

تعیین ویژگی‌ها و مدل بهینه به منظور طبقه بندی اجزای کیفی گندم بذری با استفاده از پردازش تصاویر دوربین دیجیتال صنعتی

ضرغام فاضل نیاری^۱، امیرحسین افکاری سیاح^{۲*} و یوسف عباسپور گیلانده^۳

۱، ۲ - به ترتیب: دانشجوی دکتری تخصصی؛ دانشیار؛ و استاد گروه مهندسی بیوسیستم، دانشکده علوم کشاورزی و منابع طبیعی، دانشگاه محقق اردبیلی، اردبیل، ایران
تاریخ دریافت: ۱۳۹۵/۱۱/۴؛ تاریخ پذیرش: ۱۳۹۶/۷/۱۱

چکیده

دستیابی به دانش پایه بینایی ماشینی در کنترل کیفی بذر گندم اهمیت ویژه‌ای دارد. در این پروژه، با تهیه سخت‌افزار و نرم‌افزار مناسب، تصویر ۲۱۰۰۰ عدد نمونه دانه، تهیه و با الگوریتم طراحی شده پردازش شد. نود و یک ویژگی شکل، بافت و رنگ محاسبه و رتبه‌بندی شد. از میان پنج مدل طبقه‌بندی کننده، بیشترین صحت کل، طبقه‌بندی در مدل شبکه عصبی مصنوعی با دو لایه پنهان و به کارگیری ۳۵ ویژگی برتر اول به دست آمد. در آزمون این مدل با استفاده از داده‌های مستقل، صحت طبقه‌بندی برای گندم سفید درشت، گندم سفید کوچک، گندم سفید شکسته، گندم سفید چروکیده، گندم قرمز، جو و چاودار به ترتیب ۱۰۰، ۹۶/۷، ۹۹/۳، ۹۰/۳، ۹۹، ۹۹/۷، ۹۸ درصد و میانگین آنها ۹۷/۶ درصد محاسبه شده است. صحت کل طبقه‌بندی در مدل‌های تحلیل تفکیک خطی، تحلیل تفکیک درجه دوم، K- نزدیک‌ترین همسایگی و شبکه عصبی مصنوعی با یک لایه پنهان به ترتیب ۹۵، ۹۶/۷، ۹۱/۶ و ۹۷/۳ درصد به دست آمده است. در شرایط این تحقیق، سیستم بینایی ماشینی شامل دوربین دیجیتال صنعتی و طبقه‌بندی کننده مدل شبکه عصبی مصنوعی دو لایه، با صحت خوبی قابل استفاده در بررسی کیفیت ظاهری بذر گندم تشخیص داده شده است.

واژه‌های کلیدی

داده‌کاو، طبقه‌بندی دانه، کاهش ویژگی، کنترل کیفیت، گواهی بذر

مقدمه

دانه‌های خارجی، ارزیابی دقیق بذر گندم در زمان تولید ضروری است. در حال حاضر تصمیمات درجه‌بندی گندم به بررسی ظاهری نمونه محصول توسط افراد آموزش دیده محدود است. غیر ماشینی بودن این نوع بررسی‌ها باعث می‌شود تا نتایج ارزیابی تحت تأثیر تجربیات شخصی بررسی کننده قرار گیرد (Venora *et al.*, 2009). از این‌رو، روش‌های کمی‌شده اندازه‌گیری ویژگی‌های غلات بسیار سودمند و مورد نیاز خواهد بود.

برای تولید محصول سالم و جلوگیری از کاهش عملکرد، به بذر سالم و با کیفیت بالا نیاز خواهد بود. با توجه به سطح زیر کشت سالانه گندم در ایران (۵/۶ میلیون هکتار)، هر سال به حدود ۲/۱ میلیون تن بذر نیاز است که بیش از یک سوم آن به صورت برنامه‌ریزی شده تولید می‌شود (Zareian *et al.*, 2010). میزان تولید، کیفیت و قیمت تجاری محصول نهایی تحت تأثیر کیفیت بذر قرار می‌گیرد و به منظور دستیابی به بذر سالم و عاری از

امکان‌پذیر است (Xia et al., 2010). دوبای و همکاران (Dubey et al., 2006) پتانسیل شبکه‌های عصبی مصنوعی را در تشخیص سه رقم دانه‌های گندم با استفاده از پردازش تصویر و شکل‌شناسی مطالعه و اعلام کردند که شبکه عصبی مصنوعی (ANN) در ترکیب با پردازش تصویر، در تشخیص ارقام گندم پتانسیل عالی دارد. صحت تشخیص را می‌توان با اضافه کردن دیگر ویژگی‌ها، مانند رنگ و بافت، افزایش داد (Savakar, 2012).

در مطالعات گذشته، تفاوت اندازه دانه گندم سالم، شکستگی دانه، و دانه‌های چروکیده به‌همراه دیگر دانه‌ها به‌طور همزمان بررسی نشده است و در طبقه‌بندی‌ها، محققان دو گروه سالم و معیوب یا چند گروه دانه‌های سالم را در نظر داشته‌اند. همچنین، مطالعات گذشته با استفاده از دوربین عکاسی دیجیتال، دوربین فیلم‌برداری، یا اسکنر بوده است. در به‌کارگیری دوربین دیجیتال صنعتی برای طبقه‌بندی غلات گزارشی ارائه نشده است.

در این پروژه قابلیت سیستم بینایی ماشین بر اساس دوربین دیجیتال صنعتی جهت طبقه‌بندی هفت گروه اجزای کیفی گندم بذری شامل دانه‌های گندم سفید درشت، گندم سفید کوچک، گندم سفید شکسته، گندم سفید چروکیده، گندم قرمز، جو، و چاودار بررسی شده و هدف اصلی آن استخراج و شناسایی ویژگی‌های برتر و ساخت مدل طبقه‌بندی بهینه است.

مواد و روش‌ها

کلیات سیستم بینایی ماشین مورد استفاده شامل یک بخش سخت‌افزاری برای تهیه تصویر و انتقال آن به کامپیوتر و یک بخش نرم‌افزاری برای استخراج ویژگی‌ها و مدل‌های طبقه‌بندی است. مراحل کار به‌ترتیب شامل تهیه

از مهم‌ترین کارها در زمینه شناسایی ارقام مختلف دانه گندم با استفاده از پردازش تصویر و خصوصیات ظاهری، می‌توان به تحقیقات (Dubey et al., 2006; Mahesh et al., 2008; Pourreza et al., 2011; Zapotocny, 2012) اشاره کرد. این محققان به صحت قابل قبولی در شناسایی و طبقه‌بندی ارقام دست یافته‌اند. ویژگی بافت از بهترین گروه ویژگی‌ها در شناسایی ارقام گندم معرفی شده است (Pourreza et al., 2012).

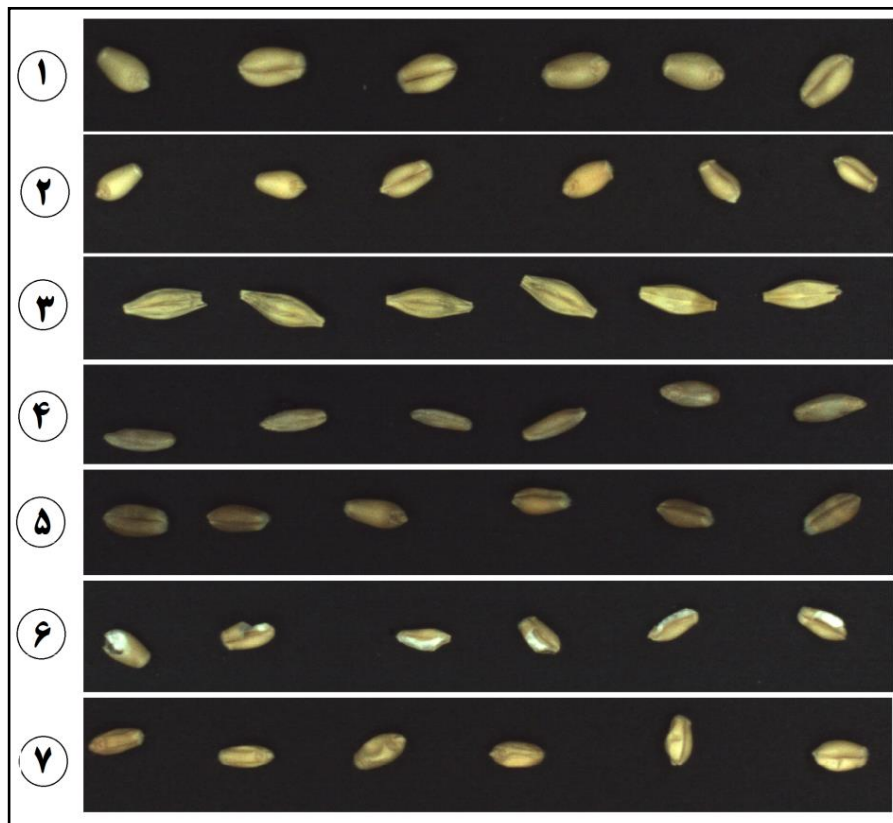
برای طبقه‌بندی تک‌دانه‌ای گندم قرمز، گندم دوروم، جو، و چاودار با استفاده از سه مجموعه ویژگی (شکل، رنگ، و بافت)، ماجومدار و جایاس (Majumdar & Jayas, 2000a) با استفاده از دوربین فیلم‌برداری مطالعاتی جامع انجام داده‌اند. در تحقیقات آنها بیشترین صحت طبقه‌بندی در مدل مرکب از سه گروه ویژگی شکل، رنگ، و بافت حاصل شد. در بیشتر تحقیقات، برای طبقه‌بندی دانه‌های گندم، روش‌های تشخیص تحلیلی خطی، k -نزدیکترین همسایگی و شبکه‌های عصبی مصنوعی نتایج خوب و قابل قبولی داشته‌اند (Paliwal et al., 2003; Wang et al., 2003; Delwiche et al., 2013; Pazoki et al., 2013; Khazaei et al., 2016). تعیین درصد دانه‌های شفاف، ناشاسته‌ای، دو رنگ، و چروکیده در دانه گندم با استفاده از تصویر نور عبوری در توده دانه تک لایه نیز امکان‌پذیر است (Venora et al., 2009). اگر هدف، تشخیص گندم و جو با بینایی ماشین باشد، فقط با به‌کارگیری سه ویژگی می‌توان به حداکثر صحت در طبقه‌بندی بین گندم و جو دست یافت (Guevara-Hernandez & Gomez-Gil, 2011). در بررسی کیفیت ظاهری دانه‌های گندم با استفاده از اسکنر، تشخیص دانه‌های جوانه زده و کپک‌زده به‌راحتی

