

تشخیص بیماری‌های درخت سیب با بهره‌گیری از روش‌های تحلیل بافت و یادگیری ماشین

سجاد نصیری^۱، مصطفی خجسته‌نژاد^{۲*}

۱. دانشکده فنی و مهندسی، دانشگاه بناب، بناب، ۵۵۵۱۷۶۱۱۶۷ (s.nasiri.cs@gmail.com)

۲. گروه مهندسی مکانیک، دانشکده فنی و مهندسی، دانشگاه بناب، بناب، ۵۵۵۱۷۶۱۱۶۷، ایران (khojasteh@ubonab.ac.ir)

چکیده

بیماری و آفت هر گیاهی بعنوان یک عامل اصلی و تاثیرگذار در کیفیت و کمیت محصول تولیدی است. روش‌های مرسوم تشخیص این آفتها براساس روش سنتی استفاده از فرد خبره است که مشکلات خاص خود را دارد. برای حل این مشکلات از روش خودکار و غیر مخرب بینایی ماشین استفاده می‌شود که باعث تشخیص سریع و کنترل بیماری خواهد شد و در پی آن کاهش هزینه‌های نظارت در سطح وسیع را در پی خواهد داشت. در این پژوهش، از یادگیری ماشین بر مبنای ویژگی‌های بافت هیستوگرام سطح خاکستری (GLH)، ماتریس هم-رخداد سطح خاکستری (GLCM)، ماتریس طول بردار سطح خاکستری (GLRM) و الگوی دودویی محلی (LBP) برای تشخیص بیماری‌های پوسیدگی سیاه، خال‌زدگی و لکه سیاه برگ درخت سیب استفاده شد. با قطعه‌بندی تصاویر بوسیله روش Kmeans، مدل‌های طبقه‌بند کننده درخت تصمیم (D-Tree)، K-نزدیک‌ترین همسایگی (KNN)، آنالیز تفکیک خطی (LDA)، ماشین بردار پشتیبان (SVM) و شبکه عصبی (ANN) بر اساس ویژگی‌های استخراج شده از بافت تصاویر نمونه‌ها در دو مرحله آموزش و آزمون مدل‌ها بکار گرفته شدند. مقایسه نتایج داده‌های آزمون نشان داد؛ مدل درخت تصمیم با دقت ۸۴/۴۲ درصد و مدل شبکه عصبی با دقتی برابر با ۹۳/۳۰ درصد بترتیب پایین‌ترین و بالاترین عملکرد را در بین سایر روش‌ها داشتند. K-نزدیک‌ترین همسایگی، ماشین بردار پشتیبان و آنالیز تفکیک خطی نیز به ترتیب ۸۷/۳۰، ۹۱/۳۳ و ۹۲/۶۲ درصد از تصاویر آزمون را بدرستی دسته‌بندی کردند. نتایج نشان داد، می‌توان از روش پردازش تصویر برای تشخیص سریع آفت‌های موجود در برگ درخت سیب استفاده نمود.

کلمات کلیدی:

پوسیدگی سیاه سیب، خال‌زدگی سیب، لکه سیاه سیب، بافت، پردازش تصویر، یادگیری ماشین

* نویسنده مسئول: مصطفی خجسته‌نژاد، khojasteh@ubonab.ac.ir

تشخیص بیماری‌های درخت سیب با بهره‌گیری از روش‌های تحلیل بافت و یادگیری ماشین

مقدمه

درختان و گیاهان، به دلیل شرایط آب و هوایی، شرایط کاشت و داشت محصول و همین‌طور عوامل غیر مترقبه، با بیماری‌های گوناگونی درگیر هستند که بعنوان یک عامل مهم در کیفیت محصول نهایی اثرگذار است. این بیماری‌ها، اگر به موقع و در زمان مناسبی شناسایی نشوند، تاثیر منفی بر روی کیفیت و کمیت محصول تولیدی خواهند گذاشت [۲۰]. پر واضح است که تشخیص سریع و دقیق بیماری می‌تواند مانع از گسترش آن شده و مدیریت و کنترل آن را برای کشاورز سهل و آسان نماید. بر این اساس در سال‌های گذشته، از روش سنتی برای تشخیص این بیماری‌ها استفاده شده است. در این روش عمل تشخیص توسط یک کارشناس آموزش دیده و خبره انجام می‌پذیرد که نظرات او، بدون خطا نیست. از جمله معایب روش سنتی می‌توان به زمانبر بودن آن علی‌الخصوص در مزارع و باغات بزرگ، هزینه‌بر بودن و احتمال خطا در تشخیص به دلیل خستگی فرد اشاره نمود. به همین دلیل، شناسایی سریع و دقیق بیماری‌ها با روشی خودکار و غیر مخرب، به عنوان یکی از نیازهای اساسی باغداری محسوب می‌شود [۱۷]. از جمله روش‌های غیر مخرب و سریع که کاربردهای فراوانی در بخش کشاورزی و مواد غذایی داشته است، روش بینایی ماشین با کمک گرفتن از روش‌های مختلف پردازش تصویر است.

در تحقیقات قبلی نیز به منظور بررسی بیماری گیاهان و درختان، از تصویر برگ آنها استفاده شده است [۳ و ۱۶]. چرا که برگ از جمله اندام‌های گیاهی است که با نفوذ بیماری در گیاه، تغییرات رنگی و بافتی در آن اتفاق می‌افتد [۴ و ۵]. در روش بینایی ماشین، با استخراج خصوصیات مختلفی از جمله خصوصیات رنگ؛ شکل، اندازه و بافت نمونه‌ها و تشکیل پایگاه داده، با روش‌های مختلف مدلسازی، دسته‌بندی بین نمونه‌ها انجام می‌پذیرد [۱۱]. براساس تحقیقات بی شماری که در این زمینه‌ها ارائه گردیده است، روش بینایی ماشین بعنوان روشی قابل اعتماد با دقت بالا معرفی شده است. بعد از استخراج ویژگی مورد نظر، بایستی دسته‌بندی و کلاس‌بندی نمونه‌ها انجام پذیرد. دسته‌بندی بیماری‌ها و تشخیص آنها توسط مدل‌های مختلفی می‌تواند انجام گیرد که براساس یافته‌های قبلی روش‌های شبکه عصبی مصنوعی (ANN)، ماشین بردار پشتیبان (SVM)، آنالیز تفکیک خطی (LDA)، K-نزدیکترین همسایگی (KNN) و درخت تصمیم (D-Tree) نتایج خوب و قابل قبولی را از خود نشان داده‌اند [۱۳، ۱۴، ۱۸، ۲۱ و ۲۳].

