



Identification and Localization of Chickpea Impurities Using SVM and KNN Classifiers

Hossein Bagherpour¹ | Siavash Shamohammadi²

1. Corresponding Author, Department of Biosystems Engineering, Faculty of Agriculture, Bu-Ali Sina University, Hamedan, Iran. E-mail: H.bagherpour@basu.ac.ir
2. Department of Biosystems Engineering, Faculty of Agriculture, Bu-Ali Sina University, Hamedan, Iran. E-mail: s.shamohammadi@agr.basu.ac.ir

Article Info

Article type: Research Article

Article history:

Received: Sep. 27, 2025

Revised: Dec. 12, 2025

Accepted: Dec. 24, 2025

Published: Spring. 2026

Keywords:

Classification,
Image processing,
Beans,
Pea impurities

During chickpea harvesting, various types of impurities are present in the product, which must be identified and removed before market distribution or use as seed. Although pneumatic and mechanical methods can eliminate a substantial portion of these impurities, conventional techniques are insufficient for separating objects such as small stones of similar size to chickpeas or unripe and discolored grains. The objective of this study was to identify the type and determine the location of different chickpea impurities using two intelligent classifiers: Support Vector Machine (SVM) and k-Nearest Neighbors (KNN). For this purpose, 400 RGB images were acquired, encompassing six classes: healthy, green, black, colored, stones, and split chickpeas. After object segmentation and classification into six groups, the total number of samples reached 3,840. Features extracted included mean, median, variance, skewness, histogram, entropy, and texture descriptors derived from the gray-level co-occurrence matrix (GLCM), such as contrast, correlation, energy, and homogeneity. In the SVM model, the RBF kernel exhibited superior performance compared to other kernels. For KNN, the optimal results were obtained with $k = 13$, the City Block distance metric, and a weighting scheme of $1/(c + D^2)$ with $c = 1$. Object localization was performed in MATLAB by determining the coordinates of each object's center. Based on the results, the highest classification accuracy for the SVM and KNN models at a resolution of 250×250 pixels were 98.09% and 90.88%, respectively.

Cite this article: Bagherpour, H., Shamohammadi, S., (2026) Identification and Localization of Chickpea Impurities Using SVM and KNN Classifiers, *Iranian Journal of Biosystem Engineering*, 57 (1),1-15.
<https://doi.org/10.22059/ijbse.2025.403138.665619>

© The Author(s).

Publisher: University of Tehran Press.



DOI: <https://doi.org/10.22059/ijbse.2025.403138.665619>



EXTENDED ABSTRACT

Introduction

Chickpeas are a high-protein legume whose consumption has increased worldwide. Ensuring their quality requires the removal of foreign impurities and uniformity in size, shape, and color. Manual separation is labor-intensive, making automated systems necessary for efficient and accurate impurity detection. Machine vision has shown great potential for identifying impurities and classifying varieties in small-seeded legumes. However, the effective deployment of such systems requires crop-specific algorithms tailored to local cultivars. This study aimed to develop an algorithm for native Iranian chickpea varieties that can accurately detect and classify impurities while determining their spatial locations—a factor largely unaddressed in previous research. The ultimate goal is to reduce reliance on manual labor and enhance the performance of automatic grading machines.

Methods

Images were pre-processed to remove background noise and isolate the primary objects. Various color spaces (HSV, Lab, YCbCr) were evaluated, and YCbCr provided the best discrimination between objects and background. Following conversion from RGB to YCbCr, thresholds in the Cr channel were tested, and $Cr > 105$ was selected for effective segmentation. Small particles were removed using the `bwareaopen` function, and final binary masks were multiplied with the original images to produce segmented outputs. Each object was assigned to one of six classes: healthy, green, black, colored, stone, and split chickpeas, resulting in 3,840 images (640 per class). Feature extraction combined color, texture, and geometric descriptors. Texture features (correlation, energy, homogeneity, entropy) were derived from grayscale images using the gray-level co-occurrence matrix (GLCM) at four angles (0° , 45° , 90° , 135°) and pixel distances of 1–10, yielding 160 features. Statistical descriptors (kurtosis, skewness, maximum, median, mean, variance) were computed for RGB channels (18 features) and grayscale images (5 features), along with perimeter-to-area ratio and entropy (2 features), totaling 185 features. Seventy percent of the dataset was used for training and 30% for testing. All features were normalized before classification. To identify and classify pea impurities, two well-known machine learning models—Support Vector Machine (SVM) and K-Nearest Neighbors (KNN)—were employed, allowing a comparative evaluation of their performance in distinguishing between the six classes.