چاودار (رقم سنآباد) بودند. نمونه‌های جو از مخلوط کردن چهار رقم جو (بهمن، دشت، خرم، ماکویی) متداول در منطقه به دست آمد. هفت گروه دانه‌های مذکور از چهار کارخانه فرآوری بذر در نقاط مختلف استان اردبیل تهیه شد. از هر گروه تشکیل دهنده، ۱۰۰ نمونه ۳۰ تایی (۳۰۰۰ دانه از هر نوع) جداسازی و بسته‌بندی شد. رطوبت نمونه‌های گندم، جو و چاودار به ترتیب ۱۳/۵، ۱۲/۸ و ۱۳ درصد بود. نمونه تصویر یک ردیف از هر گروه دانه در شکل ۱ ارائه شده است.

نمونه، گرفتن تصویر، پردازش تصویر، استخراج و رتبه‌بندی ویژگی‌ها، مدل‌سازی، و مقایسه طبقه‌بندی-کننده‌هاست.

تهیه نمونه‌های دانه

گندم رقم مروارید به عنوان گندم سفید و بذر پایه در نظر گرفته شد و چهار گروه دانه شامل گندم سفید درشت، گندم سفید کوچک، گندم سفید شکسته، گندم سفید چروکیده از توده اصلی نمونه‌برداری شده تهیه شد. سه گروه دیگر شامل دانه‌های گندم قرمز (رقم زاگرس)، جو، و



شکل ۱- نمونه‌هایی از هر گروه دانه (۱) گندم سفید درشت، (۲) گندم سفید کوچک، (۳) جو، (۴) چاودار، (۵) گندم قرمز، (۶) گندم سفید شکسته، (۷) گندم سفید چروکیده

Fig. 1- Samples 1) Coarse white wheat, 2) Small white wheat, 3) Barley, 4) rye, 5) Red wheat, 6) Broken white wheat, 7) Shrunken white wheat

فیلم‌برداری متداول، اندازه پیکسل در سنسور این دوربین در راستای افقی و عمودی یکسان و برابر ۲/۲ میکرومتر و قدرت تفکیک آن در راستای افقی و عمودی به ترتیب ۲۵۹۲ و ۱۹۴۴ پیکسل است. برای

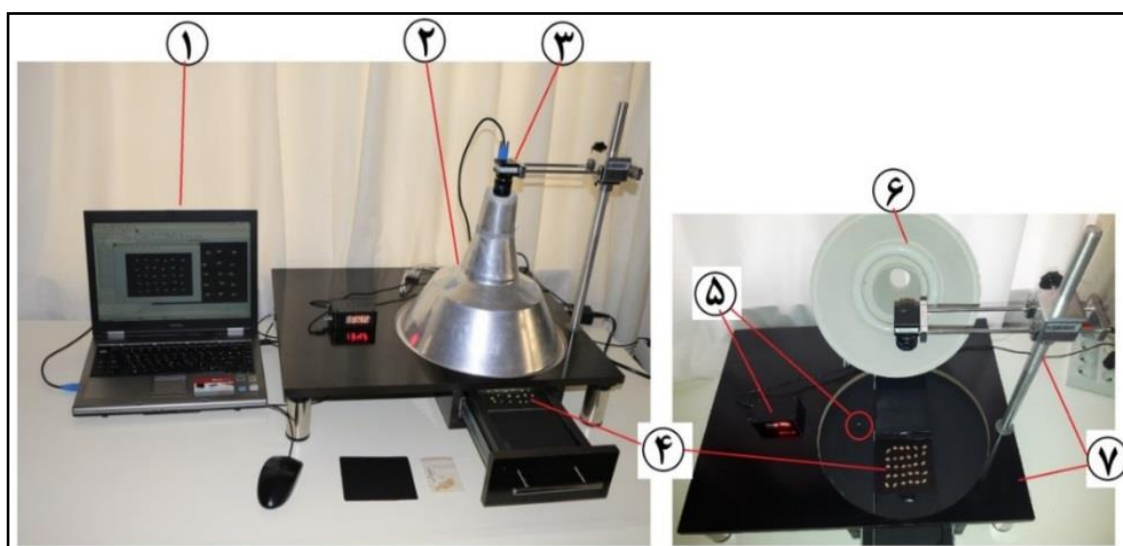
تهیه و پردازش تصاویر

برای تصویربرداری، از یک دوربین دیجیتال صنعتی با حسگر نوع CMOS مدل DFK72AUC02 استفاده شد. برخلاف دوربین‌های عکاسی و

مدار حسگر فتو رزیستور کنترل گردید. مجموعه پایه فلزی، دوربین، محفظه نورپردازی، و حسگر نورسنج روی میز طراحی شده مخصوص نصب و برای قرار دادن نمونه‌های دانه در محفظه نورپردازی، از کشوی ریلی استفاده شد. دوربین از طریق کابل USB به کامپیوتر قابل حمل متصل گردید (شکل ۲). به دلیل قابلیت بالای نرم‌افزار متلب و دارا بودن جعبه ابزارها و توابع مفید، تمامی برنامه‌ها و الگوریتم‌های ارتباط با دوربین، پردازش تصویر، مدل‌سازی، و ارزیابی روش‌های طبقه‌بندی در محیط نرم‌افزار متلب نسخه Matlab 2015a پیاده‌سازی شد.

برنامه‌ها در کامپیوتر همراه مدل Toshiba TECRA Series تدوین و اجرا شد. کامپیوتر مذکور دارای پردازشگر دو هسته‌ای GHZ 2.20 اینتل (Intel(R), Core(TM) Duo CPU 2)، 1 GB حافظه رم و سیستم عامل Windows xp sp3 است.

جلوگیری از اعوجاج تصویر در حاشیه‌های ناحیه تصویربرداری، لنز تله سنتریک ساخت شرکت کامپیوتار^۱ ژاپن، مدل MP1614_mp2, Type 2/3 با فاصله کانونی ۱۶ میلی‌متر به کار گرفته شد. دوربین به فاصله ۳۰ سانتی‌متر از صفحه قرارگیری دانه نصب شد. برای کالیبره کردن مکانی دوربین از سه سکه با مساحت مشخص استفاده و ارتباط بین تعداد پیکسل‌ها و اندازه‌ها روی نمونه واقعی مشخص شد. به دلیل عملکرد خوب لامپ‌های مهتابی (Mahesh *et al.*, 2008) برای نورپردازی نمونه‌ها، از لامپ مهتابی حلقوی نور سفید با توان ۲۲ وات و محفظه مخروطی شکل استفاده شد. برای یکسان‌سازی نورپردازی، دیواره داخلی محفظه نورپردازی سفید رنگ‌آمیزی و با یک لایه پودر اکسید منیزیم پوشانده شد. بین محل دید دوربین و لامپ حلقوی، دیواره فلزی قرار داده شد تا فقط نور انعکاسی از سطح نمونه‌های دانه به لنز دوربین برسد. ثبات شدت نور لامپ از طریق



شکل ۲- اجزای سیستم: (۱) کامپیوتر قابل حمل، (۲) محفظه نور پردازی، (۳) دوربین دیجیتال صنعتی، (۴) کشویی و صفحه نمونه دانه (داخل محفظه)، (۵) نمایشگر شدت نور و نورسنج (داخل محفظه)، (۶) لامپ حلقوی (داخل محفظه)، (۷) پایه دوربین و میز

Fig. 2- System configuration: 1) computer, 2) illumination chamber, 3) industrial digital camera, 4) sample board and rail drawer, 5) photo resistor sensor and display, 6) fluorescent lamp, 7) camera stand and table

محاسبه دیگر ویژگی‌های شکل، روابط زیر در محیط متلب تدوین شد.

