter diagram of PCA scores on different days after inoculation

۳-۲-۲- نتایج حاصل از طبقه بندی

در این بخش نتایج حاصل از تحلیل طبقه بندی های جنگل تصادفی ، ماشین بردار پشتیبان و روش دسته بندی تقویت گرادیان برای جدا کردن گیاه بیمار از گیاه سالم با استفاده از روش های مختلف پیش پردازش آورده شده است.

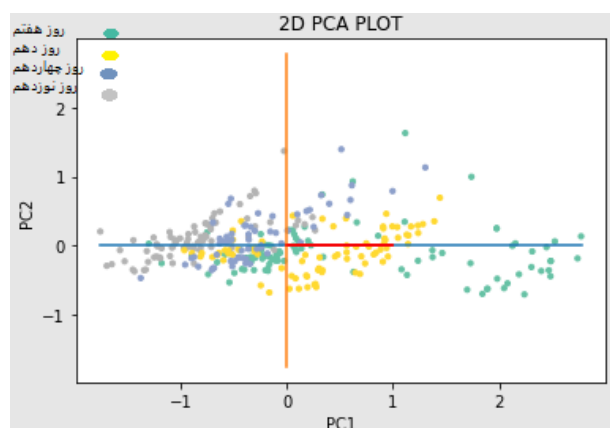
۳-۲-۱- نتایج الگوریتم طبقه بندی جنگل تصادفی

همان طور که گفته شد در طبقه بندی با استفاده از الگوریتم جنگل تصادفی مقدار ۳۰ درصد از نمونه های موجود در هر دسته برای آزمایش و ۷۰ درصد به منظور مدلسازی استفاده شدند. لازم به ذکر است که به منظور جلوگیری از بیش برازش یا کم برازش داده ها تعداد ۱۵ مولفه اصلی برای کاهش ابعاد داده ها در نظر گرفته شد. با توجه به نتایج حاصل از تحلیل طبقه بندی جنگل تصادفی در جدول ۱، مشاهده شد که روش پیش پردازش مشتق اول در روز هفتم و دهم با صحتی معادل ۰/۹۸، روش MSC در روز چهاردهم با صحتی معادل ۰/۹۶ و روش SD در روز نوزدهم با صحتی معادل ۰/۹۷ بهترین نتیجه را برای طبقه بندی گیاه سالم و بیمار ارائه کردند.

در نهایت از آنجایی که تشخیص هرچه زودتر بیماری گیاهی زمان بیشتری برای مقابله با آن را فراهم می کند، لذا روش FD که بهترین نتیجه را در روز هفتم و دهم به همراه داشت به عنوان بهترین روش پیش پردازش الگوریتم جنگل تصادفی در نظر گرفته شد.

ماوراء بنفش تشکیل شده است. امضای طیفی حاصل از این طیف ها در واکاوی واکنش تنش های گیاهی مختلف بر روی گیاهان بسیار کارآمد است. جذب و بازتاب طول موج های مختلف نور به ساختار سلولی برگ در گیاه مربوط می شود. بدین صورت که نور قرمز و آبی توسط رنگدانه های کلروفیل جذب می شوند و نور سبز توسط این رنگدانه ها بازتاب می شود. تنش های گیاهی بر میزان بازتاب نور در محدوده های طول موج مختلف تاثیر می گذارد (Al-Saddik et al., 2017). به طور کلی هر ناحیه از طول موج های بازتابی از گیاه بیانگر یک سری ویژگی های خاص گیاه است. ناحیه ۴۰۰ تا ۷۰۰ نانومتر (مرئی) مربوط به رنگدانه های فتوسنتزی در گیاه است. طول موج های ۵۵۰ تا ۶۸۰ نانومتر و ناحیه لبه قرمز (۶۸۰ تا ۷۳۰ نانومتر)، مربوط به محتوای کلروفیل و بازتاب طیفی در محدوده ۶۷۵ و ۵۵۰ نانومتر غلظت کلروفیل را به ما نشان می دهد (Shirzadifar et al., 2020). با استفاده از میانگین طیف های برداشت شده در بازه ۸۰۰-۴۰۰ نانومتر از دسته سالم، بیمارخام و بیمار با اعمال پیش پردازش های SD و MSC در روز های هفتم، دهم، چهاردهم و نوزدهم، دریافت شد که در روز های چهاردهم و نوزدهم دو روش SD و MSC عملکرد بسیار خوبی دارند و تفاوت ظاهری طیف نمونه ها واضح تر است. بیماری های گیاهی تاثیر به سزایی بر طول موج بازتاب شده گیاهان دارد و گیاهان تحت تنش امضای طیفی متفاوتی را نسبت به گیاهان سالم دارا هستند. همان طور که در نمودار ها قابل مشاهده است اکثر طول موج های قابل تفکیک بین گیاهان سالم و بیمار مربوط به محدوده طول موج ۷۴۰ تا ۸۰۰ نانومتر است که این محدوده مربوط به بخشی از ناحیه NIR است و مربوط به تغییرات ناشی از ساختار بافت برگ است یعنی در واقع بیماری لکه باکتریایی گوجه فرنگی تاثیر بسیاری بر ساختار برگ گیاه می گذارد.