یکی از محصولات مهم کشاورزی که تولید و مصرف بالایی دارد و درگیر بیماری‌های مختلف نیز می‌باشد، درخت سیب است. از جمله بیماری‌های مرسوم برگ درخت سیب می‌توان به پوسیدگی سیاه سیب^۱، خال زدگی سیب (زنگ زدگی سیب)^۲ و لکه سیاه سیب^۳ اشاره نمود [۲۲]. تحقیقات مختلفی در رابطه با تشخیص بیماری برگ درخت سیب توسط روش‌های مختلف پردازش تصویر انجام پذیرفته است [۶، ۱۲ و ۲۲]. در تحقیقی که در کشور پاکستان بر روی بیماری‌های رایج انجام پذیرفته است، با جمع‌آوری تصاویر نمونه برگ‌های سالم و آفت زده و تشکیل پایگاه داده، اقدام به استخراج ویژگی‌هایی نموده‌اند. ویژگی‌های استخراج شده پس از یادگیری انتقال، با استفاده از رویکرد جدید کنترل شده واریانس، نمونه‌برداری شده‌اند. در نهایت، دقت متمایزترین ویژگی‌ها با استفاده از دسته‌بندی‌های مختلف ۹۷٪ گزارش گردیده است [۲۲].

براساس نتایج تحقیقات متنوعی که در زمینه بیماری‌های گیاهی انجام گرفته و نشان از توانمندی‌های روش بینایی ماشین و دسته‌بندکننده‌های مختلف در این زمینه دارند، بنابراین، در این تحقیق برای کاهش هزینه‌های نظارت و

¹ Apple black rot

² Apple rust

³ Apple scab

افزایش دقت شناسایی و دسته‌بندی بیماری‌های درخت سیب، روشی هوشمند مبتنی بر پردازش تصویر ارائه شد. برای اینکار، از استخراج ویژگی‌های بافت تصاویر برگ‌های آفت‌زده درخت سیب و الگوریتم‌های مختلف یادگیری ماشین استفاده و نتایج ارائه خواهد گردید.

روش پیشنهادی

در این تحقیق، روش هوشمندی جهت دسته‌بندی برگ‌های آفت‌زده سیب شامل سه بیماری، پوسیدگی سیاه، خال‌زدگی و لکه سیاه طراحی و ارائه شد. برای آزمایش تجربی روش پیشنهادی از پایگاه داده‌های موجود استفاده شد [۲]. این پایگاه داده در مجموع دارای ۴۸۰ تصویر، متشکل از ۱۷۰ تصویر برگ مبتلا به پوسیدگی سیاه (BR)، ۱۶۰ تصویر برگ مبتلا به خال‌زدگی (R) و ۱۵۰ تصویر برگ مبتلا به لکه سیاه (S) می‌باشد. به منظور ایجاد روشی پایدار و بهینه، در گام نخست، تصاویر نوپز زدایی و سپس با استفاده از خوشه‌بند Kmeans، برگ از پس‌زمینه جدا گردید. در گام بعدی، ویژگی‌های رنگ و بافت تصویر شامل ویژگی‌های استخراج شده از هیستوگرام سطح خاکستری^۱ (GLH)، ماتریس هم-رخدادی سطح خاکستری^۲ (GLCM)، ماتریس طول بردار سطح خاکستری^۳ (GLRM) و الگوی دودویی محلی^۴ (LBP) استخراج شد. بدین ترتیب برای هر تصویر، ۸۰ ویژگی استخراج و در بانک ویژگی ذخیره گردید. در نهایت به وسیله الگوریتم‌های شبکه عصبی، ماشین بردار پشتیبان، آنالیز تفکیک خطی^۵، k-نزدیک‌ترین همسایگی و درخت تصمیم^۶ ویژگی‌های استخراج شده مورد بررسی و مدل پیشنهادی ارائه گردید. شکل ۱ گام‌های روش پیشنهادی را نمایش می‌دهد.

پیش‌پردازش

به منظور بهبود نتایج روش پیشنهادی در تشخیص و دسته‌بندی بیماری سیب، اولین گام شامل دو فرایند، نوپز زدایی و استخراج برگ از پس‌زمینه می‌باشد. ابتدا به منظور پایداری روش نسبت به نوپز، با استفاده از فیلتر گوسی^۳* $\sigma = 3$ نوپز زدایی انجام شد. در ادامه برای استخراج ناحیه مورد نظر (برگ)، از قطعه‌بندی به وسیله خوشه‌بند Kmeans، با $k = 2$ و فاصله اقلیدسی استفاده گردید. هر دو فرایند بر روی کانال‌های رنگی تصویر انجام پذیرفت، به این معنی که از نتایج نوپز زدایی شده هر سه کانال رنگی R، G و B به عنوان ورودی الگوریتم Kmeans استفاده گردید. شکل ۲ نمونه‌ای از کلاس‌های مختلف تصویر موجود در پایگاه داده و نتایج حاصل از این مرحله را نمایش می‌دهد.