نسبت باریکی^۶: از رابطه ۱ محاسبه می‌شود و نشانگر گردی دانه است (Pazoki et al., 2013).

$$TR = \frac{P^2}{4\pi A} \quad (1)$$

که در آن،

TR = نسبت باریکی؛ P = محیط؛ و A = مساحت تصویر دانه (شکل ۳-الف).

نسبت ظاهری^۷: از تقسیم طول محور اصلی (a) بر طول کوچکترین محور (b) محاسبه می‌شود؛ مطابق شکل ۳-ب، a و b محورهای اصلی در بیضی هستند که گشتاور دوم آن برابر گشتاور دوم سطح تصویر دانه است (Delwiche et al., 2013).

نسبت ظاهری مستطیلی^۸: از تقسیم طول دانه (L) بر عرض آن (W) به دست می‌آید؛ مطابق شکل ۳-الف، طول و عرض دانه برابر ابعاد مستطیل محیطی دانه است.

نسبت مساحت^۹: از رابطه ۲ محاسبه می‌شود.

$$Area Ratio = \frac{L \times W}{A} \quad (2)$$

که در آن،

A = مساحت تصویر دانه؛ L = طول دانه؛ و W = عرض آن (شکل ۳-الف).

بزرگترین شعاع^{۱۰}: بیشترین فاصله پیکسل موجود در محیط دانه از مرکز دانه (C) در شکل ۳-الف با R_{max} نشان داده شده است و با محاسبه بیشترین مقدار پارامتر (d_k) به دست می‌آید. d_k در رابطه ۳ معرفی شده است.

از مقوای مشکی نوع اشتهنباخ (code 08) به عنوان صفحه زمینه قرارگیری دانه‌ها استفاده شد. تصاویر دانه تحت فضای RGB با تفکیک‌سازی ۵ مگاپیکسل (۲۵۶۰ در ۱۹۲۰) در کامپیوتر ذخیره شد. بدین ترتیب ۷۰۰ تصویر (۱۰۰ تصویر از هر کلاس دانه‌ها) به دست آمد که هر تصویر شامل ۳۰ دانه بود. برنامه‌ای یکپارچه نوشته شد تا عملیات پیش‌پردازش تصویر و استخراج ویژگی‌های شکل را پیش برد و با فراخوانی توابع استخراج ویژگی‌های رنگ و بافت، بردار مجموعه ویژگی‌ها برای هر دانه را محاسبه و ذخیره کند. با تحلیل و پردازش تصاویر، در مجموع ۹۱ ویژگی برای ۲۱۰۰۰ دانه محاسبه شد. ۱۰ درصد داده‌ها (۲۱۰۰ دانه) به صورت تصادفی جدا و از آنها به عنوان داده‌های مستقل برای آزمون و مقایسه عملکرد مدل‌ها استفاده شد. با استفاده از داده‌های باقیمانده (۱۸۹۰۰ دانه) تهیه مدل‌های طبقه‌بندی انجام شد.

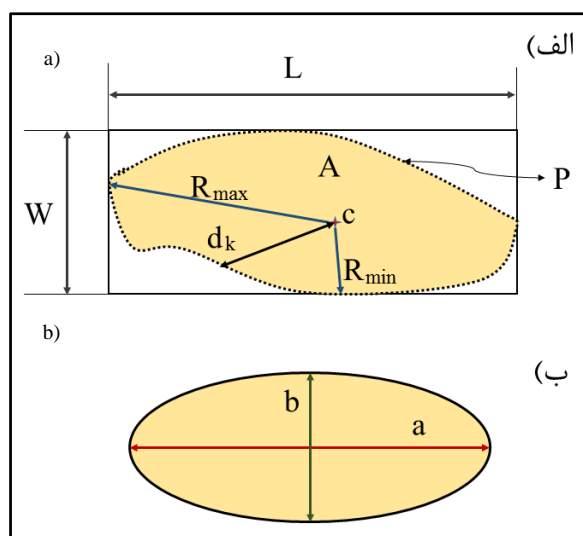
محاسبه ویژگی‌های شکل

الگوریتم محاسبه ویژگی‌های مورفولوژیکی در داخل برنامه اصلی پیش‌پردازش تصویر نوشته شد و این الگوریتم با دریافت تصویر باینری و برچسب‌گذاری شده، ویژگی‌های زیر را برای هر دانه به طور مجزا محاسبه نمود.

با استفاده از ضریب کالیبراسیون، واحد طول از پیکسل به میلی‌متر تبدیل و تمامی ویژگی‌های دارای واحد طول بر اساس میلی‌متر محاسبه گردید. ویژگی‌های مساحت، محیط، طول محور اصلی^۱، طول کوچکترین محور^۲، صلیبیت^۳، قطر معادل^۴، و خارج از مرکزی^۵ با استفاده از تابع regionprops موجود در جعبه ابزار متلب استخراج شد. برای

1- Major Axis Length
3- Solidity
5- Eccentricity
8- Rectangular Aspect Ratio
10- Maximum Radius

2- Minor Axis Length
4- Equivalent Diameter
6- Thinness Ratio
7- Aspect Ratio
9- Area Ratio



شکل ۳- الف) شماتیک دانه و مستطیل محیطی آن و ب) بیضی معادل گشتاور دوم تصویر دانه

Fig. 3- a) Schematic of the seed and its peripheral rectangle, b) equivalent second moment ellipse of the seed image

$$d_k = \left[(i_k - c_{ik})^2 + (j_k - c_{jk})^2 \right]^{1/2} \quad (3)$$

که در آن،

(i_k, j_k) = مختصات k امین پیکسل روی محیط دانه و (c_{ik}, c_{jk}) = مختصات مرکز k امین دانه.

تابع فاصله (d_k) با استفاده از ضریب کالیبراسیون $(mm/pixel)$ به میلی‌متر تبدیل شد. بزرگی توصیف‌گر فوریه از رابطه ۴ محاسبه می‌شود (Majumdar & Jayas, 2000a).

$$FD_u = [R_u^2 + I_u^2]^{1/2} \quad (4)$$

برای مقادیر u مقدار حقیقی توصیف‌گر به صورت رابطه ۵ تعریف می‌شود.

$$R_u = \sum_{k=0}^{N-1} d_k \cdot \cos \left[\frac{2\pi k u}{N} \right] \quad (5)$$

مقدار مجازی توصیف‌گر به صورت رابطه ۶ تعریف می‌شود.

$$I_u = \sum_{k=0}^{N-1} d_k \cdot \sin \left[\frac{2\pi k u}{N} \right] \quad (6)$$

که در آن،

کوچکترین شعاع^۱: کمترین فاصله بین پیکسل روی محیط دانه از مرکز دانه (C) در شکل ۳- الف با R_{min} نشان داده شده است و با محاسبه کمترین مقدار پارامتر (d_k) به دست می‌آید.

نسبت شعاعی^۲: از تقسیم بزرگترین شعاع (R_{max}) بر کوچک‌ترین شعاع (R_{min}) به دست می‌آید. میانگین شعاع‌ها^۳: میانگین فاصله پیکسل‌های روی محیط دانه از مرکز (C) آن است و با μ_R نشان داده می‌شود. با محاسبه مقدار میانگین پارامتر (d_k) تعیین می‌شود.

انحراف استاندارد شعاع‌ها^۴: انحراف استاندارد فاصله تمام پیکسل‌های روی لبه از مرکز دانه است که با σ_R نشان داده می‌شود. با محاسبه مقدار انحراف استاندارد پارامتر (d_k) تعیین می‌شود.

نسبت هارالیک^۵: از تقسیم میانگین شعاع‌ها به انحراف استاندارد آنها (μ_R/σ_R) محاسبه می‌شود (Majumdar & Jayas, 2000b).

توصیف‌گرهای فوریه^۶: ابتدا تابع فاصله تک جهت d_k برای تمام پیکسل‌های لبه هر دانه با رابطه ۳ محاسبه می‌شود.

1- Minimum Radius
3- Mean Radius
5- Haralick Ratio

2- Radius Ratio
4- Standard Deviation of all Radius
6- Fourier Descriptors

استفاده از آن سه پارامتر میانگین، واریانس، و انحراف استاندارد برای شش تکرنگ (R, G, B, H, S, I) محاسبه شد و در مجموع ۱۸ ویژگی رنگ برای هر دانه به دست آمد.

