یازدهمین کنگره ملی مهندسی مکانیک بیوسیستم و مکانیزاسیون ایران

شناسایی ارقام مختلف دانه لوبیا با استفاده از پردازش تصویر و الگوریتم جنگل تصادفی

حامد توکلی[1]، رضا محمدی گل[2]

[1]استادیار، گروه مهندسی مکانیک بیوسیستم، دانشگاه اراک؛ h-tavakoli@araku.ac.ir

[2]استادیار، گروه مهندسی مکانیک بیوسیستم، دانشگاه اراک؛ r-mohammadigol@araku.ac.ir

چکیده

در این تحقیق از پردازش تصویر به همراه الگوریتم طبقه‌بندی جنگل تصادفی، برای طبقه‌بندی ارقام مختلفی از دانه لوبیا استفاده شد. با استفاده از یک دوربین دیجیتال از ارقام مختلف لوبیای قرمز، سفید و چیتی تصویربرداری شد. سپس تصاویر مورد پردازش قرار گرفت و تعداد ۳۴ شاخص مورفولوژیکی از آنها استخراج شد و برای طبقه‌بندی استفاده شد. الگوریتم طبقه‌بندی جنگل تصادفی با تعداد ۱۰۰ تا ۵۰۰ درخت توسعه داده شد و بهترین نتیجه گزارش شد. بر اساس نتایج بدست آمده دقت طبقه‌بندی ۸۹، ۸۱، ۷۲ و ۷۹ درصد به ترتیب برای طبقه‌بندی دانه ارقام لوبیای قرمز، سفید، چیتی و کل ارقام بدست آمد. نتایج نشان داد روش تصویربرداری دیجیتال به همراه الگوریتم جنگل تصادفی می‌تواند ابزار قدرتمندی برای شناسایی و طبقه‌بندی محصولات کشاورزی فراهم آورد.

کلمات کلیدی: لوبیا، پردازش تصویر، طبقه‌بندی، جنگل تصادفی

Identification of different cultivars of bean using digital image processing and random forest

Hamed Tavakoli[1], Reza Mohammadi Gol[2]

[1]Assistant Professor, Department of Mechanical Engineering of Biosystems, Arak University; E-mail: h-tavakoli@araku.ac.ir

[2]Assistant Professor, Department of Mechanical Engineering of Biosystems, Arak University; E-mail: r-mohammadigol@araku.ac.ir

ABSTRACT

In this research, image processing along with Random Forest (RF) classification algorithm were used to classify different cultivars of been seeds. Using a digital camera, images were taken from different cultivars of red, white and pinto beans. Then the images were processed and 34 morphological indices were extracted and used for the classification. The RF algorithm was developed using 100 to 500 trees and the best results were reported. According to the results obtained, the classification accuracy was 89, 81, 72 and 79 percent, for red, white, pinto and all beans, respectively. The results showed that the imaging method along with the RF algorithm can provide a strong tool for identification and classification of agricultural products.

Keywords: Bean, Image Processing, Classification, Random Forest

یازدهمین کنگره ملی مهندسی مکانیک بیوسیستم و مکانیزاسیون ایران

[1] نویسنده مسئول: حامد توکلی؛ آدرس: اراک-خیابان شهید بهشتی-دانشگاه اراک-گروه مهندسی مکانیک بیوسیستم. شماره تماس: ۰۸۶۳۲۶۲۳۸۰۰

۱- مقدمه

بینایی انسان یک حس تشخیصی و اکتشافی مهم است. در تحقیقات و همچنین در صنعت تحلیل‌ها و ارزیابی‌های زیادی به این وسیله انجام می‌گیرد. اما تشخیص انسان به راحتی می‌تواند دچار اشتباه شود؛ همچنین با سرعت کمی انجام شده و هزینه‌بر است. اشکال دیگر تشخیص انسان این است که استاندارد سازی نتایج بدست آمده دشوار است. بنابراین نیاز به استفاده از سامانه‌های اندازه‌گیری و تشخیصی مکانیزه جایگزین انسان وجود دارد. علوم کامپیوتر و الکترونیک امکان جایگزینی بینایی انسان را با وسایل مکانیزه بوجود آورده‌اند. در این راستا، تحقیقات زیادی به منظور بکارگیری روش‌های ماشین بینایی (Machine Vision) در صنایع مختلف صورت گرفته است (Venora et al., 2007).