Results

The SVM classifier demonstrated high stability and reliability due to its convergence to a global minimum. Analysis of the cost parameter C indicated that the optimal performance was achieved at $C = 10$. Furthermore, the RBF kernel outperformed both linear and polynomial kernels, yielding approximately 14.9% and 13.8% higher accuracy, respectively. For the KNN classifier, k values from 3 to 21 were examined, and the best performance (90.79% accuracy) was obtained using $k = 13$, the City Block distance metric, and the weighted scheme $1/(c + D^2)$. Evaluation of distance metrics showed that weighted approaches performed better than simple distance measures. Investigating the effect of image size revealed that increasing the resolution did not significantly improve classification accuracy, and the 250×250 resolution provided the optimal trade-off between speed and accuracy. Confusion matrix results indicated that the black chickpea class exhibited the highest separability, while misclassifications mainly occurred in classes with similar characteristics. Overall comparison demonstrated that SVM achieved superior performance, with an accuracy of 98.09%, outperforming KNN, which achieved 90.88%. In the SVM model, the F1-score for all classes—except the stone class—exceeded 97%. The slightly lower performance for the stone class was attributed to its visual similarity to certain chickpea types. Comparison with previous studies showed that the improved SVM model increased accuracy by approximately 6.8%. Additionally, object localization and bounding box generation were successfully accomplished, with an average processing time of 25.6 ms per image, which can be further reduced using more advanced hardware.

Conclusion

In this study, two well-known classifiers—Support Vector Machine (SVM) and K-Nearest Neighbors (KNN)—were used to identify and differentiate healthy chickpea seeds from various impurities. SVM outperformed KNN, with KNN performance influenced by the number of neighbors, distance metrics, and weighting schemes, while SVM accuracy depended on the kernel function. A key feature of this study is the inclusion of classes with highly similar characteristics, improving discrimination of subtle differences and enabling high-purity chickpea batches. Although common Iranian varieties and typical impurities were used, variations in color, texture, or shape in other cultivars may limit generalizability. Future studies should include a broader range of cultivars and impurities to enhance robustness. With the growing adoption of robotic and automated systems, the algorithm can be integrated into automatic grading machines for detecting both the type and spatial location of impurities. Further testing in laser-based grading systems is recommended to assess industrial performance.

Author Contributions

Both authors contributed equally to the research and the preparation of the manuscript, under the guidance and supervision of the corresponding author.

Data Availability Statement

Data available on request from the authors

Acknowledgments

The authors would like to express their sincere gratitude to Bu-Ali Sina University for providing the facilities and support necessary to conduct this research.

Ethical considerations

The authors avoided data fabrication, falsification, plagiarism, and misconduct.

Conflict of interest

The author declares no conflict of interest

شناسایی و تعیین موقعیت مکانی ناخالصی‌های نخود با استفاده کلاسبندهای SVM و KNN

حسین باقرپور^۱ | سیاوش شامحمدی^۲

۱. نویسنده مسئول، گروه مهندسی بیوسیستم، دانشکده کشاورزی، دانشگاه بوعلی سینا، همدان، ایران. رایانامه:

h.bagherpour@basu.ac.ir

۲. گروه مهندسی بیوسیستم، دانشکده کشاورزی، دانشگاه بوعلی سینا، همدان، ایران. رایانامه:

s.shamohammadi@agr.basu.ac.ir

اطلاعات مقاله

چکیده

نوع مقاله: مقاله پژوهشی

تاریخ دریافت: ۱۴۰۴/۷/۵

تاریخ بازنگری: ۱۴۰۴/۹/۲۱

تاریخ پذیرش: ۱۴۰۴/۱۰/۳

تاریخ انتشار: بهار ۱۴۰۵

واژه‌های کلیدی:

درجه بندی،

پردازش تصویر،

حبوبات،

ناخالصی نخود

در زمان برداشت نخود، انواع مختلفی از ناخالصی‌ها در محصول وجود دارد که لازم است پیش از عرضه به بازار، شناسایی و جداسازی شوند. اگرچه بخش زیادی از این ناخالصی‌های براحتی قابل حذف هستند، اما جداسازی مواردی مانند سنگ‌ریزه‌های هم‌اندازه نخود یا نخودهای نارس و بدرنگ با روش‌های مرسوم امکان‌پذیر نیست. هدف این پژوهش، تشخیص نوع و تعیین موقعیت ناخالصی‌های مختلف نخود با استفاده از دو مدل هوشمند ماشین‌بردار پشتیبان (SVM) و K نزدیک‌ترین همسایه (KNN) است. برای این منظور، ۴۰۰ تصویر RGB تهیه شد که هر کدام از تصاویر شامل شش کلاس نخود سالم، سبز، سیاه، رنگی، سنگ و لپه بودند. برای شناسایی نوع کلاس هر کدام از اشیای موجود در تصویر و استخراج ویژگی‌ها، بعد از تعیین موقعیت مکانی، هر یک از ۶ کلاس از تصاویر اصلی جدا گردیدند و به صورت مجزا در ۶ دسته مختلف طبقه‌بندی شدند. با این عملیات، در مجموع کل تعداد تصاویر اشیای به ۳۸۴۰ رسید. ویژگی‌هایی شامل میانگین، میانه، واریانس، چولگی، هیستوگرام، آنتروپی و نیز ویژگی‌های بافتی حاصل از ماتریس هم‌وقوع سطح خاکستری شامل کنتراست، همبستگی، انرژی و همگنی استخراج شد. در مدل SVM، تابع RBF بهترین عملکرد را در مقایسه با توابع دیگر نشان داد. در مدل KNN نیز بهترین نتایج با $k=13$ ، معیار فاصله City Block و وزن‌دهی $(c+D^2)/1$ با $c=1$ حاصل شد. تعیین موقعیت مکانی اشیای بر اساس مختصات مرکز آن‌ها در محیط MATLAB انجام گرفت. بر اساس نتایج، بیشترین دقت مدل‌های SVM و KNN در رزولوشن 250×250 به ترتیب برابر با $98/09$ و $90/88$ درصد به‌دست آمد.