محاسبه ویژگی‌های بافت

تابع نوشته شده برای استخراج ویژگی‌های بافت، تصویر رنگی و افقی داخل چهار ضلعی محیطی هر دانه را دریافت و ویژگی‌های بافت را محاسبه می‌کند. الگوریتم مذکور ابتدا سه رنگ اصلی قرمز، سبز و آبی (R, G, B) و ترکیب‌های ممکن آنها شامل، $\{X_1=(3R+2G+B)/6\}$ ، $\{X=(R+G+B)/3\}$ ، $\{X_2=(2R+1G+3B)/6\}$ و $\{X_3=(1R+3G+2B)/6\}$ را استخراج می‌کند. با این کار، باند رنگ ویژه یا ترکیب‌های آنها که صحت بهتری در طبقه بندی نسبت به دیگری ایجاد می‌کنند تعیین خواهد شد. برای کاهش زمان محاسباتی، مقدار سطح خاکستری (یا شدت) در تصاویر هر هفت باند مذکور از ۲۵۶ به ۸ کاهش داده شد. ماتریس سطح خاکستری هم‌رویدادی (GLCM) مستقل از جهت، برای تمامی آنها محاسبه شد. با تقسیم ماتریس شدت هم‌رویدادی^۱ (GLCM) هر یک از باندهای رنگی به مجموع عناصر خود ماتریس، مقادیر نرمال شده آن به دست آمد. با استفاده از آن، شش ویژگی بافت برای هر یک از هفت باند رنگی به ترتیب زیر محاسبه شد و در مجموع ۴۲ ویژگی بافت برای هر دانه به دست آمد.

کنتراست: مقیاسی از کنتراست، شدت بین یک پیکسل و همسایه آن روی کل تصویر است (رابطه ۱۰).

$$Contrast = \sum_{i=1}^K \sum_{j=1}^K (i-j)^2 p_{ij} \quad (10)$$

که در آن،

N = تعداد پیکسل‌های موجود در محیط دانه.

بدین طریق، هفت توصیف‌گر اولیه فوریه با اختصاص مقادیر صفر تا شش برای u محاسبه شدند. محاسبه گشتاورها: گشتاورهای ثابت مکانی^۱ مجموعه‌ای از ثابت‌های ممانی غیر حساس به انتقال، تجانس، انعکاس و دوران هستند (Guevara-Hernandez & Gomez-Gil, 2011). این گشتاورها اندازه‌گیری‌های آماری صفات یک شیء هستند. هفت گشتاور ثابت و استاندارد با استفاده از گشتاورهای مرکزی نرمال، معرفی شده توسط گنزالس و همکاران (Gonzalez et al., 2009) برای تصویر باینری هر دانه محاسبه شدند و در مجموع ۳۱ ویژگی شکل برای هر دانه به دست آمد.

محاسبه ویژگی‌های رنگ

تابع نوشته شده برای استخراج ویژگی‌های رنگ، تصویر رنگی هر دانه را دریافت کرده پس از تفکیک‌سازی سه رنگ اصلی قرمز، سبز و آبی (R, G, B) محاسبات را به ترتیب زیر انجام داد. ابتدا مقادیر (HSI) از روابط ۷، ۸ و ۹ محاسبه شد (Gonzalez et al., 2009).

$$I = \frac{R + G + B}{3} \quad (7)$$

$$S = 1 - \frac{3 \text{Min}(R, G, B)}{I} \quad (8)$$

$$H = \cos^{-1} \left\{ \frac{\frac{1}{2}[(R-G)+(R-B)]}{[(R-G)^2+(R-B)(G-B)]^{1/2}} \right\} \quad (9)$$

که در آنها،

I = شدت رنگ؛ S = اشباع؛ و H = اصل رنگ (پرده رنگ). مقادیر RGB قبل از به‌کارگیری در روابط بالا با تقسیم به عدد (R+G+B) نرمال‌سازی شدند. تعداد پیکسل‌های هر دانه (مساحت دانه) به‌عنوان فراوانی هر یک از تکرنگ‌ها در نظر گرفته شد و با

$$Variance = \sum_{i=1}^K \sum_{j=1}^K (i - m)^2 p_{ij} \quad (15)$$

که در آن،

m = میانگین ماتریس هم‌رویدادی است (Majumdar & Jayas, 2000c).

رتبه‌بندی ویژگی‌ها

به‌کارگیری ویژگی‌های بیشتر از تعداد بهینه، نه تنها صحت طبقه‌بندی را بهبود نمی‌بخشد بلکه موجب بدتر شدن عملکرد طبقه‌بندی نیز می‌شود (Paliwal et al., 2003; Reunanen, 2003; Chelladurai et al., 2012). از دیدگاه آماری، مثال‌هایی وجود دارد که در آنها ویژگی‌های غیر وابسته، ولی نویزدار، اطلاعات کمی فراهم می‌کنند. در میان روش‌های مختلف کاهش ویژگی، الگوریتم Relief برای ویژگی‌های دارای نویز یا ویژگی‌های دارای همبستگی خوب کار می‌کند و می‌تواند وابستگی بین ویژگی‌ها و گروه‌های مورد نظر را در سیستم‌های طبقه‌بندی محاسبه کند (Robnik-Šikonja & Kononenko, 2003). به‌طور کلی، باید به‌دنبال ویژگی‌هایی باشیم که نویز کمتری دارند و مقادیر آنها در هر یک از گروه‌ها تفاوت معنی‌داری داشته باشند. به‌منظور تعیین اهمیت هر یک از ویژگی‌ها و انتخاب بهترین آنها از میان کل ۹۱ ویژگی، رتبه‌بندی ویژگی‌ها با پیاده‌سازی الگوریتم Relief_F در نرم‌افزار متلب دنبال شد. زیرمجموعه پیش‌فرض مناسب برای شروع برنامه پس از بررسی حالت‌های مختلف، برابر ۱۵ قرار داده شد.

تهیه مدل‌های طبقه‌بندی

روش آماری: تحلیل تفکیک‌کننده خطی آماری را ابتدا فیشر (Fisher, 1936) معرفی کرد و پس از او توسعه داده شد. این مدل‌ها با استفاده از متغیرهای

i و j = به‌ترتیب شاخص‌های سطر و ستون ماتریس هم‌رویدادی؛ K = مقیاس سطح خاکستری تصویر که در اینجا برابر ۸ قرار داده شد؛ و P_{ij} = عناصر نرمال ماتریس هم‌رویدادی. همگنی^۱: درجه نزدیکی توزیع عناصر در ماتریس هم‌رویدادی به قطر آن را نشان می‌دهد (رابطه ۱۱).

$$Homogeneity = \sum_{i=1}^K \sum_{j=1}^K \frac{p_{ij}}{1 + |i - j|} \quad (11)$$

انرژی^۲: مجموع مربعات عناصر ماتریس هم‌رویدادی نرمال را محاسبه می‌کند (رابطه ۱۲).

$$Energy = \sum_{i=1}^K \sum_{j=1}^K (p_{ij})^2 \quad (12)$$

آنترپی^۳: میزان تصادفی بودن شدت پیکسل‌ها را نشان می‌دهد. نرمال آن طبق رابطه ۱۳ محاسبه می‌شود.

$$Entropy = - \sum_{i=1}^K \sum_{j=1}^K p_{ij} \log_2 p_{ij} \quad (13)$$

همبستگی^۴: نشان می‌دهد که یک پیکسل چگونه با همسایه‌اش روی کل تصویر همبستگی دارد (رابطه ۱۴).

$$Correlation = \sum_{i=1}^K \sum_{j=1}^K \frac{(i - m_r)(j - m_c)p_{ij}}{\sigma_r \sigma_c} \quad (14)$$

که در آن،

m_r و m_c = میانگین‌های سطری و ستونی ماتریس هم‌رویدادی؛ و σ_r و σ_c = به‌ترتیب انحراف استاندارد سطری و ستونی ماتریس هم‌رویدادی (Gonzalez et al., 2009).

واریانس^۵: واریانس ماتریس هم‌رویدادی را محاسبه می‌کند (رابطه ۱۵).

1- Homogeneity
3- Entropy
5- Variance

2- Energy
4- Correlation

در میان ساختارهای مختلف شبکه‌های عصبی مصنوعی، نوع چند لایه پیشخور بیشترین کاربرد را دارند (Szczypiński *et al.*, 2015).