¹ Gray Level Histogram

² Gray Level Co-occurrence Matrix

³ Gray Level Run Length Matrix

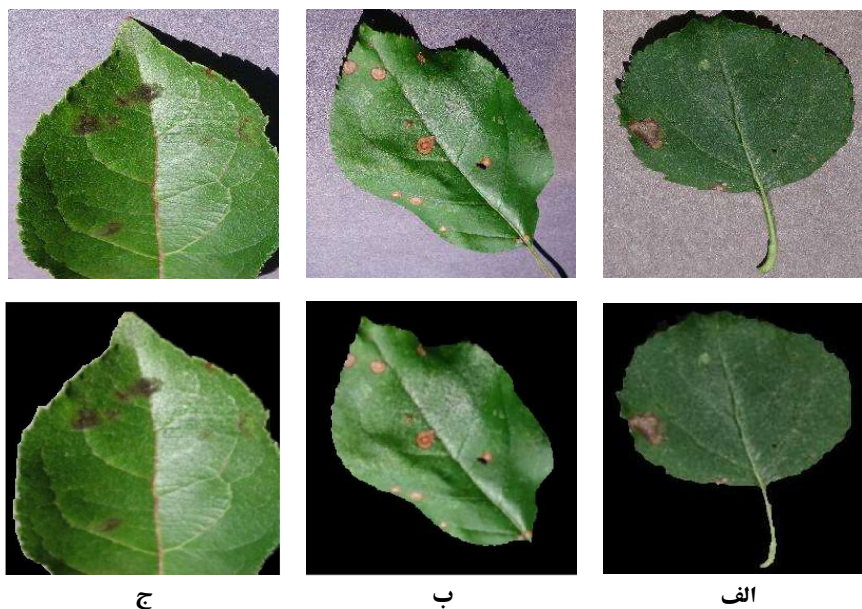
⁴ Local Binary pattern

⁵ Linear Discriminant Analysis

⁶ Decision Tree



شکل (۱) نمای کلی روش پیشنهادی



شکل (۲) نمونه‌ای از کلاس‌های مختلف تصاویر (ردیف بالا) و نتایج حاصل از پیش پردازش برای تصاویر متناظر (ردیف پایین).
الف: برگ مبتلا به پوسیدگی سیاه. ب: برگ مبتلا به خال زدگی. ج: برگ مبتلا به لکه سیاه.

استخراج ویژگی

ویژگی‌های ظاهری اصلی‌ترین پارامترهایی می‌باشند که اغلب افراد خبره با توجه به آن‌ها، درختان را از لحاظ ابتلا به بیماری مورد ارزیابی قرار می‌دهند. از این رو، برای ایجاد روش هوشمند و غیرمخرب از استخراج و آنالیز ویژگی‌های بافت و رنگ استفاده گردید. شباهت‌های موجود در پیکسل‌های همسایگی عامل اصلی ایجاد بافت تصاویر می‌باشند. انتخاب ویژگی‌هایی که اطلاعات اساسی تصویر را توصیف کند، چالش برانگیزترین موضوع در تحلیل بافت می‌باشد. ویژگی‌های تصاویر به چند روش مختلف، استخراج و در آموزش مدل‌ها استفاده گردید. ویژگی‌های مورد استفاده در ادامه شرح داده می‌شود.

هیستوگرام سطح خاکستری (GLH)

تحلیل بافت به وسیله مشخصات آماری هیستوگرام تصویر از ساده‌ترین و پرکاربردترین روش‌های تحلیل بافت می‌باشد [۷]. این روش نحوه قرارگیری پیکسل‌ها نسبت به همسایگی‌های خود را مورد تجزیه و تحلیل قرار نمی‌دهد، اما بصورت کلی توزیع بافت تصویر را مشخص می‌کند. در این تحقیق، برای هر یک از تصاویر شش ویژگی میانگین، انحراف معیار، همواری، گشتاور سوم، یکنواختی و انتروپی بر اساس احتمال وقوع شدت‌های رنگ سطح خاکستری محاسبه گردید، که به ترتیب مقادیر کمی، میانگین شدت بافت، میانگین کنتراست بافت، همواری نسبی شدت نواحی، چولگی هیستوگرام، میزان یکنواختی سطوح شدت و میزان تصادفی بودن را مشخص می‌کند.

ماتریس هم‌رخدادی سطح خاکستری (GLCM)

ماتریس هم‌رخدادی که برای تشخیص الگوهای تصویر کاربرد دارد، از توصیف‌کننده‌های آماری بافت می‌باشد. برای یک تصویر خاکستری، GLCM میزان احتمال مشترک وقوع شدت‌های i و j را در فواصل و زاوایی معین که قابل تنظیم هستند، محاسبه می‌کند [۸]. برای فاصله همسایگی با مقدار یک، فقط الگوهای موجود در پیکسل‌های مجاور نسبت به پیکسل مرکزی تشخیص داده می‌شود. در کار حاضر، ماتریس GLCM در چهار زاویه صفر، ۴۵، ۹۰ و ۱۳۵ درجه برای پیکسل‌های مجاور ایجاد شد و ویژگی‌های: انرژی، انتروپی، بیشترین احتمال، کنتراست، همبستگی و همگنی استخراج گردید. بنابراین برای هر تصویر بیست و چهار ویژگی از GLCM در بردار ویژگی‌های تصویر ذخیره می‌گردد.

ماتریس طول بردار سطح خاکستری (GLRM)

یکی دیگر از روش‌های تحلیل آماری بافت، ماتریس GLRM می‌باشد [۱۱]. به هر یک از مولفه‌های این ماتریس طول اجرا اطلاق می‌شود. هر کدام از این مولفه‌ها مجموع تعداد پیکسل‌های همجواری هستند که در زاویه‌ای مشخص، شدت رنگ برابری دارند. در این پژوهش، ۱۱ ویژگی: short run, long run, gray level non-uniformity, run length non-uniformity, run ratio, low gray level run, high gray level run, short run low gray level, short run high gray level, long run low gray level and Long run high gray level برای هر تصویر از GLRM استخراج گردید، که با احتساب چهار زاویه صفر، ۴۵، ۹۰ و ۱۳۵ درجه مورد بررسی، برای هر تصویر در مجموع ۴۴ ویژگی از این روش در بردار ویژگی‌های تصویر ذخیره گردید.