در حوزه کشاورزی نیز از بینایی انسان برای کاربردهایی استفاده می‌شود که مکانیزه کردن آنها می‌تواند باعث کاهش خطا و هزینه‌ها و افزایش سرعت انجام فرآیندها شود. پردازش تصاویر دیجیتال در تحقیقات برای کاربردهای مختلفی در حوزه کشاورزی بکار رفته است. شناسایی علف هرز (Tellaeche et al., 2011)، تشخیص بیماری و آفت در محصول (Leiva-Valenzuela and Aguilera, 2013)، تشخیص کمبود مواد غذایی در محصول (Wang et al., 2013)، تخمین عملکرد محصول (Zaman et al., 2008) و شناسایی ارقام محصول (Venora et al., 2009a) نمونه‌هایی از این کاربردها هستند.

لوبیا (*Phaseolus vulgaris L.*) یکی از منابع مهم پروتئین و تولید انرژی برای انسان می‌باشد. ارقام مختلفی از این محصول موجود می‌باشد. شناسایی ارقام لوبیا، یکی از عملیات‌هایی است که توسط کارشناسان مجرب صورت می‌گیرد. با توجه به تنوع ارقام لوبیا، شناسایی ارقام مختلف برای کارشناسان و کشاورزان دشوار است. همچنین جداسازی ارقام مخلوط شده بصورت دستی، وقت‌گیر و هزینه‌بر است. در صورت توسعه روشی مکانیزه برای انجام این کار، تشخیص ارقام لوبیا با سهولت و دقت بیشتری قابل انجام خواهد بود. الگوریتم‌های آنالیز تصویر که به روش‌های طبقه‌بندی مجهز شده‌اند، به نظر می‌رسد روشی مفید و مناسب برای شناسایی اتوماتیک محصولات کشاورزی ارائه می‌دهند (Granitto et al., 2003; Bacchetta et al., 2008; Venora et al., 2009b).

محققان زیادی از آنالیز تصاویر دیجیتال برای شناسایی دانه غلات (Carter et al., 2006; Choudhary et al., 2008; Zapotoczny, 2011) و دانه بقولات استفاده کرده‌اند (Kılıç et al., 2007; Venora et al., 2007; Venora et al., 2009a).

در این تحقیق نیز امکان استفاده از ماشین بینایی و تصویربرداری دیجیتال برای تشخیص ارقام مختلف لوبیا مورد بررسی قرار می‌گیرد. شاخص‌های مورفولوژیکی از تصاویر دانه‌های ارقام لوبیا استخراج شد و برای طبقه‌بندی از الگوریتم جنگل تصادفی[1] (RF) (Wold et al., 1983) استفاده شد. در صورت امکان تشخیص دانه‌های ارقام لوبیا با استفاده از این روش، می‌توان برای فرآیندهایی مانند درجه‌بندی لوبیا، جداسازی ارقام مخلوط شده و یا جداسازی مواد خارجی موجود در لوبیا نیز از این روش بطور اتوماتیک استفاده کرد.

۲- مواد و روش‌ها

در این تحقیق ارقام مختلفی از لوبیای قرمز (گلی، ناز، اختر و درخشان)، لوبیای سفید (دانشکده، درسا، پاک و شکوفا) و لوبیای چیتی (محلی خمین، صدری، کوشا، cos16 و لاین 682) از مرکز ملی تحقیقات لوبیا خمین تهیه شده و مورد استفاده قرار گرفت.

با استفاده از یک دوربین دیجیتال کانن[2] مدل PowerShot SX230 HS که دارای تفکیک‌پذیری ۱۲ مگاپیکسل است، تصاویر رنگی (RGB[3]) از دانه‌های ارقام لوبیا تهیه گردید. عکس‌ها با قرار دادن دوربین در حالت عکسبرداری Program تهیه شدند. دوربین در حالت تنظیم اتوماتیک سرعت شاتر[4] و ایزو[5] قرار داده شد. تصاویر به فرمت JPEG ثبت شده و برای پردازش با کمک نرم افزار متلب به کامپیوتر شخصی منتقل گردیدند.

نمونه‌ها در داخل یک جعبه در دار که داخل آن نورپردازی شده بود، قرار داده شد و از آنها عکسبرداری صورت گرفت. جهت تسهیل در امر جداسازی دانه از زمینه، از زمینه‌های مصنوعی (کاغذ آبی رنگ) در حین عکسبرداری استفاده شد.