در این تحقیق، از ساختار مذکور با روش آموزش استاندارد پس انتشار خطا و تابع تبدیل تانژانت سیگموئید هایپربولیک استفاده شد. تابع اجرا از نوع حداقل مربعات خطا (MSE) انتخاب و برای کاهش مدت زمان آموزش، در شبکه اول فقط از یک لایه مخفی و در شبکه دوم از دو لایه مخفی استفاده شد. برای بررسی اثر تعداد ویژگی‌های رتبه بندی شده بر صحت طبقه بندی شبکه، تعداد نرون‌های لایه پنهان در هر دو ساختار مورد نظر به صورت اتوماتیک و با رابطه ۱۶ محاسبه شد (Chen *et al.*, 2010).

$$h = (m + n)^{1/2} + a, \quad a \in [0,10] \quad (16)$$

که در آن،

m = تعداد نرون‌های خروجی برابر تعداد گروه‌های دانه (تعداد کلاس‌ها)؛ n = تعداد نرون‌های ورودی برابر با تعداد ویژگی‌های ورودی؛ a = مقدار ثابت انتخابی؛ و h = مجموع تعداد نرون‌های لایه یا لایه‌های پنهان. با آزمون و خطا و ملاک قرار دادن حداکثر صحت طبقه بندی، مقدار a برای ساختار اول با تک لایه پنهان و ساختار دوم با دو لایه پنهان به ترتیب ۸ و ۱۰ قرار داده شد. بدین ترتیب، ساختار کلی شبکه با یک لایه پنهان به شکل (n_h_m) و با دو لایه پنهان به شکل $(n_h/2_h/2_m)$ طراحی شد.

ارزیابی عملکرد طبقه بندی کننده‌ها

بررسی منابع علمی متعدد نشان می‌دهد که متداول‌ترین معیار بیان صحت طبقه بندی در مسائل چند کلاسه، محاسبه درصد طبقه بندی صحیح است و به معنای نسبت تشخیص صحیح کلاس‌های هدف

غیر وابسته، تابع تفکیک کننده ایجاد می‌کنند و با استفاده از آنها گروه‌های جدید طبقه بندی و شناسایی می‌شوند (Guo *et al.*, 2007). در اکثر مطالعات، طبقه بندی کننده‌های آماری صحت مورد قبولی در طبقه بندی گروه‌های مختلف دانه‌ها داشته‌اند (Shahin & Symons, 2001; Savakar, 2012; Khazaei *et al.*, 2016). الگوریتم‌های طبقه بندی آماری شامل تحلیل تفکیکی آماری خطی و درجه دوم در متلب ایجاد شدند و با استفاده از آنها مدل‌های تحلیل تفکیک خطی (LDA) و تحلیل تفکیک درجه دوم (QDA) برای تمام ویژگی‌های رتبه بندی شده آموزش دیدند.

روش K- نزدیک‌ترین همسایگی (KNN): در این روش تصمیم‌گیری این‌که نمونه‌ای جدید در کدام کلاس قرار گیرد، با بررسی تعداد (k) از شبیه‌ترین نمونه‌ها یا همسایه‌ها تعیین می‌شود. در بین این k نمونه، تعداد نمونه‌ها برای هر کلاس شمرده می‌شود و نمونه جدید به کلاسی نسبت داده می‌شود که تعداد بیشتری از همسایه‌ها به آن تعلق دارند. این روش در شرایط مختلف استفاده و عملکرد آن در طبقه بندی داده‌های نوین دار ثابت شده است (Silva *et al.*, 2015). در تحلیل مقدماتی قابلیت طبقه بندی این روش، تابع محاسبه فاصله از نوع اقلیدسی و تعداد همسایگی ۱۰ بیشترین صحت را داشت. از این‌رو، در مدل سازی از حالت مذکور استفاده شد.

شبکه عصبی مصنوعی (ANN): توانایی مدل شبکه عصبی مصنوعی با الگوریتم پس انتشار خطا با یک یا دو لایه پنهان در حل مسائل طبقه بندی ثابت شده است. اکثر محققان برای تعیین تعداد نرون‌های لایه پنهان از روش آزمون و خطا استفاده کرده‌اند (Azizi *et al.*, 2015; Khazaei *et al.*, 2016).

نتایج و بحث

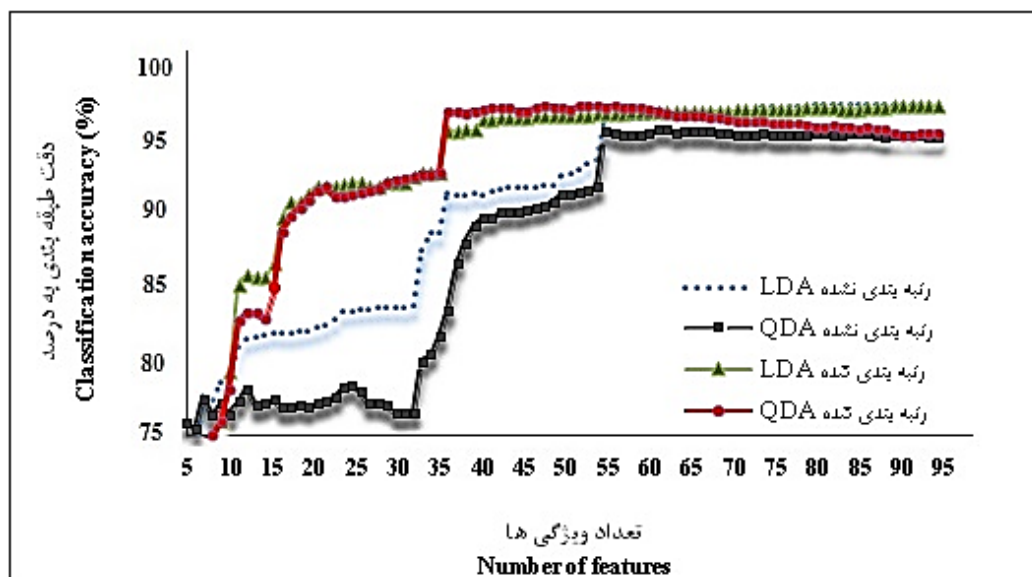
در شکل ۴، تأثیر رتبه‌بندی ۹۱ ویژگی با الگوریتم Relief_F بر صحت طبقه‌بندی و شناسایی ویژگی‌های مؤثر ارائه شده است. به دلیل سرعت بالای مدل‌سازی و طبقه‌بندی روش‌های آماری از این مدل‌ها برای ارزیابی استفاده شده است. این نمودار روند افزایش صحت طبقه‌بندی در اعمال دو گروه ویژگی‌های رتبه‌بندی نشده و رتبه‌بندی شده را نشان می‌دهد و مشخص است در حالت استفاده از ۳۵ ویژگی رتبه‌بندی شده صحت طبقه‌بندی در مدل QDA حداکثر مقدار شده است. ولی این مقدار در حالت ویژگی‌های رتبه‌بندی نشده بیشتر از ۶۰٪ است. این نتیجه، قابلیت الگوریتم مورد استفاده را در حذف ویژگی‌های کم‌اهمیت تأیید می‌کند.

است. با در نظر گرفتن ماتریس اغتشاش هر یک از مدل‌های طبقه‌بندی، صحت طبقه‌بندی از رابطه ۱۷ محاسبه شد (Vanloot *et al.*, 2014).

$$ccr = \frac{N_c}{N_c + N_{ic}} \times 100 \quad (17)$$

که در آن،

N_c = تعداد نمونه‌هایی که کلاس واقعی آنها درست تشخیص داده شده‌اند؛ و N_{ic} = تعداد نمونه‌های کلاس هدف اشاره می‌کند که کلاس آنها نادرست تخمین زده شده است. با اعمال رابطه ۱۷ به هر سطر ماتریس اغتشاش، صحت طبقه‌بندی هر کلاس و با اعمال آن به کل ماتریس اغتشاش صحت طبقه‌بندی کل مدل محاسبه می‌شود.



شکل ۴- تأثیر رتبه‌بندی ویژگی‌ها بر صحت کل طبقه‌بندی‌کننده‌های آماری بر اساس تعداد ویژگی

Fig. 4- Influence of feature rankings on the classification accuracy results of the statistical classifiers, obtained in respect to the number of ranked features

هستند که نسبت مساحت، مؤثرترین آنهاست. دلیل آن تفاوت‌های دانه‌ها از لحاظ شکل لبه دانه و یکنواختی این ویژگی‌های شکل در اکثر گروه‌های

۳۵ ویژگی برتر بر حسب وزن اشتراک آنها در مدل طبقه‌بندی، به صورت نزولی در جدول ۱ آمده است. ۱۰ ویژگی مهم اول جزء ویژگی‌های شکل

تعیین ویژگی‌ها و مدل بهینه به منظور طبقه بندی اجزای ...

شد سهم کمتری از ویژگی‌های برتر را داشته باشند. تفاوت‌های بافتی سطح دانه‌ها و همبستگی کمتر بین آنها باعث شد تا ویژگی‌های بافت سهم خوبی از ویژگی‌های برتر را داشته باشند. از میان ویژگی‌های تاثیرگذار بافت، آنتروپی تک‌رنگ‌ها و ترکیب رنگ‌های مختلف بیشترین سهم را دارند. دلیل آن تفاوت بافت سطحی دانه‌ها از لحاظ میزان تصادفی بودن شدت رنگ پیکسل‌ها نسبت به هم است.