الگوی دودویی محلی (LBP)

LBP ساختار محلی همسایگی یک پیکسل را توصیف می‌کند [۱۵]. به این معنا که مقادیر پیکسل‌های همسایگی در مقایسه با پیکسل مورد بررسی به حالت دودویی تغییر می‌یابند. به عنوان مثال پیکسل‌های با شدت کم‌تر یا مساوی از پیکسل مرکزی به صفر و مابقی به یک تبدیل می‌شوند. سپس کد دودویی حاصل از این مرحله به عدد صحیح تبدیل شده و به عنوان مقدار LBP پیکسل مرکزی استفاده می‌شود. برای استخراج ماتریس LBP و بدست آوردن مقادیر یکتا در الگوهای یکسان، پس از ایجاد کد دودویی، آن را با یک حرکت دورانی در راستای پادساعتگرد جابجا کرده و در نهایت بیشترین مقدار محاسبه شده در چرخش‌های مختلف به عنوان LBP پیکسل مورد نظر انتخاب می‌شود [۱۰]. شش ویژگی میانگین، انحراف معیار، همواری، گشتاور سوم، یکنواختی و انتروپی از این ماتریس برای هر تصویر محاسبه و در بردار ویژگی‌های تصویر ذخیره گردید.

روش‌های مختلفی برای دسته‌بندی وجود دارد، اما در صورت وجود برچسب داده‌ها، یادگیری نظارت‌شده نتایج بهتری ارائه خواهد کرد. این رویکرد داده‌های ورودی را در دو مرحله به کلاس‌هایی خاص نگاشت می‌کند. ابتدا در مرحله آموزش، مدل مجموعه داده‌هایی را که از قبل برچسب‌گذاری شده‌اند را تحت یک الگوریتم آموزش فرا می‌گیرد. در مرحله بعدی مدل آموزش یافته در مرحله قبل، برای دسته‌بندی داده‌های جدید بکار گرفته می‌شود. برای بررسی بهتر ویژگی‌های استخراج شده و همچنین ارائه‌ی مدلی کارا تر در دسته‌بندی و تشخیص بیماری‌های سیب، الگوریتم‌های دسته‌بندی نظارت‌شده مختلفی مورد استفاده قرار گرفت.

درخت تصمیم (D-Tree)

مزیت این روش نسبت به سایر روش‌ها، قابل تفسیر بودن نتایج آن برای انسان می‌باشد. بطوری که حتی برای افراد غیر متخصص نیز قابل درک می‌باشد. این روش، داده‌ها را به گروه‌های کوچک‌تر با ساختار درختی دسته‌بندی می‌کند. بطور معمول از معیارهای جینی یا آنتروپی برای اندازه‌گیری عدم شباهت داده‌ها استفاده می‌گردد. اگر محدودیتی از لحاظ عمق درخت در نظر گرفته نشود، تولدایی تولید میزان خطای بسیاری پایینی برای داده‌های آموزش را دارد. الگوریتم‌های مختلفی از جمله ID3، C4.5 و CART برای آموزش این روش وجود دارند [۱۸].

K-نزدیک‌ترین همسایگی (KNN)

یک روش یادگیری ماشین غیر پارامتری است که از قانون نزدیک‌ترین همسایگی برای دسته‌بندی داده‌ها استفاده می‌کند. در این روش، دسته‌بندی بر اساس شباهت بین الگوهای داده‌های آزمون با الگوهای داده‌های آموزش استفاده می‌گردد. در واقع هر داده جدید بر اساس برچسب نزدیک‌ترین همسایگان خود که بیشترین فراوانی را در محدوده تعیین شده دارند، دسته‌بندی می‌شود [۱ و ۱۹].

ماشین بردار پشتیبان (SVM)

این روش که مبتنی بر ابرصفحه جداساز است؛ در ساختار پایه، یک ابر صفحه بهینه که حاشیه بین داده‌های دو کلاس را به بیشترین مقدار می‌رساند جست‌وجو می‌کند. برای دسته‌بندی داده‌های چند کلاسی نیز مجموعه‌ای از ابر صفحه‌ها ایجاد می‌شوند. برای این کار از دو روش استفاده می‌شود. در روش اول از آموزش یک SVM چند کلاسی و در روش دیگر از چند SVM دو کلاسی استفاده می‌گردد. در این تحقیق از هسته چند جمله‌ای درجه ۳ استفاده گردید [۹].

آنالیز تفکیک خطی (LDA)

آنالیز تفکیک خطی برای دسته‌بندی صحیح داده‌ها، راستای موثر را جست‌جو می‌کند. در راستای بهینه، میانگین داده‌های کلاس‌های مختلف بیشترین فاصله ممکن و واریانس داده‌های داخل هر کلاس کمترین مقدار را خواهد داشت. یعنی بیشترین میزان تفکیک در این راستا قابل دستیابی است [۲۳]. برای دسته‌بندی چند کلاسی نیز می‌توان از دو روش یکی در مقابل بقیه یا کلاس‌بندی جفت کلاس‌ها استفاده و در نهایت با ترکیب دسته‌بندها به یک دسته‌بند چند کلاسی دست یافت.

شبکه عصبی (ANN)

شبکه عصبی که برگرفته از شبکه‌های عصبی زیستی می‌باشد، که با توجه به عملکردی که در پژوهش‌های مختلف داشته است از معروف‌ترین روش‌های یادگیری ماشین می‌باشد. در حالت کلی شبکه عصبی شامل چند لایه از نودهایی می‌باشد که به عنوان نورون شناخته می‌شوند. هر یک از نورون‌ها دارای یک مقدار وزن هستند که در فرایند یادگیری، مقادیر این وزن‌ها تنظیم می‌شود. انواع مختلفی از شبکه‌های عصبی برای کاربردهای مختلفی ارائه شده‌اند که از آن جمله می‌توان به شبکه عصبی پیش‌خور اشاره کرد. این شبکه در عین سادگی ساختاری، در کارهای دسته‌بندی از جمله کارترین روش‌ها می‌باشد [۱۰].