استناد: باقرپور؛ حسین، شامحمدی؛ سیاوش، (۱۴۰۵) شناسایی و تعیین موقعیت مکانی ناخالصی‌های نخود با استفاده کلاسبندهای SVM و KNN،

مجله مهندسی بیوسیستم ایران، ۵۷ (۱)، ۱۵-۱. <https://doi.org/10.22059/ijbse.2025.403138.665619>

© نویسندگان.

ناشر: مؤسسه انتشارات دانشگاه تهران.

DOI: <https://doi.org/10.22059/ijbse.2025.403138.665619>

مقدمه

وجود درصد بالای پروتئین در نخود باعث مصرف روزافزون آن برای عموم مردم دنیا شده است. نخود به دلیل توانایی آن در تثبیت نیتروژن اتمسفر و تحمل آن در برابر عوامل استرس‌زای محیطی از جمله خشکی، گرما، شوری و خاک‌های کم‌نیتروژن برای کشاورزان مهم است (Bazrafkan et al., 2023). جایگزینی جزئی برای پروتئین‌های حیوانی با پروتئین‌های گیاهی می‌تواند تقاضای فزاینده را در چارچوب توسعه پایدار سیستم‌های کشاورزی و غذایی برآورده کند (Voisin et al., 2014). حدود ۷۵ درصد نخود برای مصارف انسانی و مابقی نیز برای مصارف دام و تولید بذر بکار می‌رود (Kanouni et al., 2018).

نبود ناخالصی‌های خارجی در نخود و یکنواختی آن از لحاظ اندازه، شکل و رنگ لازمه بکارگیری سیستم خودکاری است که بتواند کار شناسایی را با صحت و سرعت مناسب انجام دهد. استفاده از بینایی ماشین راهکار مناسبی برای تفکیک و شناسایی ناخالصی‌ها و سایر ارقام در حبوبات دانه‌ریز مانند نخود و سایر محصولات کشاورزی است، به طوری که نتایج به دست آمده از محققان مختلف نشان از توانایی بالای این تکنیک در شناسایی واریته‌ها و ناخالصی‌های مختلف در محصولات دانه‌ریز است. از طرفی یکی از چالش‌هایی که هنوز بسیاری از دستگاه‌های درجه‌بندی مدرن با آن مواجه هستند، وجود خطا در شناسایی ناخالصی‌هایی است که از نظر ویژگی بسیار شبیه نخود سالم هستند و در صورت عدم شناسایی باعث کاهش ارزش نهایی محصول مخصوصاً برای صادرات می‌گردد (Salam et al., 2022).