دسته بندی طیف های مختلف با استفاده از ۱۵ مولفه اصلی نخست انجام شد و در مجموع ۹۵ درصد از واریانس کل را پوشش دادند. نمودار دو بعدی و سه بعدی پراکندگی امتیازهای PCA که از اعمال بهترین پیش پردازش بدست آمده است به منظور مشاهده بهتر در شکل ۱۰ و ۱۱ نشان داده شده است.



۲-۳- نتایج الگوریتم طبقه‌بندی SVM

در این پژوهش از روش غیرخطی الگوریتم SVM استفاده شد. نتایج حاصل از اعمال این طبقه‌بندی بر روی داده‌های پیش پردازش شده در بازه‌ی ۸۰۰-۴۰۰ نانومتر در جدول ۲ نشان داده شده است. بهترین نتایج به دست آمده با توجه به روز عکسبرداری و روش پیش پردازش متفاوت بود. در روز هفتم روش مشتق اول با صحتی معادل ۱/۰۰، در روز دهم روش مشتق اول و مشتق دوم با صحتی معادل ۱/۰۰، در روز چهاردهم روش مشتق دوم با صحتی معادل ۰/۹۶ و در روز نوزدهم روش MSC با صحتی معادل ۰/۹۸ بهترین نتیجه را به همراه داشتند. بطور کلی در این الگوریتم تمام روش‌های پیش پردازش صحتی بالاتر از ۰/۹۲ را ارائه کردند.

۳-۲-۳ نتایج الگوریتم طبقه‌بندی تقویت گرادیان

همان طور که گفته شد این روش (تقویت گرادیان) ترکیبی خطی از یک سری مدل‌های ضعیف است که به صورت تناوبی برای ایجاد یک مدل نهایی قوی ساخته شده است. نتایج حاصل از روش تقویت گرادیان در جدول ۳ ارائه شده است. همانند هر دو الگوریتم طبقه‌بندی دیگر در این الگوریتم نیز روش پیش پردازش FD بهترین نتایج را برای روز هفتم و دهم به همراه داشت و صحتی معادل ۱/۰۰ و ۰/۹۸ را به ترتیب ارائه کرد. در روز چهاردهم و نوزدهم نیز روش SNV با صحتی معادل ۰/۹۴ و ۰/۹۷ بهترین نتیجه را داشت.

پس از بررسی نتایج مربوط به طبقه‌بندی اعمال شده بر روی داده‌ها، هر سه روش جنگل تصادفی، SVM و تقویت گرادیان نتایج بسیار خوبی را ارائه کردند اما این نتایج با توجه به روش پیش پردازش اعمال شده و روز عکسبرداری تا حدی متفاوت بود. در روز هفتم هر دو الگوریتم SVM و تقویت گرادیان با پیش پردازش FD، در روز دهم الگوریتم SVM با پیش پردازش FD، در روز چهاردهم الگوریتم جنگل تصادفی با پیش پردازش MSC و روز نوزدهم الگوریتم SVM با پیش پردازش MSC بهترین نتایج با بالاترین مقدار صحت را به همراه داشتند. در نهایت روش SVM به عنوان روشی با بهترین عملکرد برای جداسازی گیاه سالم از گیاه آلوده پیش از ظهور علائم ظاهری شناخته شد. SVM با نگاشت داده‌ها به یک فضای ویژگی با ابعاد بالا کار می‌کند تا در نهایت بتوان نقاط داده را طبقه‌بندی کرد، حتی زمانی که داده‌ها به طور خطی قابل جداسازی نیستند. با بررسی چندین پژوهش مختلف در این زمینه دیده شد که روش SVM در طبقه‌بندی انواع بیماری‌های گیاهی بر روی گیاهان مختلف نیز از دقت بالایی برخوردار بوده است. برای مثال کاظمی در سال ۱۴۰۰ بعد از بررسی چندین روش طبقه‌بندی به منظور جداسازی گیاه آلوده به ویروس موزایک زرد کدو، روش SVM را به عنوان بهترین روش