دانه، غیر از گندم شکسته، است. در بین ۳۵ ویژگی برتر اول، ۱۶ ویژگی شکل، ۸ ویژگی رنگ و ۱۱ ویژگی بافت قرار دارد. سهم ویژگی‌های رنگ کمتر از سهم ویژگی‌های شکل و بافت است. ویژگی‌های رتبه یازدهم و پانزدهم تا هجدهم میانگین و واریانس رنگ‌ها هستند؛ دلیل آن، اختلاف ظاهر دانه‌ها از لحاظ ترکیب رنگی سطح است. ولی واریانس بیشتر ویژگی‌های رنگ در داخل هر گروه از دانه‌ها باعث

جدول ۱- رتبه‌بندی ویژگی‌ها با روش Relief F: شناسه ویژگی مورفولوژیکی (M)، شناسه ویژگی رنگ (CC) و شناسه ویژگی بافت (T)

Table 1- Ranked features by relief F algorithm; [Morphological (M); Color (C), Texture (T)]					
Ranking	Feature	Weight	Ranking	Feature	Weight
رتبه	ویژگی	وزن	رتبه	ویژگی	وزن
1	Area Ratio (M)	96	19	Red GLCM Contrast (T)	55
2	Area (M)	91	20	Red GLCM Entropy (T)	53
3	First invariant moment (M)	91	21	Blue GLCM Entropy (T)	50
4	Minor Axis (M)	88	22	Haralik Ratio (M)	47
5	Mean Radius (M)	87	23	X2 GLCM Entropy (T)	45
6	Major Axis (M)	83	24	X3 GLCM Entropy (T)	44
7	Radius Standard Deviation (M)	80	25	Sixth Fourier descriptor (M)	44
8	Solidity (M)	80	26	Blue GLCM Correlation (T)	44
9	Maximum Radius (M)	77	27	X GLCM Entropy (T)	44
10	Perimeter (M)	75	28	Hue Mean (C)	44
11	Saturation Mean (C)	74	29	Green GLCM Entropy (T)	42
12	Minimum Radius (M)	73	30	Blue Mean (C)	42
13	First Fourier descriptor (M)	71	31	Sixth invariant moment (M)	42
14	Third Fourier descriptor (M)	66	32	Red Standard Deviation (C)	42
15	Saturation Standard deviation (C)	64	33	Green GLCM Contrast (T)	41
16	Red Mean (C)	61	34	X GLCM Homogeneity (T)	40
17	Green Mean (C)	61	35	Red GLCM Homogeneity (T)	40
18	Saturation Variance (C)	59			

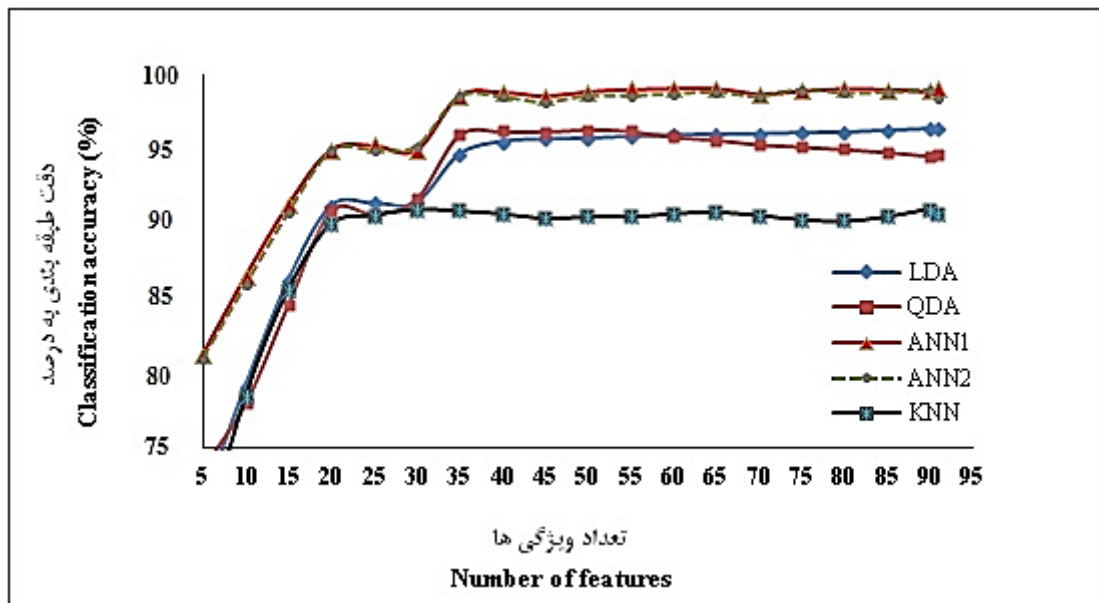
افزایش تعداد ویژگی‌های برتر به ۲۰، صحت کل تا بیش از ۹۰ درصد برای تمامی مدل‌ها افزایش یافته و تا به کارگیری ۳۰ ویژگی تقریباً ثابت مانده است. با ادامه افزایش ویژگی‌ها تا ۳۵ ویژگی، صحت کل برای تمام مدل‌ها غیر از مدل k- نزدیک‌ترین همسایگی (KNN)، افزایش یافته است. بعد از آن در مدل‌های شبکه عصبی تقریباً روند ثابت دارد، در مدل LDA افزایش ناچیز و در مدل QDA کاهش آن مشاهده

پنج مدل طبقه‌بندی توضیح داده شده در بخش مواد و روش‌ها، با به کارگیری ویژگی‌های رتبه‌بندی شده و تقسیم داده‌ها به پنج قسمت بر اساس روش اعتبارسنجی متقابل، مدل‌سازی شدند. در شکل ۵، صحت کل طبقه‌بندی کننده‌ها نسبت به تعداد ویژگی‌ها، به ترتیب رتبه، آمده است. در حالت استفاده فقط از ۱۰ ویژگی اول، صحت کل طبقه‌بندی ضعیف و کمتر از ۸۵ درصد است. با

گروه دانه‌های (گندم سفید درشت، گندم سفید کوچک، گندم سفید شکسته، گندم سفید چروکیده، گندم قرمز، جو، و چاودار) به‌کار گرفت. صحت کل طبقه‌بندی بر اساس داده‌های آموزشی در هریک از روش‌های طبقه‌بندی (مدل شده بر اساس ۳۵ ویژگی برتر) شامل شبکه عصبی مصنوعی تک لایه (ANN1)، شبکه عصبی مصنوعی دو لایه (ANN2)، k- نزدیکترین همسایگی (KNN)، تحلیل تفکیک خطی (LDA)، و تحلیل تفکیک درجه دوم (QDA)، به ترتیب ۹۸/۷، ۹۸/۶، ۹۰/۹، ۹۴/۷ و ۹۶ درصد است (شکل ۵).

شده است. مدل K- نزدیکترین همسایگی از ویژگی بیستم به بعد افزایش محسوسی در صحت کل طبقه‌بندی ندارد (شکل ۲).

افزایش بیشتر تعداد ویژگی‌های مورد استفاده در طبقه‌بندی، در برخی مواقع نه تنها موجب بهبود صحت طبقه‌بندی نشده بلکه به بدتر شدن صحت انجامیده است. این مسئله در مدل (QDA) مشاهده شد. غیر از ۳۵ ویژگی اول، دیگر ویژگی‌ها وزن ضعیفی در مدل‌ها دارند. بنابراین، بدون تأثیر بر صحت طبقه‌بندی، قابل حذف هستند و می‌توان تنها ۳۵ ویژگی برتر (جدول ۲) را در طبقه‌بندی هفت



شکل ۵- صحت کل طبقه‌بندی تحلیل تفکیک خطی (LDA)، تحلیل تفکیک غیر خطی (QDA)، شبکه عصبی مصنوعی تک لایه (ANN1)، شبکه عصبی مصنوعی دو لایه (ANN2) و k- نزدیکترین همسایگی (KNN) براساس تعداد ویژگی‌های رتبه بندی شده
 Figure 5- Classification accuracy results obtained in respect to the number of ranked features employed, using five classifiers (LDA, QDA, ANN1, ANN2, and KNN)

(۷-۱۴-۳۵) و شبکه عصبی دو لایه (ANN2) به‌صورت (۷-۸-۸-۳۵) ایجاد شد. نتایج پنج مدل انتخاب شده با اعمال داده‌های مستقل در جدول ۲ آمده است. صحت کل طبقه‌بندی برای هفت کلاس دانه در مدل K- نزدیکترین همسایگی، کمترین

برای پیش‌بینی هفت کلاس دانه بر اساس داده‌های مستقل که در آموزش و تولید مدل‌ها شرکت ندارند، فقط مدل‌های ایجاد شده با ۳۵ ویژگی برتر به‌کار گرفته شدند. در این حالت، ساختار شبکه عصبی تک لایه (ANN1) به‌صورت

مقدار و عامل کاهش صحت کل طبقه بندی در تمامی مدل‌هاست. دانه‌های گندم سفید درشت، گندم قرمز، و جو در تمامی مدل‌ها با بیشترین صحت تشخیص داده شده‌اند. دانه‌های جو در سه مدل با صحت ۱۰۰ درصد شناسایی شده‌اند (جدول ۲). قابلیت خوب تشخیص دانه‌های جو از دانه‌های گندم مطابق با نتایج مطالعات گواوارا-هرناندز و گومز-گیل (Guevara-Hernandez & Gomez-Gil, 2011) است.