با بررسی پیشینه تحقیق مطالعات متعددی در ارتباط با موضوع حاضر یافت شد. در تحقیقی که صرفاً بر اساس شکل‌های مختلف نخود، طبقه‌بندی آن‌ها انجام شد، با استفاده از ۱۴ ویژگی موثر، رنگ، بافت و شکل، طبقه‌بندی انواع نخود (مطلوب، سبز، قهوه‌ای، لپه و چروکیده) و ناخالصی‌های آن (سنگ و ساقه) به کمک شبکه عصبی صحت ۹۱/۹ درصد حاصل شد (سلام و خیرعلی‌پور، ۱۴۰۰). خطای بالای این مدل، می‌تواند یکی از ایرادات اصلی آن در کاربردهای صنعتی باشد. از طرفی گزارشی مبنی بر سرعت تشخیص و امکان بکارگیری آن در تشخیص‌های سریع یافت نشد. با استفاده از بینایی ماشین و شبکه عصبی مصنوعی، پنج کلاس لوبیا به دسته‌های سفید (سالم)، زرد-سبز (آسیب دیده)، سیاه (آسیب دیده)، کم‌آسیب دیده و بسیار آسیب دیده طبقه‌بندی شدند (Kılıç et al., 2007). این مطالعه از یک مجموعه داده نسبتاً کوچک و محدود به شرایط آزمایشگاهی استفاده کرده است که تعمیم‌پذیری مدل به شرایط واقعی‌تر را کاهش می‌دهد. همچنین با روش‌های یادگیری جدیدتر مقایسه نشده‌اند و عملکرد مدل در برابر سایر شبکه‌ها نامشخص است. با استفاده از شبکه یادگیری انتقالی VGG16 تعمیم‌یافته دوازده نوع بادام زمینی با میانگین صحت ۹۶/۷ درصد طبقه‌بندی شد (Yang et al., 2021). اگرچه VGG16 عملکرد خوبی داشته است، اما مدل بسیار سنگین است و مطالعه مقایسه‌ای با معماری‌های سبک‌تر و مدرن‌تر انجام نشده است. علاوه بر این، تنوع محدود نمونه‌های بادام‌زمینی می‌تواند موجب سوگیری و کاهش پایداری مدل در داده‌های واقعی شود. با استفاده از طول موج بهینه و کلاسبند KNN شناسایی پوست گردو از مغز آن انجام شد (Zhu et al., 2007). در این پژوهش تنها بر یک ویژگی طیفی و یک کلاسبند ساده (KNN) تکیه شده است که ممکن است در برابر نویز، تغییرات محیطی و تفاوت‌های بین نمونه‌های عملکرد پایداری نداشته باشد. بالاترین صحت در تفکیک چهار کلاس مختلف برنج آسیاب شده شامل دانه‌های کم‌فرآوری شده، دانه‌های شکسته کم‌فرآوری شده، و دانه‌های شکسته با فرآوری بالا با شبکه MLP و صحت ۹۸/۷۲ درصد بدست آمد. پس از MLP، به ترتیب ماشین‌های بردار پشتیبان، درخت تصمیم و شبکه بیزین دارای صحت بالاتری بودند (Zareiforoush et al., 2016). انواع مختلف سویا و همچنین دانه‌های سالم در مقابل ناسالم با استفاده از پردازش تصویر شناسایی شد (Liu et al., 2015). این تحقیق عمدتاً به ویژگی‌های هندسی و رنگی سنتی متکی است که در شرایط نوری و محیطی متفاوت ممکن است پایداری لازم را نسبت به ویژگی‌های بافتی نداشته باشند. سه رقم برنج از هم و از دانه‌های شکسته دیگر (چهار رقم) به کمک ویژگی‌های هندسی با استفاده از چندین الگوریتم یادگیری ماشین کلاسبندی شدند که الگوریتم نزدیکترین همسایه با میانگین صحت ۹۰/۵ درصد بهتر از دیگر الگوریتم‌ها شناخته شد (Ozan et al., 2015). با استفاده از تصویربرداری لیزری چهار مخلوط نمونه از سیب‌زمینی، کلوخ و سنگ برای انتخاب ویژگی استفاده شد. از ماشین بردار پشتیبان (SVM)، فاصله ماله‌انوبیس (MDD) و LDA برای جداسازی نمونه‌ها به دو دسته (با توجه به پنج ویژگی انتخاب‌شده) استفاده شد (Geng et al., 2021). اگرچه تصویربرداری لیزری دقیق است و می‌تواند صحت قابل قبولی را ارائه دهد، اما هزینه و پیچیدگی سخت‌افزاری این روش موجب محدودیت کاربرد صنعتی آن می‌شود.

سیستم‌های تجاری مدرن می‌توانند محصولات را بر اساس پارامترهای فیزیکی خارجی و عیوب خارجی با استفاده از سیستم بینایی ماشین طبقه‌بندی کنند (Lefebvre et al., 1995). از نظر اقتصادی و فنی، اجرای روش‌های خودکار تکرارپذیر و سریع برای شناسایی و طبقه‌بندی بذرها حائز اهمیت است و همچنین بازرسی خودکار میوه‌ها و سبزیجات، قابل اطمینان‌تر و عینی‌تر از بازرسی انسانی است