طبقه‌بندی معرفی کرد (Kazemi, 2021). همچنین در سال ۲۰۲۰ گائو و همکاران از تصویربرداری فراطیفی برای تشخیص غیر مخرب بیماری پیچیدگی برگ انگور توسط ویروس (GLRaV-3) در مراحل پیش از بروز علائم استفاده کردند. عکس برگ‌های انگور آلوده و غیر آلوده در ۵ مرحله فنولوژیکی توسط دوربین فراطیفی تهیه شده بود. در این پژوهش برای طبقه‌بندی از روش ماشین بردار پشتیبان استفاده شده که دقتی در حدود ۶۶.۶۷ تا ۸۹.۹۳ درصد بدست آمد و از این رو نتیجه گرفته شده که تصویربرداری فراطیفی روش مناسبی برای تشخیص غیر مخرب بیماری پیچیدگی برگ انگور پیش از بروز علائم می‌باشد (Z. Gao et al., 2020).

در سال ۲۰۱۰ رامپف و همکاران، روش SVM را با دقت طبقه‌بندی ۹۷ درصد به عنوان یک روش بسیار کارآمد به منظور تمایز بین بیماری‌های سفیدک پودری، زنگ زدگی و لکه برگ چغندر قند پیش از ظهور علائم ظاهری ارائه کردند (Rumpf et al., 2010). با بررسی جدول‌های مربوط به نتایج طبقه‌بندی نتیجه گرفتیم که با وجود اختلاف بسیار کم در نتایج حاصل از روزهای مختلف، طیف برداری در روزهای ابتدایی آلودگی کارآمدتر است. این مسئله باعث تشخیص سریع‌تر بیماری و زمان بیشتر به منظور کنترل بیماری پیش از گسترش آن است.

در این پژوهش سینتیک خشک‌شدن، مدل‌سازی خشک‌کردن لایه نازک و انتقال جرم داخلی برای خشک‌شدن برگ‌های سیب در داخل یک خشک‌کن خورشیدی از نوع RW-PTC ارائه شد. آزمایش‌ها به سه روش مرسوم (RW)، ترکیبی (PRW) و کاملاً خورشیدی (SRW) انجام شد. نتایج نشان داد که با افزایش دما از ۶۵ به ۸۵°C زمان خشک‌کردن در روش اول ۵۶.۲۵٪ (از ۳۲۰ دقیقه به ۱۴۰) و در روش دوم ۳۸.۴۶٪ (از ۲۶۰ دقیقه به ۱۶۰) کاهش یافت. زمان خشک‌کردن محصول در روش سوم که کاملاً با انرژی خورشیدی فرایند گرم‌کردن آب انجام شد ۲۴۰ دقیقه (۴ ساعت) بود. در برازش داده‌ها با مدل میدلی مقدار میانگین R²، SSE و RMSE به ترتیب ۰.۰۳۱، ۰.۰۶۳ و ۰.۰۶۲ بدست آمد. بعد از مدل میدلی به ترتیب مدل‌های پیچ اصلاح‌شده، هندرسون، لگاریتمی و نیوتن دقت بیشتری در توصیف خروج رطوبت از محصول داشتند. کمترین و بیشترین ضریب نفوذ مؤثر رطوبت به دست‌آمده در آزمایش‌های انجام شده به ترتیب در روش‌های RW-۶۵ و RW-۸۵ به مقدار 8.93×10^{-13} و 4.82×10^{-12} بدست آمد. مقدار ضریب نفوذ مؤثر در روش‌های خشک‌کردن ترکیبی (PRW) در محدوده $1.49 - 3.87 \times 10^{-12}$ قرار داشت. همچنین مقدار ضریب نفوذ مؤثر رطوبت در روش خورشیدی SRW به میزان 1.74×10^{-12} بدست آمد. بطور کلی نتایج نشان داد که خشک‌کن ترکیبی خورشیدی ارائه شده قابلیت مناسبی برای خشک‌کردن محصول و کاهش مصرف انرژی از منابع تجدیدناپذیر دارد.