مقدار (۹۱/۶ درصد) است. بیشترین مقدار آن ۹۷/۶ درصد نیز در مدل شبکه عصبی دو لایه (ANN2) با ۸ نرون در هر لایه مخفی مشاهده شده است. نتایج طبقه بندی با استفاده از داده‌های مستقل تفاوت محسوسی با صحت طبقه بندی در استفاده از داده‌های آموزشی نشان نمی‌دهد و نشانگر قابلیت تعمیم مدل‌ها در پیش‌بینی کلاس داده‌های جدید است. صحت طبقه بندی دانه‌های سفید چروکیده، در مقایسه با دیگر دانه‌ها در تمامی مدل‌ها، کمترین

جدول ۲- صحت طبقه بندی (به درصد) مدل‌های ایجاد شده بر اساس ۳۵ ویژگی (با اعمال داده‌های مستقل) برای دانه‌های

۱- گندم سفید درشت، ۲- گندم سفید کوچک، ۳- جو، ۴- چاودار، ۵- گندم قرمز، ۶- گندم سفید شکسته، ۷- گندم سفید چروکیده

Table 2- The results of classification accuracy produced by five models on test data, using the 35 first features; kernel classes are: 1) coarse white wheat, 2) small white wheat, 3) barley, 4) rye, 5) red wheat, 6) broken white wheat, 7) shrunken white wheat.

classification models مدل طبقه بندی	kernels classes & classification accuracy طبقه بندی دانه‌ها و صحت طبقه بندی							average of classification accuracy [%] میانگین صحت طبقه بندی (درصد)
	1	2	3	4	5	6	7	
LDA	99.7	93.7	99.7	95.7	98.7	94.7	83	95
QDA	98	98.3	100	98	98.3	97.7	83.3	96.7
ANN1	99	96.7	100	97	99.7	99.7	89.3	97.3
ANN2	100	96.7	99.7	98	99	99.3	90.3	97.6
KNN	98	95	100	97	98.7	84	68.3	91.6

در جدول ۳ دیده می‌شود که دانه‌های گندم سفید درشت، گندم سفید کوچک، گندم سفید شکسته، گندم قرمز، جو، و چاودار با صحت بالایی تشخیص داده شده‌اند. صحت طبقه بندی خوب و تداخل ناچیز دانه‌های گندم سفید درشت و جو، ناشی از تفاوت معنی‌دار آنها از نظر شکل، رنگ، و بافت است. دانه‌های گندم سفید چروکیده با صحت ۹۰/۳ درصد طبقه بندی شده‌اند. مشابهت شکل و رنگ گندم سفید چروکیده با دانه‌های گندم سفید کوچک، گندم سفید شکسته و چاودار صحت طبقه بندی را تحت تأثیر قرار داده و از آن کاسته است به طوری که از ۳۰۰ دانه گندم سفید چروکیده ۱۳ دانه به اشتباه در کلاس دانه‌های گندم سفید شکسته طبقه بندی شده‌اند.

زمان مورد نیاز برای آموزش مدل‌های شبکه عصبی مصنوعی، در مقایسه با مدل‌های آماری طبقه بندی، خیلی بیشتر است، ولی از نظر سرعت طبقه بندی مدل ساخته شده در برخورد با داده‌های مستقل، این دو تفاوت محسوسی ندارند. بنابراین، بسته به این که زمان آموزش نسبت به صحت طبقه بندی چقدر اهمیت داشته باشد مدل مناسب را می‌توان مطابق آن انتخاب کرد. در اینجا مدل تهیه شده به روش شبکه عصبی مصنوعی با دو لایه پنهان (ANN2)، به دلیل بیشترین صحت طبقه بندی در مواجهه با دانه‌های جدید، به عنوان مدل برتر انتخاب شد. ماتریس اغتشاش این مدل در جدول ۳ برای شناسایی دانه‌های جدید (غیر از دانه‌های به کارگیری شده در آموزش) ارائه شده است.

جدول ۳- ماتریس اغتشاش مدل شبکه عصبی دو لایه (ANN2) بر اساس ۳۵ ویژگی (با اعمال داده‌های مستقل) برای دانه‌های ۱- گندم سفید درشت، ۲- گندم سفید کوچک، ۳- جو، ۴- چاودار، ۵- گندم قرمز، ۶- گندم سفید شکسته، ۷- گندم سفید چروکیده
 Table 3- Confusion matrix of ANN2 classifier using the 35 first features on test data; kernel classes are: 1) coarse white wheat, 2) small white wheat, 3) barley, 4) rye, 5) red wheat, 6) broken white wheat, 7) shrunken white wheat.

Target classes طبقه های هدف	Test samples نمونه های آزمایشی	Output classes طبقه های خروجی							classification [%] طبقه بندی (%)
		1	2	3	4	5	6	7	
1	300	300	0	0	0	0	0	0	100
2	300	0	290	0	0	0	0	10	96.7
3	300	0	0	299	1	0	0	0	99.7
4	300	0	1	3	294	2	0	0	98
5	300	0	1	0	2	297	0	0	99
6	300	0	0	0	0	0	298	2	99.3
7	300	0	11	0	4	1	13	271	90.3
classification accuracy of model صحت طبقه بندی مدل									97.6

می‌توان نتیجه گرفت که تصویربرداری با تفکیک‌سازی ۵ مگاپیکسل (۲۵۶۰ در ۱۹۲۰) در ارزیابی ظاهری دانه‌های گندم کافی است و سیستم نورپردازی با لامپ مهتابی به دلیل ثبات در شدت نور، مناسب تصویر برداری در شرایط بسته (کنترل شده) است. به کارگیری تمامی ویژگی‌ها ممکن است صحت طبقه‌بندی را کم کند. محدوده صحت طبقه‌بندی مناسب با تعداد ویژگی کمتر، قابلیت خوب به کارگیری روش Relief_F را در رتبه‌بندی ویژگی‌ها نشان می‌دهد، هرچند مقایسه روش‌های مختلف از جمله الگوریتم ژنتیک نیز پیشنهاد می‌شود. مدل‌های طبقه‌بندی آماری و شبکه عصبی مصنوعی با به کارگیری فقط ۳۵ ویژگی (۱۶ ویژگی شکل، ۱۱ ویژگی بافت و ۸ ویژگی رنگ) با صحت مناسبی تهیه شده‌اند. کمترین میانگین صحت طبقه‌بندی در تشخیص هفت گروه دانه با استفاده از داده‌های مستقل در مدل K- نزدیک‌ترین همسایگی (KNN) مشاهده شده است. مدل شبکه عصبی دو لایه (ANN2) قابلیت بیشتری در شرایط این مطالعه نشان داده است. دانه‌های گندم سالم و جو به دلیل تفاوت‌های ظاهری بیشتر با دیگر گروه‌ها با بیشترین

ونورا و همکاران (Venora et al., 2009)، با استفاده از تصویر نور عبوری از توده تک لایه در تشخیص و طبقه‌بندی دانه‌های گندم چروکیده به صحت ۹۵ درصد دست یافته‌اند البته تعداد کلاس‌های مورد نظر آنها کمتر و متفاوت از مطالعه حاضر بوده است. صحت طبقه‌بندی به دست آمده در این مطالعه، در مقایسه با نتایج بررسی‌های مجومدار و جایاس (Majumdar & Jayas, 2000a)، کمتر است. در بررسی‌هایی که این محققان به آنها پرداخته‌اند، فقط پنج گروه دانه غلات سالم بررسی شده است. آنها از دوربین فیلم‌برداری برای تصویربرداری استفاده کرده بودند؛ به همین دلیل به صحت طبقه‌بندی نسبتاً بالاتری دست یافته بودند. ولی در مقایسه با نتایج تحقیقات پالیوال و همکاران (Paliwal et al., 2003)، صحت طبقه‌بندی بالاتری در این مطالعه به دست آمده است. در مطالعات آنها، به دلیل وجود ۱۰ کلاس طبقه‌بندی، صحت طبقه‌بندی نسبتاً کم بوده است.

نتیجه‌گیری

از صحت طبقه‌بندی قابل قبول به دست آمده

تعیین ویژگی‌ها و مدل بهینه به منظور طبقه بندی اجزای ...

صحت و دانه‌های گندم سفید چروکیده به دلیل شباهت با دانه‌های دیگر با صحت کمتری قابل تشخیص بوده‌اند. برای طبقه‌بندی اجزای کیفی گندم بذری می‌توان گفت که دوربین دیجیتال صنعتی و روش طبقه‌بندی شبکه عصبی مصنوعی دو لایه (ANN2) یا روش‌های آماری، پتانسیل کافی را دارند. در صورت نیاز با آموزش گروه‌های جدید به سیستم، تعداد گروه‌های قابل تشخیص را می‌توان افزایش داد.