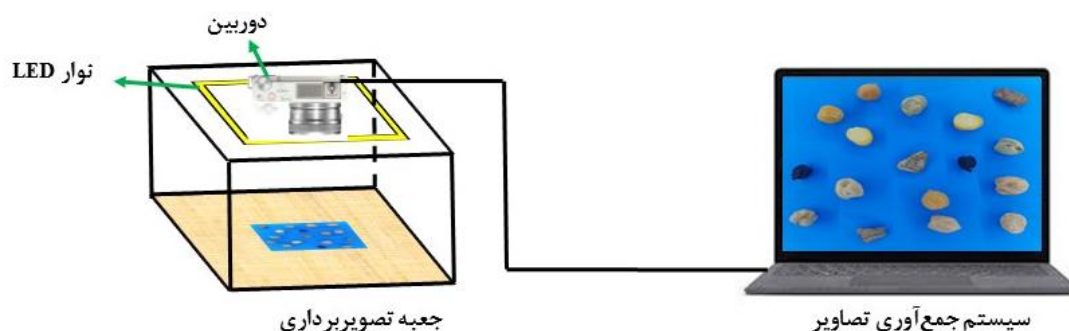
(Cubero et al., 2011). در سیستم‌های خودکار تشخیص محصولات مختلف و یا بذرها، اصولاً تفکیک بر اساس تحلیل تصاویر بدست آمده از آن‌ها و استخراج ویژگی‌هایی مانند اندازه، شکل، رنگ و بافت انجام می‌گردد (Venora et al., 2007). یکی از موضوعات بسیار مهم در تجاری‌سازی محصولات کشاورزی، چه در رابطه با خرید و چه در فروش، کیفیت بصری آن‌ها است که از ویژگی‌هایی مانند رنگ، شکل و اندازه ارزیابی می‌شود که به طور کلی بر قیمت بازار محصول تأثیر می‌گذارد. بنابراین، بازرسی کیفیت بصری یک وظیفه بسیار مهم برای اکثر محصولات کشاورزی است (Aggarwal and Mohan., 2010; Stegmayer et al., 2013). از طرفی علاقه به سیستم‌های بینایی ماشین در صنعت فناوری بذر به دلیل افزایش قابلیت‌های محاسباتی، سرعت پردازش بالاتر الگوریتم‌های موجود و ماهیت غیرمخرب و کارآمد این فناوری‌ها در سراسر جهان رو به رشد بوده است (Brosnan and Sun., 2002).

با توسعه تکنولوژی‌های جدید و وجود دستگاه‌های درجه‌بندی و هوشمند تفکیک ناخالصی‌ها در کشاورزی، بکارگیری این دستگاه‌ها برای تفکیک، درجه‌بندی یا شناسایی ناخالصی‌ها در هر محصولی نیازمند الگوریتم ویژه مربوط به همان محصول است. در نتیجه به دلیل اهمیت بالای حذف دقیق ناخالصی‌ها از محصول نخود و بهبود کیفیت محصول نهایی، در این مطالعه سعی گردید که برای نخود بومی استان کرمانشاه، الگوریتمی توسعه داده شود که بتوان با عملکرد بالایی ناخالصی‌های موجود در محصول نخود را شناسایی و مکان‌یابی کرد. در این مطالعه به کمک الگوریتم پردازش تصویر و استفاده از کلاسبندهای SVM^1 و KNN^2 ، علاوه بر تفکیک نوع ناخالصی‌های نخود، موقعیت مکانی آنها نیز محاسبه و تعیین گردید. موقعیت مکانی یک فاکتور مهم در شناسایی بوده که در اینجا استفاده شده ولی در تحقیقات مشابهی که تا به الان در مورد ناخالصی‌های نخود به کمک الگوریتم‌های دیگر انجام شده است، مورد بررسی قرار نگرفته بود. هدف اصلی این مطالعه کمک به افزایش عملکرد ماشین‌های درجه‌بندی اتوماتیک در شناسایی ناخالصی‌های نخود بومی کشور می‌باشد.

مواد و روش‌ها

آماده‌سازی نمونه‌ها

در این مطالعه نمونه‌های نخود که شامل مخلوط نخود و ناخالصی‌های آن (نخود سالم، نخود سبز، نخود سیاه، نخود رنگی، سنگ و نخود شکسته) بودند از مرکز تحقیقات کشاورزی و منابع طبیعی استان همدان جمع‌آوری شد. در هنگام تصویربرداری، نمونه‌ها کاملاً به صورت کاملاً تصادفی انتخاب و عکس‌برداری شدند. به منظور یکنواخت کردن شدت نور و حذف سایه در تصویربرداری و ایجاد شرایط تقریباً یکسان از باکس تصویربرداری و نور LED، نواری با شدت روشنایی 270 lm/ft استفاده شد (شکل ۱). در این پژوهش، ابتدا ۴۰۰ تصویر RGB، که هر یک شامل شش کلاس مذکور بودند، با سایز اولیه ۲۹۹۲×۲۹۹۲ پیکسل و با استفاده از دوربین تلفن همراه (Samsung Galaxy A6+, South Korea) تهیه شد. همچنین، در تمامی مراحل تصویربرداری، فاصله دوربین تا سطح نمونه‌ها ثابت و برابر ۳۰ سانتی‌متر در نظر گرفته شد.