جدول ۱- نتایج حاصل از طبقه‌بندی الگوریتم جنگل تصادفی برای دو گروه برگ گیاه گوجه‌فرنگی سالم و بیمار
Table 1. Classification results of random forest algorithm for two groups of healthy and diseased tomato plant leaves

روز	طول موج	پیش پردازش	کلاس	PCs	صحت	حساسیت	اختصاصیت	دقت	معیار F
روز هفتم	۴۰۰-۸۰۰	RAW	سالم	۱۵	۰/۹۰	۰/۹۱	۰/۸۶	۰/۹۱	۰/۹۱
		MSC	(n=۲۵)		۰/۸۱	۰/۸۱	۰/۷۲	۰/۸۳	۰/۸۱
		SNV	بیمار (n=۲۷)		۰/۹۲	۰/۹۲	۰/۸۷	۰/۹۲	۰/۹۲
		FD			۰/۹۸	۰/۹۸	۰/۹۶	۰/۹۸	۰/۹۸
		SD			۰/۹۰	۰/۹۰	۰/۸۵	۰/۹۱	۰/۹۰
روز دهم	۴۰۰-۸۰۰	RAW	سالم	۱۵	۰/۸۸	۰/۸۸	۰/۹۰	۰/۸۹	۰/۸۸
		MSC	(n=۲۷)		۰/۸۲	۰/۸۲	۰/۸۸	۰/۸۳	۰/۸۲
		SNV	بیمار (n=۲۴)		۰/۸۴	۰/۸۴	۰/۸۵	۰/۸۶	۰/۸۴
		FD			۰/۹۸	۰/۹۸	۰/۹۸	۰/۹۸	۰/۹۸
		SD			۰/۹۱	۰/۹۱	۰/۹۰	۰/۹۱	۰/۹۱
روز چهاردهم	۴۰۰-۸۰۰	RAW	سالم	۱۵	۰/۸۶	۰/۸۷	۰/۹۰	۰/۹۰	۰/۸۷
		MSC	(n=۲۹)		۰/۹۶	۰/۹۶	۰/۹۹	۰/۹۷	۰/۹۶
		SNV	بیمار (n=۲۴)		۰/۹۴	۰/۹۴	۰/۹۸	۰/۹۸	۰/۹۴
		FD			۰/۹۲	۰/۹۲	۰/۹۵	۰/۹۴	۰/۹۲
		SD			۰/۹۲	۰/۹۲	۰/۹۶	۰/۹۳	۰/۹۲
روز نوزدهم	۴۰۰-۸۰۰	RAW	سالم	۱۵	۰/۹۱	۰/۹۱	۰/۹۵	۰/۹۲	۰/۹۲
		MSC	(n=۲۸)		۰/۹۴	۰/۹۴	۰/۹۴	۰/۹۴	۰/۹۴
		SNV	بیمار (n=۱۹)		۰/۸۹	۰/۸۹	۰/۹۰	۰/۹۲	۰/۸۹
		FD			۰/۸۹	۰/۸۹	۰/۸۴	۰/۸۹	۰/۸۹
		SD			۰/۹۷	۰/۹۸	۰/۹۵	۰/۹۸	۰/۹۸

جدول ۲- نتایج حاصل از طبقه‌بندی الگوریتم SVM برای دو گروه برگ گیاه گوجه‌فرنگی سالم و بیمار
Table 2. The results of SVM algorithm classification for two groups of healthy and diseased tomato plant leaves