ارزیابی مدل‌ها

مدل‌ها به وسیله پارامترهای آماری حساسیت^۱، ویژگی^۲ و دقت^۳ ارزیابی شد. معادلات (۱)، (۲) و (۳) به ترتیب نحوه محاسبه این پارامترها را به نمایش می‌گذارند که TP ، TN ، FP و FN به ترتیب نشان دهنده مقادیر؛ نرخ مثبت صحیح، نرخ منفی صحیح، نرخ مثبت کاذب و نرخ منفی کاذب می‌باشند. مقدار کارایی مدل به وسیله کمیت حساسیت اندازه‌گیری می‌شود که در واقع توانایی شناسایی صحیح کلاس‌ها را نمایش می‌دهد. در طرف دیگر ویژگی، نرخ پاسخ‌های منفی صحیح را بیان می‌کند، یعنی مقدار پیش‌بینی عدم تعلق به یک کلاس را می‌سنجد. آخرین معیار مورد ارزیابی برای مدل‌های پیشنهادی یعنی دقت، میزان کل پیش‌بینی‌های دقیق را نمایش می‌دهد [۱۱].

$$Sensitivity = \frac{TP}{TP + FN} \quad (1)$$

$$Specificity = \frac{TN}{TN + FP} \quad (2)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

تصاویر استفاده شده، بصورت تصادفی و نسبت ۷۰ به ۳۰ به ترتیب برای مراحل آموزش و آزمون اختصاص داده شد. بنابراین ۳۳۶ تصویر برای آموزش و ۱۴۴ تصویر دیگر نیز برای آزمون مدل‌ها استفاده گردید. از ۱۴۴ تصویر آزمون، ۵۵ تصویر مربوط به پوشیدگی سیاه (BR)، ۴۵ تصویر خال زدگی (R) و ۴۴ لکه سیاه (S) می‌باشد. تمام مراحل پردازش تصویر، استخراج ویژگی و مدل‌سازی‌ها در محیط برنامه متلب (Matlab) نسخه شماره ۹ انجام و نتایج استخراج گردید.

نتایج و بحث

در پژوهش حاضر دسته‌بندی تصاویر برگ درخت سیب مبتلا به بیماری بر مبنای ویژگی‌های استخراج شده توسط الگوریتم‌های درخت تصمیم (D-Tree) با استفاده از الگوریتم CART و معیار جینی، k-نزدیک‌ترین همسایگی (KNN) با

¹ Sensitivity

² Specificity

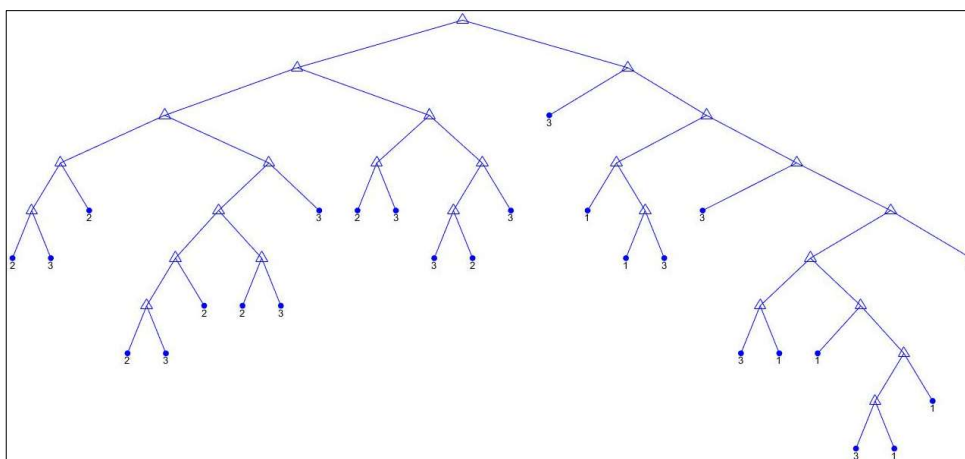
³ Accuracy

$k = 6$ ، آنالیز تفکیک خطی (LDA)، ماشین بردار پشتیبان (SVM) چند کلاسی با هسته چند جمله‌ای درجه ۳ و شبکه عصبی (ANN) پیشخور با یک لایه مخفی شامل ۵۰ نورون انجام گردید. بردار ویژگی هر تصویر که دارای ۸۰ ویژگی شامل ویژگی‌های استخراج شده از الگوریتم‌های GLH، GLCM، GLRM و LBP بود، در مدلسازی‌ها مورد استفاده قرار گرفت. ماتریس اغتشاش و میزان شاخص‌های ارزیابی شده برای فاز آموزش الگوریتم‌های مورد استفاده در جدول ۱ نمایش داده شده است. تمام روش‌ها دقت مناسب و قابل اطمینانی در تشخیص و دسته‌بندی کلاس‌های مختلف داشته‌اند، که عملکرد مناسب تحلیل بافت و ویژگی‌های استخراج شده را نمایش می‌دهد. با توجه به عدم اعمال محدودیت ارتفاع، همان‌گونه که پیش‌بینی می‌شد درخت تصمیم، دقت بسیار بالایی در یادگیری الگوهای داده‌های آموزش داشته است. گراف شکل ۳، رسم درخت تصمیم بدست آمده است که ۹۸/۴۵ درصد از برچسب داده‌های آموزش را حداکثر در ارتفاع ۹، بصورت صحیح فراگرفته است. در طرف مقابل شبکه عصبی با دقت ۹۴/۴۰ درصد، کم‌ترین عملکرد را در بین سایر الگوریتم‌ها بر روی تصاویر آموزش داشت. اما آنچه در روش‌های یادگیری ماشین اهمیت بیشتری دارد، تعمیم‌پذیری مدل ایجاد شده می‌باشد. به عبارت دیگر مدل باید قدرت تفکیک داده‌های جدید (داده‌های آزمون) را به همان میزان داده‌های آموزش داشته باشد.