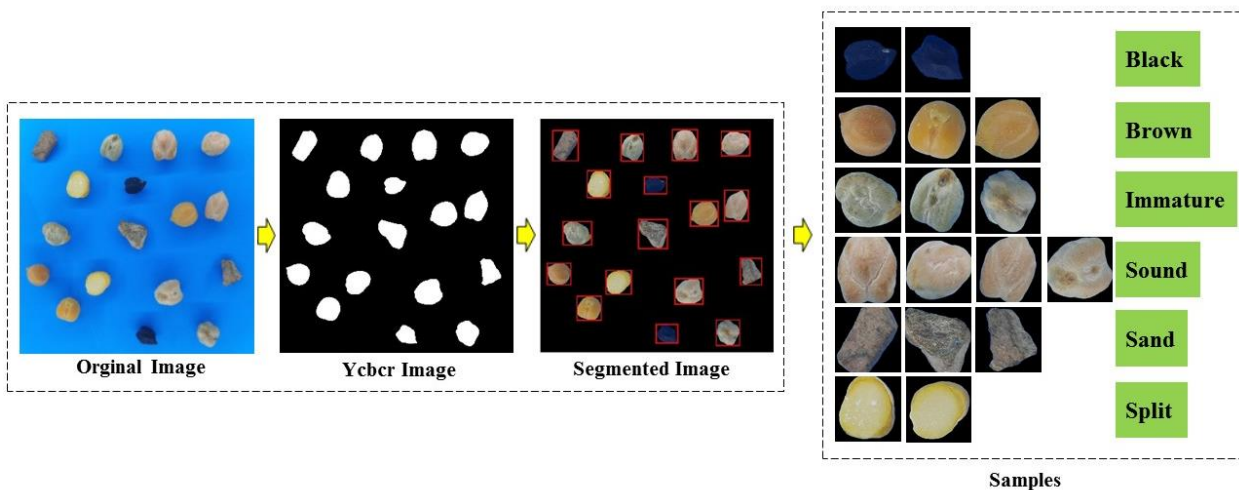


شکل ۱. سیستم تصویربرداری اخذ تصاویر نخود و ناخالصی‌های آن

پیش‌پردازش و دسته‌بندی تصاویر

پس از تهیه تصاویر برای حذف پس‌زمینه اولیه و نویزهای ناخواسته و رسیدن به آبجکت‌های اصلی در تصاویر، بعد از بررسی اولیه فضای رنگی دیگر مانند ابتدا HSV، Lab، YCBCR و فضای YCBCR قابلیت بالاتری نسبت به سایر فضاهای رنگی دیگر در تفکیک اشیاء از

پس‌زمینه داشت. لذا همین فضا برای تحلیل‌های بعدی استفاده شد. بعد از تبدیل تمامی تصاویر از فضای RGB به فضای YCBCR، با بررسی کانال‌های مختلف رنگی در این فضا، تمامی آستانه‌های مختلف با گام افزایشی ۱۰ مورد ارزیابی قرار گرفت و با ارزیابی بصری خیلی دقیق در تصویرهای مختلف، در نهایت در کانال CR و با آستانه $CR > 105$ و حذف ذرات ریز توسط دستور (bwareaopen) تمامی پس‌زمینه حذف گردید و ماسک مورد نظر حاصل شد. با ضرب ماسک بدست آمده در تصویر اصلی، تصویر سگمنت شده ایجاد شد. ادامه کلاس‌های مختلف در تصاویر از هم جدا شده و در دسته‌های مجزایی قرار گرفت (شکل ۲). چون تعداد تصاویر دسته‌بندی شده از هر کلاس یکسان نبود در نظر گرفتن کمترین مقدار آن‌ها، به منظور بالانس کردن تعداد کلاس‌ها، ۶۴۰ عدد تصویر از هر کلاس مورد استفاده قرار گرفت و در نهایت، تعداد ۳۸۴۰ تصویر از ۶ کلاس مختلف تهیه گردید. در این پژوهش تمامی عملیات پردازش تصویر و مدلسازی در محیط متلب (MATLAB R2023b, MathWorks, Natick, MA, USA) انجام شد.



شکل ۲. مراحل پیش‌پردازش تصاویر در محیط متلب و استخراج کلاس‌های مختلف از تصویر اصلی

استخراج ویژگی‌های موثر و پیش‌پردازش داده‌ها

با استفاده از ماتریس GLCM می‌توان روابط مکانی شدت پیکسل‌ها را در یک تصویر توصیف کرد. این روش با بررسی فراوانی جفت‌پیکسل‌هایی با مقادیر شدت مشخص و روابط مکانی معین در تصویر تعریف می‌شود. (Haralick et al., 2007) در این پژوهش، برای استخراج ویژگی‌های بافتی، ابتدا تصویر به فضای خاکستری تبدیل شد و ماتریس GLCM در چهار جهت با زوایای ۰، ۴۵، ۹۰ و ۱۳۵ درجه تشکیل شد. سپس در هر زاویه و با تغییر فاصله پیکسلی از ۱ تا ۱۰، ویژگی‌هایی از جمله همبستگی، انرژی، همگنی و آنتروپی استخراج گردید که در مجموع ۱۶۰ ویژگی را شامل می‌شود. در فضای RGB، از هر یک از کانال‌های R، G و B مقادیر کشیدگی، چولگی، بیشینه، میانه، میانگین و واریانس محاسبه شد (۱۸ ویژگی). همچنین در فضای خاکستری مقادیر کشیدگی، چولگی، میانه، میانگین و واریانس (۵ ویژگی) و دو ویژگی محیط بر مساحت و آنتروپی استخراج گردید. در مجموع، ۱۸۵ ویژگی به دست آمد. در نهایت، ۷۰ درصد داده‌ها برای آموزش مدل و ۳۰ درصد آن‌ها برای ارزیابی عملکرد (تست) مورد استفاده قرار گرفت. پیش از وارد کردن ویژگی‌ها به طبقه‌بند، تمامی ویژگی‌های عددی به منظور مقایسه‌پذیر شدن در مقیاس‌های مختلف، نرمال‌سازی شدند. نرمال‌سازی با استفاده از روش Min-Max انجام گرفت که به صورت رابطه ۱ تعریف می‌شود. این رابطه تمامی مقادیر ویژگی‌ها را به بازه (۰, ۱) نگاشت می‌کند. این مرحله به این دلیل انجام شد که ویژگی‌هایی با دامنه عددی بزرگ‌تر موجب غلبه بر فرآیند یادگیری نشوند و مقایسه عادلانه‌ای میان تمام ویژگی‌های استخراج‌شده برقرار شود.