برای هر رقم، تعداد ۱۰۰ دانه بطور تصادفی انتخاب شد. دانه‌ها در جهتی که روی زمین آزادانه استقرار پیدا می‌کردند، قرار داده شدند و از آنها عکسبرداری شد. دو بعد عمود بر هم دانه‌ها (در حالتی که روی زمین آزادانه قرار می‌گرفتند) با استفاده از یک کولیس با دقت ۰/۰۱ میلی‌متر

[1] Random Forest
[2] Canon
[3] Red-Green-Blue
[4] Shutter speed
[5] ISO

اندازه‌گیری شد. بعد بزرگتر، بعنوان طول دانه و بعد کوچکتر بعنوان عرض دانه ثبت گردید.

برای استخراج شاخص‌های مورفولوژیکی، تصاویر تهیه شده به کمک نرم افزار متلب (Matlab) (R2014b, Mathworks Company, U.S.) پردازش شدند. دوربین دیجیتال مورد استفاده تصاویر را در محدوده مرئی طیف نور با استفاده از کانال‌های[1] قرمز، سبز و آبی ذخیره می‌کرد. هر کانال دارای ۸ بیت است (۲۵۶ سطح شدت نور). از اختلاف کانال‌های رنگی بین دانه و پیش‌زمینه، برای جدا کردن محصول از پیش‌زمینه استفاده شد. همچنین از یک برگ کاغذ مشبک با ابعاد مشخص برای کالیبره کردن تعداد پیکسل‌ها با ابعاد واقعی استفاده شد.

شکل دانه‌ها بصورت بیضی در نظر گرفته شد و از روابط بیضی برای محاسبه‌ی خصوصیات هندسی بر اساس اندازه‌گیری مستقیم ابعاد دانه استفاده شد. شاخص‌های هندسی محاسبه شده با استفاده از اندازه‌گیری‌های مستقیم و همچنین با استفاده از پردازش تصاویر در جدول۱ آورده شده است.

جدول۱- شاخص‌های هندسی استخراج شده از تصاویر ارقام لوبیا

Table 1. The geometric indices extracted from the images of bean cultivars

رابطه شاخص*	رابطه شاخص
$Pe=\frac{\pi}{2}(L+W)$	$PA=\frac{\pi}{4}(L*W)$
$D=\sqrt{(4*PA)/\pi}$	$Sh1=(4\pi*PA)/(Pe^2)$
$Sh2=Pe^2/PA$	$Sh3=L/W$
$Sh4=2*\sqrt{PA/\pi}$	$Sh5=Pe/\pi$
$Sh6=Pe/(2*\sqrt{\pi*PA})$	$Sh7=L/PA$
$Sh8=PA/L^3$	$Sh9=(4*PA)/(\pi*L^2)$
$Sh10=(4*PA)/(\pi*L*W)$	$Sh11=(Pe-Pe*\sqrt{Pe^2-4\pi*PA})/(Pe+Pe*\sqrt{Pe^2-4\pi*PA})$

*L: طول دانه (میلی‌متر)، W: عرض دانه (میلی‌متر)، Pe: محیط سطح تصویر شده دانه (میلی‌متر)، PA: مساحت سطح تصویر شده دانه (میلی‌متر مربع)، D: قطر معادل سطح تصویر شده دانه (میلی‌متر)، $Sh1$ تا $Sh11$: شاخص‌های شکل دانه

همبستگی بسیار بالایی بین مقادیر محاسبه شده‌ی این شاخص‌ها با استفاده از داده‌هایی که با اندازه‌گیری کولیس بدست آمده بود، و داده‌های حاصل از پردازش تصویر وجود داشت (r=0.84–0.98).

شاخص‌های دیگری بر اساس رنگ دانه‌های لوبیا استخراج شد که در جدول۲ آورده شده است.

[1]Channel

یازدهمین کنگره ملی مهندسی مکانیک بیوسیستم (ماشین‌های کشاورزی) و مکانیزاسیون ایران