نتایج حاصل از اعمال مدل‌های آموزش دیده بر روی تصاویر آزمون در جدول ۲ مشخص شده است. بر اساس نتایج فاز آزمون، ANN با دقتی برابر با ۹۳/۳۰ درصد، بیشترین و D-Tree با دقتی معادل با ۸۴/۴۲ کم‌ترین میزان دقت دسته‌بندی برای داده‌های آزمون را نتیجه داده است. K-نزدیک‌ترین همسایگی، ماشین بردار پشتیبان و آنالیز تفکیک خطی نیز به ترتیب با دقت ۸۷/۳۰، ۹۱/۳۳ و ۹۲/۶۲ درصدی در تشخیص و دسته‌بندی بیماری‌های سیب (پوسیدگی سیاه، خال‌زدگی و لکه سیاه) قابل قبول و مناسب ارزیابی می‌شود، در این بین نتایج حاصل

جدول ۱) نتایج حاصل از آموزش الگوریتم‌های مورد استفاده و ماتریس اغتشاش مربوطه

	ANN			LDA			SVM			KNN			D-Tree		
	BR	R	S	BR	R	S	BR	R	S	BR	R	S	BR	R	S
BR	۱۱۲	۰	۳	۱۱۵	۰	۰	۱۱۵	۰	۰	۱۱۳	۱	۱	۱۱۳	۱	۱
R	۰	۱۰۲	۸	۰	۱۱۵	۰	۰	۱۱۴	۱	۰	۱۱۴	۱	۰	۱۱۴	۱
S	۷	۷	۹۲	۷	۵	۹۴	۴	۴	۹۸	۲۶	۲۹	۵۱	۳	۱	۱۰۲
شاخص	Sen	Spe	Acc	Sen	Spe	Acc	Sen	Spe	Acc	Sen	Spe	Acc	Sen	Spe	Acc
BR	۹/۱۲	۹/۶۲	۹/۴۰	۹/۳۰	۱۰۰	۹/۱۳	۹/۶۴	۱۰۰	۹/۳۲	۸/۳۰	۹/۰۰	۹/۱۴	۹/۴۱	۹/۱۰	۹/۲۵
	۴	۸	۶	۴	۱۰۰	۷	۶	۱۰۰	۸	۱	۹	۰	۷	۹	۸
R	۹/۹۰	۹/۴۰	۹/۱۳	۹/۸۳	۱۰۰	۹/۹۲	۹/۶۱	۹/۵۴	۹/۱۰	۹/۲۰	۹/۵۰	۸/۳۲	۹/۳۰	۹/۵۵	۹/۹۱
	۳	۶	۵	۵	۱۰۰	۷	۶	۹	۸	۶	۹	۹	۸	۹	۸
S	۸/۳۲	۹/۰۰	۹/۷۰	۱۰۰	۹/۰۴	۹/۵۲	۹/۰۰	۹/۶۲	۹/۸۱	۹/۲۳	۸/۶۰	۸/۴۰	۹/۱۰	۹/۳۰	۹/۲۰
	۹	۴	۱	۱۰۰	۵	۷	۹	۶	۷	۶	۰	۸	۸	۸	۸
میانگین	۹/۴۳	۹/۳۴	۹/۴۰	۹/۷۰	۹/۳۵	۹/۵۲	۹/۴۱	۹/۷۲	۹/۱۰	۸/۶۰	۹/۰۱	۸/۳۰	۹/۹۲	۹/۰۰	۹/۴۵
بین	۲	۶	۴	۶	۸	۷	۷	۸	۸	۵	۳	۹	۷	۹	۸



شکل ۳) درخت تصمیم، اعداد بر روی برگ‌ها مشخص کننده برجسب‌ها می‌باشند؛ (۱) برگ مبتلا به پوسیدگی سیاه، (۲) برگ مبتلا به خال زدگی، (۳) برگ مبتلا به لکه سیاه.

از آنالیز تفکیک خطی با توجه به اختلاف بسیار پایین با شبکه عصبی قابل قیاس می‌باشد. با مقایسه اختلاف دقت نتایج آزمون (جدول ۲) و آموزش (جدول ۱)، مشخص می‌شود شبکه عصبی و K-نزدیک‌ترین همسایگی بیشترین میزان تعمیم‌پذیری را داشته‌اند. در طرف مقابل اما درخت تصمیم بر روی داده‌های آزمون نتایج خوبی نداشته است.

در تحقیق [۲۲] که بر روی بیماری‌های سیب انجام پذیرفته است، برای دسته‌بندی تصاویر در چهار کلاس متشکل از برگ سالم و برگ‌های مبتلا به ۳ نوع بیماری، دقت ۹۷ درصدی را برای آنالیز تفکیک خطی گزارش شده است. آن‌ها البته نتایج ارائه شده را به تفکیک برای داده‌های آموزش و آزمون ارائه نکرده‌اند و دقت اعلامی برای کل داده‌ها می‌باشد. اما بر اساس نتایج بدست آمده در این پژوهش (جدول ۱ و ۲)، مشخص است هر چند آنالیز تفکیک خطی بطور کلی بر روی کل داده‌ها نتیجه مناسبی ارائه کرده است، بدلیل عدم تعمیم‌پذیری مناسب، روشی قابل اعتماد ارائه نمی‌دهد.

با بررسی دقیق‌تر جداول ۱ و ۲ مشخص می‌شود تمام دسته‌بندها برای دو کلاس پوسیدگی سیاه (BR) و خال زدگی (R) نتایج بهتری در مقایسه با کلاس لکه سیاه (S) داشته‌اند. عدم دقت مناسب در تشخیص صحیح این کلاس می‌تواند به علت شباهت آن به سایر نمونه‌ها باشد. چنانچه از ماتریس‌های اغتشاش نیز مشخص است نرخ مثبت صحیح دو کلاس دیگر بسیار بالا است، اما تصاویر مربوط به برگ‌های مبتلا به لکه سیاه بر روی دو کلاس دیگر پخش شده‌اند.