روز	طول موج	پیش پردازش	کلاس	PCs	صحت	حساسیت	اختصاصیت	دقت	معیار F
روز هفتم	۴۰۰-۸۰۰	RAW	سالم	۱۵	۰/۹۸	۰/۹۸	۰/۹۹	۰/۹۸	۰/۹۸
		MSC	(n=۲۵)		۰/۹۶	۰/۹۶	۰/۹۸	۰/۹۷	۰/۹۶
		SNV	بیمار (n=۲۷)		۰/۹۶	۰/۹۶	۰/۹۶	۰/۹۶	۰/۹۶
		FD			۱/۰۰	۱/۰۰	۱/۰۰	۱/۰۰	۱/۰۰
		SD			۰/۹۲	۰/۹۲	۰/۸۶	۰/۹۳	۰/۹۲
روز دهم	۴۰۰-۸۰۰	RAW	سالم	۱۵	۰/۹۸	۰/۹۸	۱/۰۰	۰/۹۸	۰/۹۸
		MSC	(n=۲۷)		۰/۹۲	۰/۹۲	۰/۹۶	۰/۹۲	۰/۹۲
		SNV	بیمار (n=۲۴)		۰/۹۸	۰/۹۸	۰/۹۶	۰/۹۸	۰/۹۸
		FD			۱/۰۰	۱/۰۰	۱/۰۰	۱/۰۰	۱/۰۰
		SD			۱/۰۰	۱/۰۰	۱/۰۰	۱/۰۰	۱/۰۰
روز چهاردهم	۴۰۰-۸۰۰	RAW	سالم	۱۵	۰/۹۲	۰/۹۲	۰/۹۷	۰/۹۴	۰/۹۲
		MSC	(n=۲۹)		۰/۹۴	۰/۹۴	۰/۹۸	۰/۹۵	۰/۹۴
		SNV	بیمار (n=۲۴)		۰/۹۲	۰/۹۲	۰/۹۷	۰/۹۴	۰/۹۲
		FD			۰/۹۲	۰/۹۲	۰/۹۸	۰/۹۷	۰/۹۶
		SD			۰/۹۶	۰/۹۶	۰/۹۹	۰/۹۲	۰/۸۹
روز نوزدهم	۴۰۰-۸۰۰	RAW	سالم	۱۵	۰/۸۹	۰/۸۹	۰/۹۹	۰/۹۲	۰/۸۹
		MSC	(n=۲۸)		۰/۹۸	۰/۹۸	۰/۹۸	۰/۹۸	۰/۹۸
		SNV	بیمار (n=۱۹)		۰/۹۵	۰/۹۶	۰/۹۹	۰/۹۶	۰/۹۶
		FD			۰/۹۳	۰/۹۴	۰/۹۵	۰/۹۴	۰/۹۴
		SD			۰/۹۶	۰/۹۶	۰/۹۹	۰/۹۶	۰/۹۶

جدول ۳- نتایج حاصل از طبقه‌بندی الگوریتم تقویت گرادیان برای دو گروه برگ گیاه گوجه فرنگی سالم و بیمار

Table 3. The classification results of the gradient boosting algorithm for two groups of healthy and diseased tomato plant leaves

روز	طول موج	پیش پردازش	کلاس	PCs	صحت	حساسیت	اختصاصیت	دقت	معیار F
روز هفتم	۴۰۰-۸۰۰	RAW	سالم	۱۵	۰/۹۰	۰/۹۱	۰/۹۳	۰/۹۱	۰/۹۱
		MSC	(n=۲۵)		۰/۸۷	۰/۸۷	۰/۸۶	۰/۸۷	۰/۸۷
		SNV	بیمار		۰/۹۶	۰/۹۶	۰/۹۸	۰/۹۶	۰/۹۶
		FD	(n=۲۷)		۱/۰۰	۱/۰۰	۱/۰۰	۱/۰۰	۱/۰۰
		SD			۰/۸۹	۰/۸۸	۰/۸۶	۰/۸۸	۰/۸۸
روز دهم	۴۰۰-۸۰۰	RAW	سالم	۱۵	۰/۹۶	۰/۹۶	۰/۹۵	۰/۹۶	۰/۹۶
		MSC	(n=۲۵)		۰/۹۰	۰/۹۰	۰/۸۹	۰/۹۰	۰/۹۰
		SNV	بیمار		۰/۹۲	۰/۹۲	۰/۸۸	۰/۹۲	۰/۹۲
		FD	(n=۲۶)		۰/۹۸	۰/۹۸	۰/۹۶	۰/۹۸	۰/۹۸
		SD			۰/۹۸	۰/۹۸	۰/۹۶	۰/۹۸	۰/۹۸
روز چهاردهم		RAW	سالم	۱۵	۰/۸۸	۰/۸۹	۰/۹۲	۰/۹۱	۰/۸۹
		MSC	(n=۲۹)		۰/۹۲	۰/۹۲	۰/۹۳	۰/۹۴	۰/۹۲
		SNV	بیمار (n=۲۴)		۰/۹۴	۰/۹۴	۰/۹۶	۰/۹۵	۰/۹۴
		FD			۰/۹۲	۰/۹۲	۰/۹۷	۰/۹۴	۰/۹۲
		SD			۰/۹۰	۰/۹۱	۰/۹۱	۰/۹۱	۰/۹۱
روز نوزدهم		RAW	سالم	۱۵	۰/۹۱	۰/۹۱	۰/۹۴	۰/۹۲	۰/۹۲
		MSC	(n=۲۸)		۰/۹۰	۰/۹۱	۰/۸۹	۰/۹۱	۰/۹۰
		SNV	بیمار		۰/۹۷	۰/۹۸	۰/۹۹	۰/۹۸	۰/۹۸
		FD	(n=۱۹)		۰/۹۳	۰/۹۴	۰/۹۵	۰/۹۴	۰/۹۴
		SD			۰/۹۳	۰/۹۴	۰/۹۴	۰/۹۴	۰/۹۴