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad \text{رابطه ۱}$$

کلاسبند SVM

الگوریتم ماشین بردار پشتیبان (SVM) یکی از روش‌های قدرتمند در یادگیری نظارت‌شده است که به‌طور گسترده برای مسائل طبقه‌بندی و رگرسیون مورد استفاده قرار می‌گیرد. این الگوریتم با هدف یافتن یک ابرصفحه‌بهنینه عمل می‌کند که داده‌های دو کلاس را با بیشترین حاشیه ممکن از یکدیگر جدا کند. مزیت اصلی SVM در توانایی آن برای مدیریت داده‌های با ابعاد بالا و همچنین استفاده از هسته‌های

غیرخطی جهت تبدیل فضای ویژگی به فضای با بعد بالاتر است، به گونه‌ای که داده‌هایی که در فضای اولیه تفکیک پذیر نیستند، در فضای جدید به خوبی از یکدیگر جدا شوند. الگوریتم SVM در برابر بیش‌برازش مقاوم است، به‌ویژه در مواردی که حجم داده‌ها کم و تعداد ویژگی‌ها زیاد است. از این‌رو، این مدل در کاربردهایی نظیر تشخیص چهره، شناسایی بیماری از تصاویر و دسته‌بندی متن عملکرد بسیار مطلوبی دارد. انتخاب مناسب تابع هسته^۴ پارامتر هزینه یا جریمه^۵ C (که تعادل بین حاشیه و خطای طبقه‌بندی را تعیین می‌کند)، و پارامتر گاما در هسته‌های غیرخطی، نقش مهمی در کارایی نهایی مدل ایفا می‌کنند (Bazrafkan et al., 2023). یکی دیگر از مزایای SVM، فرموله‌سازی ریاضی دقیق و قابلیت تفسیرپذیری نسبی آن در مقایسه با بسیاری از مدل‌های یادگیری ماشین عمیق است. این ویژگی‌ها موجب شده‌اند که SVM همچنان به‌عنوان یک گزینه‌ی قابل اعتماد در مسائل یادگیری ماشین، به‌ویژه در حوزه‌هایی با داده‌های ساختاریافته و حجم محدود، مطرح باشد (Venkataraman and Mangayarkarasi., 2017).

کلاسبند KNN

الگوریتم KNN یا همان الگوریتم K-Nearest Neighbors) یکی از ساده‌ترین و در عین حال پرکاربردترین الگوریتم‌های یادگیری نظارت شده در حوزه یادگیری ماشین است. KNN هم برای مسائل رگرسیون و هم مسائل طبقه‌بندی کاربرد دارد. KNN همچنین به‌عنوان یک مدل مبتنی بر نمونه شناخته می‌شود، زیرا یک مدل داخلی ایجاد نمی‌کند و فقط نمونه‌های آموزشی را حفظ می‌کند که به‌عنوان دانش برای مرحله‌ی پیش‌بینی استفاده می‌شود. در این الگوریتم، فاصله مهم‌ترین و در عین حال ساده‌ترین فاکتور در کارکرد این کلاسبند است. KNN نزدیک‌ترین نمونه‌های k به نمونه تست را مشخص می‌کند و کلاس آن را با شناسایی برچسب کلاسی که بیشترین فراوانی را دارد، تعیین می‌کند. انتخاب و بهینه‌سازی تعداد نزدیکترین همسایگی یا K و همچنین معیار فاصله مناسب برای KNN بسیار مهم است و می‌تواند باعث بهبود عملکرد در این کلاسبند گردد (Dheer and Singh., 2019).