قدردانی

از بخش کنترل و گواهی بذر مرکز تحقیقات و آموزش کشاورزی و منابع طبیعی استان اردبیل و کارخانه‌های بوجاری بذر غلات به دلیل همکاری در جمع‌آوری نمونه‌های دانه قدردانی می‌شود.

مراجع

- Azizi, A., Abbaspour-Gilandeh, Y., Nooshyar, M., & Afkari-Sayah, A. (2015). Identifying potato varieties using machine vision and artificial neural networks. *International Journal of Food Properties*, 19, 618-635. doi:10.1080/10942912.2015.1038834.
- Chelladurai, V., Kaliramesh, S., Technology, P., & Jayas, D. S. (2012). *Detection of callosobruchus maculatus (F.) infestation in mung bean (Vigna Radiata) using thermal imaging technique. NABEC-CSBE/SCGAB 2012 Joint Meeting and Technical Conference.* July 15-18. Lakehead University, Orillia, Ontario.
- Chen, X., Xun, Y., Li, W., & Zhang, J. (2010). Combining discriminant analysis and neural networks for corn variety identification. *Computers and Electronics in Agriculture*, 71, 48-53. doi:10.1016/j.compag.2009.09.003.
- Delwiche, S. R., Yang, I. C., & Graybosch, R. A. (2013). Multiple view image analysis of free falling U. S. wheat grains for damage assessment. *Computers and Electronics in Agriculture*, 98, 62-73. doi:10.1016/j.compag.2013.07.002.
- Dubey, B. P. P., Bhagwat, S. G. G., Shouche, S. P. P., & Sainis, J. K. K. (2006). Potential of artificial neural networks in varietal identification using morphometry of wheat grains. *Biosystems Engineering*, 95, 61-67. doi:10.1016/j.biosystemseng.2006.06.001.
- Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7, 179-188.
- Gonzalez, R. C., Woods, R. E., & Eddins, S. L. (2009). *Digital Image Processing using MATLAB*. 3rd Ed. Gatesmark Publishing.
- Guevara-Hernandez, F., & Gomez-Gil, J. (2011). A machine vision system for classification of wheat and barley grain kernels. *Spanish Journal of Agricultural Research (SJAR)*, 9, 672-680. doi:10.5424/sjar/20110903-140-10.
- Guo, Y., Hastie, T., & Tibshirani, R. (2007). Regularized linear discriminant analysis and its application in microarrays. *Biostatistics*, 8, 86-100. doi:10.1093/biostatistics/kxj035.
- Khazaei, J., Goplour, I., & Ahmadi-Moghaddam, P. (2016). Evaluation of statistical and neural network architectures for the classification of paddy kernels using morphological features. *International Journal of Food Properties*, 19, 1227-1241.

doi:10.1080/10942912.2015.1071839.

- Mahesh, S., Manickavasagan, A., Jayas, D. S., Paliwal, J., & White, N. D. G. (2008). Feasibility of near-infrared hyperspectral imaging to differentiate Canadian wheat classes. *Biosystems Engineering*, 101, 50-57. doi:10.1016/j.biosystemseng.2008.05.017.
- Majumdar, S., & Jayas, D. S. (2000a). Classification of cereal grains using machine vision: Iv. combined morphology, color, and texture models. *Transactions of the ASAE*, 43, 1689-1694. doi:10.13031/2013.3069.
- Majumdar, S., & Jayas, D. S. (2000b). Classification of cereal grains using machine vision: I. Morphology Models. *Transactions of the ASAE*, 43, 1669-1675.
- Majumdar, S., & Jayas, D. S. (2000c). Classification of cereal grains using machine vision: III. Texture models. *Transactions of the ASAE*, 43, 1681-1687.
- Paliwal, J., Visen, N. S., Jayas, D. S., & White, N. D. G. (2003). Comparison of a neural network and a non-parametric classifier for grain kernel identification. *Biosystems Engineering*, 85, 405-413. doi:10.1016/S1537-5110(03)00083-7.
- Pazoki, A., Pazoki, Z., & Sorkhilalehloo, B. (2013). Rain fed barley seed cultivars identification using neural network and different neurons number. *World Applied Sciences Journal*. 22, 755-762. doi:10.5829/idosi.wasj.2013.22.05.2036.
- Pourreza, A., Pourreza, H., Abbaspour-Fard, M. H., & Sadrnia, H. (2012). Identification of nine Iranian wheat seed varieties by textural analysis with image processing. *Computers and Electronics in Agriculture*, 83, 102-108. doi:10.1016/j.compag.2012.02.005.
- Reunanen, J. (2003). Overfitting in making comparisons between variable selection methods. *The Journal of Machine Learning Research (JMLR)*, 3, 1371-1382. doi:10.1162/153244303322753715.
- Robnik-Šikonja, M., & Kononenko, I. (2003). Theoretical and empirical analysis of reliefF and RReliefF. *Machine Learning*, 53, 23-69. doi:10.1023/A:1025667309714.
- Savakar, D. (2012). Recognition and classification of similar looking food grain images using artificial neural networks. *International Journal of Applied Mathematics and Computer Science*, 13, 61-65.
- Shahin, M. A., & Symons, S. J. (2001). A machine vision system for grading lentils. *Canadian Biosystems Engineering / Le Genie des biosystems au Canada*, 43, 77-714.
- Silva, F. L., da, Grassi Sella, M. L., Francoy, T. M., & Costa, A. H. R. (2015). Evaluating classification and feature selection techniques for honeybee subspecies identification using wing images. *Computers and Electronics in Agriculture*, 114, 68-77. doi:10.1016/j.compag.2015.03.012.
- Szczypliński, P. M., Klepaczko, A., & Zapotoczny, P. (2015). Identifying barley varieties by computer vision. *Computers and Electronics in Agriculture*, 110, 1-8. doi:10.1016/j.compag.2014.09.016.
- Vanloot, P., Bertrand, D., Pinatel, C., Artaud, J., & Dupuy, N. (2014). Artificial vision and chemometrics analyses of olive stones for varietal identification of five French cultivars. *Computers and Electronics in Agriculture*, 102, 98-105. doi:10.1016/j.compag.2014.01.009.
- Venora, G., Grillo, O., & Saccone, R. (2009). Quality assessment of durum wheat storage centres in Sicily: Evaluation of vitreous, starchy and shrunken kernels using an image analysis

- system. *Journal of Cereal Science*, 49, 429-440. doi:10.1016/j.jcs.2008.12.006.
- Wang, N., Dowell, F. E., & Zhang, N. (2003). Determining wheat vitreousness using image processing and a neural network. *Transactions of the ASAE*, 46, 1143-1150.
- Xia, X., Fan, C., Lu, S. J., & Hou, L. L. (2010). *The analysis of wheat appearance quality based on digital image processing*. 2nd Conference on Environmental Science and Information Application Technology, July 17-18. Wuhan, China.
- Zapotoczny, P. (2011). Discrimination of wheat grain varieties using image analysis and neural networks. Part I. Single kernel texture. *Journal of Cereal Science*, 54, 60-68. doi:10.1016/j.jcs.2011.02.012.
- Zareian, A., Hasani, F., Sadegi, H., & Jazaery, M. R. (2010). *Wheat seed certification process and seed production instruction in Iran*. 2nd Conference on Seed Science and Technology. Oct. 26. Azad University. Mashad, Iran. (in Persian)



Determination of Features and Optimized Model to Classify the Components of Quality Wheat Seed using Industrial Digital Camera Image Processing

Z. Fazel-Niari, A. H. Afkari-Sayyah* and Y. Abbaspour-Gilandeh

*Corresponding Author: Associate Professor, Department of Biosystems Engineering, College of Agricultural Technology and Natural Resources, University of Mohaghegh Ardabili, Ardabil, Iran

Received: 23 January 2017, Accepted: 3 October 2017

Abstract

The acquisition of basic knowledge in quality control of wheat seed using machine vision technology is important. The objective of this research was to develop hardware and appropriate software to determine seven-grain groups in wheat seed samples. Ninety-one features were extracted through 21000 single seed images and the shape, texture and color features were ranked. Five classification models were investigated. The highest classification accuracy was obtained by artificial neural network with two hidden layers and the first 35 superior features. In the test run of this model with independent data, classifying accuracy for big white wheat, small white wheat, broken white wheat, wrinkled white wheat, red wheat, barley and rye were 100, 96.7, 99.3, 90.3, 99, 99.7, and 98 percent respectively with the average of 97.6 %. Shape features were more prominent and textural and color characteristics followed it respectively. Average classification accuracy in models of linear discriminant analysis, quadratic discriminant analysis, K- nearest neighbor and artificial neural network with a hidden layer were 95, 96.7, 91.6 and 97.3 % respectively. In the context of this study, the machine vision system comprising an industrial digital camera and artificial neural network with two hidden layers was identified as a valuable system in the investigation of the visual qualities of wheat seeds.

Keywords: Data Mining, Feature Reduction, Grain Classification, Quality Control, Seed Certification