بالا را در طبقه‌بندی ارائه کردند. به صورت کلی با توجه به مقایسه نتایج در جدول ۳، مشتق اول به عنوان پیش‌پردازش در آنالیز طیف‌های گیاه گوجه‌فرنگی پیشنهاد می‌شود. همچنین با توجه به این که طبقه‌بندی در روزهای هفتم و دهم بهترین نتیجه را به همراه داشت، بنابراین قادر به تشخیص هرچه زودتر بیماری و کنترل بهتر آن خواهیم بود.

منابع

- Abdulridha, J., Ampatzidis, Y., Kakarla, S. C., & Roberts, P. (2020). *Detection of target spot and bacterial spot diseases in tomato using UAV-based and benchtop-based hyperspectral imaging techniques*. Precision Agriculture, 21(5), 955–978. <https://doi.org/10.1007/s11119-019-09703-4>.
- Al-Saddik, H., Simon, J.-C., & Cointault, F. (2017). *Development of spectral disease indices for 'Flavescence Dorée' grapevine disease identification*. Sensors, 17(12), 2772. <https://doi.org/10.3390/s17122772>.
- Barbin, D. F., G. ElMasry, D. W. Sun, and P. Allen. (2012). *Predicting quality and sensory attributes of pork using near-infrared hyperspectral imaging*. Anal. Chim. Acta 719: 30-42. <https://doi.org/10.1016/j.aca.2012.01.004>.
- Bergougnoux, V. (2014). *The history of tomato: From domestication to biopharming*. Biotechnology Advances,

۴- نتیجه گیری

شناسایی و کنترل بیماری‌های گیاهی با روش گشت‌زنی و پایش در مزرعه و گلخانه توسط افراد خبره می‌باشد. این روش مناسب تولیدکنندگان باتجربه و متخصص است و کشاورزان بی‌تجربه با مشاهده بصری موفق به شناسایی بیماری گیاهی نمی‌شوند، از طرفی آزمون‌های آزمایشگاهی بسیار گران قیمت هستند و به دلیل تاخیر در شناسایی منجر به از دست رفتن بخش زیادی از محصول می‌شوند. نتایج این پژوهش نشان داد که روش تصویربرداری ابرطیفی دارای پتانسیل تشخیص سریع و غیر تهاجمی به منظور تفکیک گیاهان سالم و آلوده به بیماری لکه باکتریایی گوجه فرنگی قبل از بروز علائم ظاهری می‌باشد. بدین منظور از سه روش طبقه‌بندی جنگل تصادفی، ماشین بردار پشتیبان و تقویت گرادیان استفاده شد. هر سه روش دقت خوبی را به منظور جداسازی دو کلاس مورد نظر ارائه کردند اما روش ماشین بردار پشتیبان با دقتی بالای ۹۷ درصد در تمام روزهای مورد نظر، به عنوان بهترین روش به منظور جداسازی گیاهان سالم از گیاهان بیمار بدون علائم پیش از شیوع بیماری و در روزهای اولیه آلودگی، انتخاب شد. لازم به ذکر است اعمال روش‌های پیش پردازش تأثیر بسیاری بر کیفیت طبقه‌بندی دسته‌های مختلف از یکدیگر داشت. با توجه به نتایج بدست آمده در روز هفتم و دهم روش FD و در روز چهاردهم و نوزدهم دو روش SD و MSC عملکرد بسیار خوبی داشته‌اند و نتایج با دقت