جدول۲- شاخص‌های رنگی استخراج شده از تصاویر ارقام لوبیا

Table 2. The color indices extracted from the images of bean cultivars

رابطه شاخص*	رابطه شاخص
$Rm = mean\ of\ R$	$Gm = mean\ of\ G$
$Bm = mean\ of\ B$	$r = Rm / (Rm + Gm + Bm)$
$g = Gm / (Rm + Gm + Bm)$	$b = Bm / (Rm + Gm + Bm)$
$GMR = Gm - Rm$	$GMB = Gm - Bm$
$RMB = Rm - Bm$	$NGMR = (Gm - Rm) / (Gm + Rm)$
$NGMB = (Gm - Bm) / (Gm + Bm)$	$NRMB = (Rm - Bm) / (Rm + Bm)$
$\text{Hue} = \begin{cases} \theta & if\ B \le G \\ 360 - \theta & if\ B > G \end{cases}$ $\theta = cos^{-1}\left\{\frac{\frac{1}{2}[(R-G)+(R-B)]}{[(R-G)^2+(R-B)(G-B)]^{1/2}}\right\}$	$\text{SAT} = 1 - \frac{3}{(R+G+B)}[min(R,G,B)]$
$\text{INT} = \frac{1}{3}(R+G+B)$	$Y = 0.257*R + 0.504*G + 0.098*B + 16$
$Cb = -0.148*R - 0.291*G + 0.439*B + 128$	$Cr = 0.439*R - 0.368*G - 0.071*B + 128$

*R، G و B به ترتیب سطوح شدت نور طیف‌های قرمز، سبز و آبی هستند.

۳۲ شاخص فوق به همراه طول و عرض دانه‌ها (در مجموع ۳۴ شاخص) برای طبقه‌بندی استفاده شد.

از آنجائیکه این تحقیق شامل چندین متغیر پیشگو (ویژگی‌های مستخرج از تصاویر) بود که با یکدیگر همبستگی نیز دارند، استفاده از الگوریتم طبقه‌بندی جنگل تصادفی (RF) می‌تواند انتخاب مناسبی باشد.

روش RF در واقع توسعه یافته‌ی روش درخت تصمیم است. RF شامل ترکیبی از تعدادی درخت (یعنی جنگل) است که در آن برای هر درخت، زیر مجموعه‌ای تصادفی از متغیرهای پیشگو (در طبقه‌بندی جنگل تصادفی، جزء صحیح ریشه دوم تعداد کل متغیرهای پیشگو (Hastie et al., 2013)) استفاده می‌شود. بعلاوه، هر درخت با استفاده از بخشی از نمونه‌ها (حدود دو سوم کل نمونه‌ها) ساخته می‌شود. بقیه نمونه‌ها بعنوان مجموعه تست استفاده می‌شود که نمونه‌های خارج از کیسه (OOB) نامیده می‌شوند. هر درخت تصمیم تا رسیدن به حداقل تعداد گره از پیش تعریف شده‌ای (در مسائل طبقه‌بندی حداقل مقدار استاندارد یک گره است(Hastie et al., 2013)) توسعه می‌یابد.

با استفاده از پیش‌بینی‌های داده‌های OOB هر درخت، دقت طبقه‌بندی الگوریتم بر حسب درصد محاسبه شد. نتایجی که توسط نمونه‌های OOB بدست می‌آید تقریباً معادل آنچه توسط روش اعتبارسنجی K-fold cross-validation حاصل می‌شود، می‌باشد (Hastie et al., 2013).

در این تحقیق، الگوریتم طبقه‌بندی با تعداد ۱۰۰ تا ۵۰۰ درخت توسعه داده شد و بهترین نتیجه انتخاب و گزارش گردید. هر مدل نیز ۱۰ بار تکرار شد و میانگین مقادیر دقت طبقه‌بندی گزارش شد.

آنالیز داده‌ها با استفاده از نرم افزار متلب (Version 9.1, R2016b, Mathworks Company) انجام شد.

۳- نتایج و بحث

برای طبقه‌بندی ارقام دانه لوبیا، ابتدا ارقام لوبیای قرمز، سفید و چیتی بطور جداگانه طبقه‌بندی شدند و سپس همه ارقام با هم مورد استفاده قرار گرفتند. در جدول۳ دقت بدست آمده از الگوریتم جنگل تصادفی در طبقه‌بندی ارقام دانه لوبیا نشان داده شده است.

Out-Of-Bag samples

جدول۳- دقت الگوریتم جنگل تصادفی در طبقه‌بندی ارقام مختلف لوبیا

Table 3. The accuracy of Random Forest algorithm for classifying different cultivars of bean

	تعداد درخت	دقت طبقه‌بندی (٪)
لوبیای قرمز	۱۱۱	۸۹
لوبیای سفید	۳۹۰	۸۱
لوبیای چیتی	۲۵۳	۷۲
کل ارقام	۱۹۶	۷۹

همانطور که در جدول۳ ارائه شده است، دقت طبقه‌بندی برای لوبیای قرمز، سفید، چیتی و کل ارقام به ترتیب ۸۹، ۸۱، ۷۲ و ۷۹ درصد بدست آمد. با توجه به تعداد زیاد ارقام مورد استفاده، نتایج بدست آمده قابل قبول می‌باشد. البته در صورت افزایش تعداد نمونه‌ها، نتایج خیلی بهتری با استفاده از الگوریتم جنگل تصادفی قابل حصول است (Loosvelt et al. 2012).