جدول ۲) نتایج دسته‌بندی داده‌های آزمون توسط الگوریتم‌های مورد استفاده و ماتریس اغتشاش مربوطه

	ANN			LDA			SVM			KNN			D-Tree		
	BR	R	S	BR	R	S	BR	R	S	BR	R	S	BR	R	S
BR	۵۴	۰	۱	۵۲	۰	۳	۵۳	۱	۱	۵۴	۱	۰	۴۸	۰	۷
R	۰	۴۰	۵	۱	۴۱	۳	۰	۴۱	۴	۰	۴۳	۲	۰	۳۷	۸
S	۵	۲	۳۷	۳	۴	۳۷	۸	۳	۳۳	۱۲	۱۳	۱۹	۱۱	۴	۲۹
شاخص	Sen	Spe	Acc	Sen	Spe	Acc	Sen	Spe	Acc	Sen	Spe	Acc	Sen	Spe	Acc
BR	۹/۵۳	۹/۸۲	۹/۲۰	۹/۹۰	۹/۶۰	۹/۷۲	۸/۹۰	۹/۶۰	۹/۲۴	۸/۸۲	۹/۷۲	۹/۳۰	۸/۴۰	۹/۸۰	۸/۶۰
	۱	۸	۵	۲	۷	۴	۶	۷	۲	۱	۸	۰	۱	۱	۶
R	۹/۲۴	۹/۱۰	۹/۲۰	۹/۱۱	۹/۰۰	۹/۵۴	۹/۱۱	۹/۰۰	۹/۵۴	۷/۴۴	۹/۲۰	۸/۶۰	۹/۲۴	۹/۲۳	۹/۲۴
	۵	۵	۵	۱	۶	۳	۱	۶	۳	۵	۷	۶	۰	۲	۱

S	۸/۰۵	۹/۱۰	۸/۶۰	۸/۰۵	۹/۱۰	۸/۶۰	۸/۸۴	۸/۶۲	۸/۲۳	۹/۵۰	۷/۷۰	۸/۱۰	۶/۹۱	۸/۰۰	۷/۴۵
	۶	۳	۹	۶	۳	۹	۶	۹	۸	۰	۹	۵	۵	۵	۵
میاتک	۹/۹۴	۹/۷۰	۹/۳۰	۹/۰۲	۹/۶۰	۹/۶۲	۸/۲۸	۹/۴۰	۹/۳۳	۸/۶۰	۹/۰۴	۸/۳۰	۷/۱۸	۸/۷۰	۸/۴۲
ین	۰	۵	۳	۰	۵	۲	۸	۴	۱	۲	۲	۷	۹	۹	۴

ماشین بردار پشتیبان و آنالیز تفکیک خطی هر چند بر روی داده‌های آزمون عملکرد مناسبی از خود نشان می‌دهند، اما با مقایسه نتایج مرحله آموزش و آزمون آن‌ها مشخص است تعمیم‌پذیری مناسبی ندارند، از طرفی K- نزدیک‌ترین همسایگی علی‌رغم تعمیم‌پذیری خوبی که داشته است، در مقایسه با شبکه عصبی دقت پایین‌تری ارائه می‌دهد. بنابراین شبکه عصبی با استفاده از ویژگی‌های بافت استخراج شده در این پژوهش، مدل بهتری برای تشخیص بیماری درخت سیب ارائه می‌دهد.

نتیجه‌گیری

در این تحقیق تصاویر برگ درخت سیب که دارای سه نوع بیماری پوسیدگی سیاه، خال‌زدگی و لکه سیاه می‌باشد برای آزمایش تجربی روش پیشنهادی استفاده شد. تصاویر ابتدا نوپزداپی شده و سپس به منظور دستیابی به دقت بالاتر بخش حاوی برگ به وسیله الگوریتم Kmeans با استفاده از تمام کانال‌های رنگی از پس‌زمینه جدا شد. روش پیشنهادی بر پایه ویژگی‌های بافت طرح‌ریزی گردید. بدین منظور ویژگی‌های بافتی تصاویر از الگوریتم‌های GLH، GLCM، GLRM و LBP استخراج گردید. در ادامه برای دسته‌بندی تصاویر بر اساس بردار ویژگی هر تصویر متشکل از ۸۰ ویژگی، الگوریتم‌های درخت تصمیم، K-نزدیک‌ترین همسایگی، ماشین بردار پشتیبان، آنالیز تفکیک خطی و شبکه عصبی بکار گرفته شدند. از ۴۸۰ تصویر پایگاه داده، ۳۳۶ تصویر برای فاز آموزش و ۱۴۴ تصویر برای فاز آزمون بصورت تصادفی انتخاب شدند.

نتایج مدل‌های ایجاد شده بر روی تصاویر آزمون، نشان داد ویژگی‌های بافت در دسته‌بندی بیماری درخت سیب کارایی مناسبی دارند. در بین مدل‌های ارائه شده، درخت تصمیم علی‌رغم دقت بسیار بالا در دسته‌بندی تصاویر آموزشی و عمق مناسب درخت حاصل، اما با دقت ۸۴/۴۲ کم‌ترین میزان تشخیص صحیح را از خود نشان داد. شبکه عصبی و آنالیز تفکیک خطی به ترتیب با ۹۳/۳۰ و ۹۲/۶۲ درصد بالاترین دقت را ارائه کردند. همچنین با بررسی داده‌های حاصل از بخش آموزش مدل‌ها مشخص گردید؛ شبکه عصبی بیشترین تعمیم‌پذیری داشته است. بنابراین با توجه به تمام نتایج حاصل؛ روش پیشنهادی برای دسته‌بندی بیماری‌های درخت سیب که در برگ این گیاه ظهور می‌یابند، مناسب ارزیابی می‌شود. این روش در صورت استفاده توسط کشاورزان و باغ‌داران می‌تواند با تشخیص صحیح و به موقع این بیماری‌ها، هزینه‌ها و تلفات بیماری را تا میزان قابل ملاحظه‌ای کاهش دهد.