- Miller, S. A., Beed, F. D., & Harmon, C. L. (2009). *Plant disease diagnostic capabilities and networks*. Annual Review of Phytopathology, 47, 15–38. <https://doi.org/10.1146/annurev-phyto-080508-081743>.
- Miller, P. W., & Bollen, W. B. (1946). *Walnut bacteriosis and its control*. Technical Bulletin Oregon Aricultural Experiment Station, 9, 107 pp.
- Moffett, M. L. (1973). *Bacterial spot of stone fruit in Queensland*. Australian Journal of Biological Sciences, 26, 171- 179.
- Moghadam, P., Ward, D., Goan, E., Jayawardena, S., Sikka, P., & Hernandez, E. (2017). *Plant disease detection using hyperspectral imaging*. DICTA 2017 - 2017 International Conference on Digital Image Computing: Techniques and Applications, 2017-Decem, 1–8. <https://doi.org/10.1109/DICTA.2017.8227476>.
- Nicolai, Bart M, Beullens, Katrien, Bobelyn, Els, Peirs, Ann, Saeys, Wouter, Theron, Karen I, and Lammertyn, Jeroen. (2007). *Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review*. Postharvest Biology and Technology, 46(2): 99-118. <https://doi.org/10.1016/j.postharvbio.2007.06.024>.
- Razdan, M. K., & Mattoo, A. K. (2006). *Genetic improvement of Solanaceous crops: Tomato 2*, CRC Press.
- Rumpf, T., Mahlein, A. K., Steiner, U., Oerke, E. C., Dehne, H. W., & Plümer, L. (2010). *Early detection and classification of plant diseases with Support Vector Machines based on hyperspectral reflectance*. Computers and Electronics in Agriculture, 74(1), 91–99. <https://doi.org/10.1016/j.compag.2010.06.009>.
- Saffari, M. (2018). *Early detection of cucumber mosaic virus by visible-near-infrared spectroscopy method*, master's thesis, Isfahan University of Technology, Isfahan, Iran. (In Persian).
- Shahin, M. A., & Symons, S. J. (2011). *Detection of Fusarium damaged kernels in Canada Western Red Spring wheat using visible/near-infrared hyperspectral imaging and principal component analysis*. Computers and Electronics in Agriculture, 75(1), 107–112. <https://doi.org/10.1016/j.compag.2010.10.004>.
- Shirzadifar, A., Bajwa, S., Nowatzki, J., & Shojaeiarani, J. (2020). *Development of spectral indices for identifying glyphosate-resistant weeds*. Computers and Electronics in Agriculture, 170 (February), 105276. <https://doi.org/10.1016/j.compag.2020.105276>.
- Shirzadifar, A., Bajwa, S., Mireei, S. A., Howatt, K., & Nowatzki, J. (2018). *Weed species discrimination based on SIMCA analysis of plant canopy spectral data*. Biosystems Engineering, 171, 143–154. <https://doi.org/10.1016/j.biosystemseng.2018.04.019>.
- Siedliska, A., Baranowski, P., Zubik, M., Mazurek, W., & Sosnowska, B. (2018). *Detection of fungal infections in strawberry fruit by VNIR/SWIR hyperspectral imaging*. Postharvest Biology and Technology, 139(January), 115–126. <https://doi.org/10.1016/j.postharvbio.2018.01.018>.
- 32(1), 170–189. <https://doi.org/10.1016/j.biotechadv.2013.11.003>.
- Bock, C. H., Poole, G. H., Parker, P. E., & Gottwald, T. R. (2010). *Plant disease severity estimated visually, by digital photography and image analysis, and by hyperspectral imaging*. Critical reviews in plant sciences, 29(2), 59-107. <https://doi.org/10.1080/07352681003617285>.
- Castillejo-González, I. L., Pena-Barragán, J. M., Jurado-Expósito, M., Mesas-Carrascosa, F. J., & López-Granados, F. (2014). *Evaluation of pixel- and object-based approaches for mapping wild oat (Avena sterilis) weed patches in wheat fields using QuickBird imagery for site-specific management*. European Journal of Agronomy, 59, 57-66. <https://doi.org/10.1016/j.eja.2014.05.009>.
- Cordon, G., Andrade, C., Barbara, L., & Romero, A. M. (2022). *Early detection of tomato bacterial canker by reflectance indices*. Information Processing in Agriculture, 9(2), 184–194. <https://doi.org/10.1016/j.inpa.2021.06.004>.
- FAO. 2015. Statistical database. Available on Faostat.Fao.org
- Farber, C., Mahnke, M., Sanchez, L., & Kourouski, D. (2019). *Advanced spectroscopic techniques for plant disease diagnostics. A review*. TrAC Trends in Analytical Chemistry, 118, 43–49. <https://doi.org/10.1016/j.trac.2019.05.022>.
- Fu, Xiaping, Yibin Ying, Huishan Lu, Huirong Xu, and Haiyan Yu (2007). *FT-NIR diffuse reflectance spectroscopy for kiwifruit firmness detection*. Sensing and Instrumentation for Food Quality and Safety, 1, 29-35. <https://doi.org/10.1007/s11694-007-9004-2>.
- Gao, Z., Khot, L. R., Naidu, R. A., & Zhang, Q. (2020). *Early detection of grapevine leafroll disease in a red-berried wine grape cultivar using hyperspectral imaging*. Computers and Electronics in Agriculture, 179(April), 105807. <https://doi.org/10.1016/j.compag.2020.105807>.
- Gerszberg, A., Hnatuszko-Konka, K., Kowalczyk, T., & Kononowicz, A. K. (2015). *Tomato (Solanum lycopersicum L.) in the service of biotechnology*. Plant Cell, Tissue and Organ Culture, 120(3), 881–902. <https://doi.org/10.1007/s11240-014-0664-4>.
- Kazemi, F. (2021). *Feasibility of using visible and near-infrared spectroscopy to identify squash yellow mosaic virus infection before symptoms appear on leaves*, Master's Thesis, Faculty of Agriculture, Shiraz University, Shiraz, Iran. (In Persian).
- Madden, L.V., Hughes, G., & Bosch, F. (2007). *The Study of Plant Disease Epidemics*. American Phytopathological Society (APS Press), University of Minnesota.
- Mahlein, A.-K., Oerke, E.-C., Steiner, U., & Dehne, H.-W. (2012). *Recent advances in sensing plant diseases for precision crop protection*. European Journal of Plant Pathology, 133(1), 197–209. <https://doi.org/10.1007/s10658-011-9878-z>.