نتایج بدست آمده قابل رقابت با نتایج سایر محققان بود. Kılıç et al. (2007) دقت طبقه‌بندی ۹۰/۶ درصد را با استفاده از تصویربرداری و شبکه‌های عصبی، برای لوبیا گزارش کرده‌اند.

۴- نتیجه‌گیری

با توجه به نتایج بدست آمده می‌توان نتیجه‌گیری کرد که روش تصویربرداری دیجیتال به همراه الگوریتم جنگل تصادفی می‌تواند ابزار قدرتمندی برای شناسایی و طبقه‌بندی محصولات کشاورزی فراهم آورد.

۵- مراجع

Bacchetta, G., Grillo, O., Mattana, E., & Venora, G. (2008). Morpho-colorimetric characterization by image analysis to identify diaspores of wild plant species. *Flora, 203*, 669–682.

Carter, R. M., Yan, Y., & Tomlins, K. (2006). Digital imaging based classification and authentication of granular food products. *Measurement Science and Technology, 17*(2), 235–240.

Choudhary, R., Paliwal, J., & Jayas, D. S. (2008). Classification of cereal grains using wavelet, morphological, colour, and textural features of non-touching kernel images. *Biosystems Engineering, 99*, 330–337.

Granito, P. M., Garralda, P. A., Verdes, P. F., & Ceccato, H. A. (2003). Boosting classifiers for weed seeds identification. *Journal Cereal Science and Technology, 3*(1), 34–39.

Hastie, T., Tibshirani, R., & Friedman, J. (2013). *The Elements of Statistical Learning*. Second Edition. Springer-Verlag New York.

Kılıç, K, Boyacı, İ. H., Köksel, H., & Küsmenoğlu, İ. (2007). A classification system for beans using computer vision system and artificial neural networks. *Journal of Food Engineering, 78*, 897–904.

Leiva-Valenzuela, G. A., & Aguilera, J. M. (2013). Automatic detection of orientation and diseases in blueberries using image analysis to improve their postharvest storage quality. *Food Control, 33*, 166-173.

Loosvelt, L., Peters, J., Skriver, H., Lievens, H., Van Coillie, F. M. B., De Baets, B., & Verhoest, N. E. C. (2012). Random forests as a tool for estimating uncertainty at pixel-level in SAR image classification. *International Journal of Applied Earth Observation and Geoinformation, 19*, 173–184.

Tellaechea, A., Pajaresa, G., Burgos-Artizzub, X. P., & Ribeirob, A. (2011). A computer vision approach for weeds identification through Support Vector Machines. *Applied Soft Computing, 11*, 908–915.

Venora, G., Grillo, O., Ravalli, C., & Cremonini, R. (2009a). Identification of Italian landraces of bean (Phaseolus vulgaris L.) using an image analysis system. *Scientia Horticulturae, 121*, 410–418.

Venora, G., Grillo, O., & Saccone, R. (2009b). 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*(3), 429–440.

Venora, G., Grillo, O., Shahin, M. A., & Symons, S. J. (2007). Identification of Sicilian landraces and Canadian cultivars of lentil using an image analysis system. *Food Research International 40*, 161–166.

Wang, Y., Wang, D., Zhang, G., & Wang, J. (2013). Estimating nitrogen status of rice using the image segmentation of G-R thresholding method. *Field Crops Research, 149*, 33–39.

Wold, S., Martens, H., & Wold, H. (1983). The multivariate calibration problem in chemistry solved by the PLS method. In: Kågström, B., Ruhe, A. (Eds.), Matrix Pencils, Lecture Notes in Mathematics, vol. 973. Springer, Berlin Heidelberg, pp. 286–293. http://dx.doi.org/10.1007/BFb0062108.

Zaman, Q. U., Schumann, A. W., Percival, D. C., & Gordon, R. J. (2008). Estimation of Wild Blueberry Fruit Yield Using Digital Color Photography. *Transactions of the ASABE, 51*(5), 1539–1544.

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.

یازدهمین کنگره ملی مهندسی مکانیک بیوسیستم (ماشین‌های کشاورزی) و مکانیزاسیون ایران