منابع

1. Agrawal, D., and Dahiya, P.P. 2018. COMPARISONS OF CLASSIFICATION ALGORITHMS ON SEEDS, 7(V): 2760–2765.
2. Apple Leaf Diseases | Kaggle. (n.d.). Retrieved April 12, 2021, from <https://www.kaggle.com/mhantor/apple-leaf-diseases>

3. Argüeso, D., Picon, A., Irusta, U., Medela, A., San-Emeterio, M.G., Bereciartua, A., and Alvarez-Gila, A. 2020. Few-Shot Learning approach for plant disease classification using images taken in the field. *Computers and Electronics in Agriculture*, 175: 105542.
4. Barbedo, J.G.A. 2016. A review on the main challenges in automatic plant disease identification based on visible range images. *Biosystems Engineering*, 144: 52–60.
5. Barbedo, J.G.A. 2019. Plant disease identification from individual lesions and spots using deep learning. *Biosystems Engineering*, 180: 96–107.
6. Gargade, A., and Khandekar, S.A. 2019. A review: custard apple leaf parameter analysis and leaf disease detection using digital image processing. 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), 267–271.
7. Gonzalez, R.C.E., Woods, S.L., Gonzalez, R.E.R.E.R.C., Woods, R.E., and Eddins, S.L. 2004. Digital image processing using MATLAB (Issues 04; TA1637, G6.).
8. Haralick, R.M., Shanmugam, K., and Dinstein, I.H. 1973. Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, 6: 610–621.
9. Huang, Z., Qin, A., Lu, J., Menon, A., and Gao, J. 2020. Maize Leaf Disease Detection and Classification Using Machine Learning Algorithms. 2020 International Conferences on Internet of Things (IThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData) and IEEE Congress on Cybermatics (Cybermatics), 870–877.
10. Karimi, N., Ranjbarzadeh Kondrood, R., and Alizadeh, T. 2017. An intelligent system for quality measurement of Golden Bleached raisins using two comparative machine learning algorithms. *Measurement: Journal of the International Measurement Confederation*, 107: 68–76. <https://doi.org/10.1016/j.measurement.2017.05.009>
11. Khojastehnazhand, M., and Ramezani, H. 2020. Machine vision system for classification of bulk raisins using texture features. *Journal of Food Engineering*, 271(September 2019), 109864. <https://doi.org/10.1016/j.jfoodeng.2019.109864>.
12. Kumar, S.S., and Raghavendra, B.K. 2019. Diseases detection of various plant leaf using image processing techniques: A review. 5th International Conference on Advanced Computing & Communication Systems (ICACCS), 313–316.
13. Kurmi, Y., and Gangwar, S. 2021. A leaf image localization based algorithm for different crops disease classification. *Information Processing in Agriculture*.
14. Ngugi, L.C., Abelwahab, M., and Abo-Zahhad, M. 2021. Recent advances in image processing techniques for automated leaf pest and disease recognition—A review. *Information Processing in Agriculture*, 8(1): 27–51.
15. Ojala, T., Pietikainen, M., and Maenpaa, T. 2002. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7): 971–987.
16. Pandiyan, S., Ashwin, M., Manikandan, R., KM, K.R., & GR, A.R. 2020. Heterogeneous Internet of Things organization predictive analysis platform for apple leaf diseases recognition. *Computer Communications*, 154: 99–110.

17. Pooja, V., Das, R., and Kanchana, V. 2017. Identification of plant leaf diseases using image processing techniques. 2017 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR), 130–133.
18. Rajesh, B., Vardhan, M.V.S., and Sujihelen, L. 2020. Leaf Disease Detection and Classification by Decision Tree. 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184), 705–708.
19. Salem, S.A., Salem, N.M., and Nandi, A.K. 2007. Segmentation of retinal blood vessels using a novel clustering algorithm (RACAL) with a partial supervision strategy. Medical and Biological Engineering and Computing, 45(3): 261–273. <https://doi.org/10.1007/s11517-006-0141-2>
20. Sethy, P.K., Barpanda, N.K., Rath, A.K., and Behera, S.K. 2020. Deep feature based rice leaf disease identification using support vector machine. Computers and Electronics in Agriculture, 175: 105527.
21. Singh, J., and Kaur, H. 2018. A review on: Various techniques of plant leaf disease detection. Proceedings of the 2nd International Conference on Inventive Systems and Control, ICISC, Icisc, 232–238. <https://doi.org/10.1109/ICISC.2018.8399069>
22. Tahir, M. Bin, Khan, M.A., Javed, K., Kadry, S., Zhang, Y.-D., Akram, T., and Nazir, M. 2021. Recognition of apple leaf diseases using deep learning and variances-controlled features reduction. Microprocessors and Microsystems, 104027.
23. Zapolska, A., Kalaitzidis, C., Markakis, E., Ligoxigakis, E., and Koubouris, G. 2020. Linear Discriminant Analysis of spectral measurements for discrimination between healthy and diseased trees of *Olea europaea* L. artificially infected by *Fomitiporia mediterranea*. International Journal of Remote Sensing, 41(14): 5388–5398.

Apple Tree Diseases Detection Using Texture Analysis and Machine Learning Methods

Sajjad Nasiri¹, Mostafa Khojastehnazhand^{2*}

1. Faculty of Engineering, University of Bonab, Bonab, 5551761167, Iran
2. Mechanical Engineering Department, University of Bonab, Bonab, 5551761167, Iran

Abstract

Disease and pest of each plant is a major and influential factor in the quality and quantity of the product. Conventional methods of diagnosing these pests are based on the traditional method of using an expert who has his own problems. To solve these problems, the automated and non-destructive method of machine vision is used, which will cause rapid detection and control of the disease, and consequently reduce the costs of large-scale monitoring. In this study, machine learning based on texture features of Gray Level Histogram (GLH), Gray Level Co-occurrence Matrix (GLCM), Gray Level Run length Matrix (GLRM) and Local Binary Pattern (LBP) algorithms were used to detect apple leaves diseases such as black rot, rust and scab. By segmentation of images using kmeans method, Decision Tree (D-Tree), K-Nearest Neighborhood (KNN), Linear Discrimination Analysis (LDA), Support Vector Machine (SVM) and Artificial Neural Network (ANN) models were used in two stages of training and testing modes. The test data results shows that the decision tree model with 84.42% accuracy and the ANN model with accuracy of 93.30 had the lowest and highest performance among other methods, respectively. KNN, SVM and LDA models also correctly classified 87.30, 91.33 and 92.62% of the test images, respectively. The results showed that image processing method could be used to quickly detect pests in apple tree leaves.

Key words: Apple black rot, Apple rust, Apple scab, Texture, Image processing, Machine learning

*Corresponding author

E-mail: E-mail: khojasteh@ubonab.ac.ir