- Sivakumar, M. (2011). *Long wavelength near-infrared hyperspectral imaging for classification and quality assessment of bulk samples of wheat from different growing locations and crop years*. University of Manitoba (Canada).
- Xie, C., Shao, Y., Li, X., & He, Y. (2015). *Detection of early blight and late blight diseases on tomato leaves using hyperspectral imaging*. Scientific Reports, 5, 1–11. <https://doi.org/10.1038/srep16564>.
- Xu, H. R., Ying, Y. B., Fu, X. P., & Zhu, S. P. (2007). *Near-infrared Spectroscopy in detecting Leaf Miner Damage on Tomato Leaf*. Biosystems Engineering, 96(4), 447–454. <https://doi.org/10.1016/j.biosystemseng.2007.01.008>.
- Xuan, G., Li, Q., Shao, Y., & Shi, Y. (2022). *Early diagnosis and pathogenesis monitoring of wheat powdery mildew caused by blumeria graminis using hyperspectral imaging*. Computers and Electronics in Agriculture, 197(April), 106921. <https://doi.org/10.1016/j.compag.2022.106921>.
- Zhang, H., Paliwal, J., Jayas, D. S., & White, N. D. G. (2007). *Classification of fungal infected wheat kernels using near-infrared reflectance hyperspectral imaging and support vector machine*. Transactions of the ASABE, 50(5), 1779–1785.
- Zehr, E. I., & Sheard, D. P. (1996). *Bacterial spot of peach as influenced by water congestion, leaf wetness duration and temperature*. Plant Disease, 80, 339–340.