معیارهای ارزیابی

برای ارزیابی عملکرد مدل، از چهار شاخص متداول شامل صحت (Accuracy)، دقت (Precision)، بازخوانی یا حساسیت (Recall) و F1-score استفاده شد. شاخص دقت، نسبت تعداد نمونه‌هایی را نشان می‌دهد که مدل به‌درستی طبقه‌بندی کرده است و معیاری کلی از عملکرد مدل محسوب می‌شود. معیار Precision نشان می‌دهد که از میان تمام نمونه‌هایی که مدل به‌عنوان یک کلاس خاص تشخیص داده است، چه تعداد واقعاً متعلق به آن کلاس بوده‌اند؛ بنابراین این شاخص بیانگر میزان قابل اعتماد بودن پیش‌بینی‌های مثبت مدل است. معیار Recall توانایی مدل در شناسایی صحیح نمونه‌های مثبت واقعی را منعکس می‌کند و اهمیت ویژه‌ای در مسائلی دارد که از دست رفتن نمونه‌های مثبت هزینه‌بر یا حساس است. در نهایت، شاخص F1-score ترکیبی متوازن از Precision و Recall ارائه می‌دهد و زمانی کاربرد بیشتری دارد که داده‌ها نامتوازن باشند یا مقایسه هم‌زمان این دو معیار ضروری باشد. مجموعه این شاخص‌ها دیدی جامع از عملکرد مدل طبقه‌بندی فراهم می‌کنند. معیارهای صحت، دقت، بازخوانی یا حساسیت و F1-score از ماتریس اغتشاش (جدول ۱) بدست می‌آیند، که به ترتیب در روابط ۲ تا ۵ مشخص شده‌اند.

جدول ۱. ماتریس اغتشاش برای محاسبه معیارهای ارزیابی

		Predicted	
		Positive	Negative
Actua	Positive	TP	FN
	Negative	FP	TN

در ماتریس اغتشاش، مقدار TP تعداد پیش‌بینی‌های صحیح مثبت، TN تعداد پیش‌بینی‌های صحیح منفی، FP پیش‌بینی‌های اشتباه مثبت، و FN پیش‌بینی‌های اشتباه منفی را نشان می‌دهد. این چهار مقدار مبنای محاسبه شاخص‌های ارزیابی عملکرد مدل هستند.

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \quad \text{رابطه ۲}$$

$$Precision A = \frac{TP}{TP + FP} \quad \text{رابطه ۳}$$

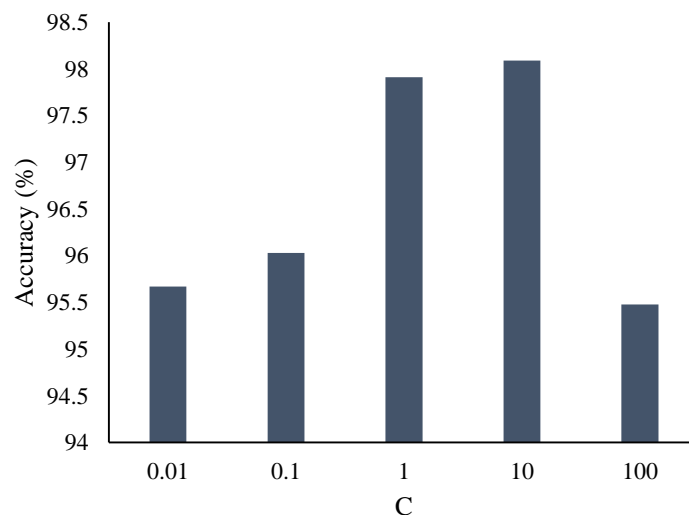
$$Recall A = \frac{TP}{TP + FN} \quad \text{رابطه (۴)}$$

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad \text{رابطه (۵)}$$

نتایج و بحث

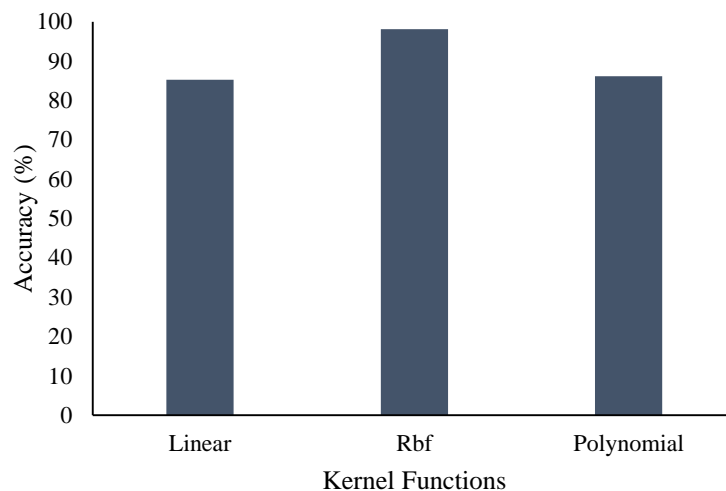
کلاسبند SVM

با توجه به اینکه در مدل SVM بر خلاف مدل‌های دیگر مانند شبکه عصبی که امکان گیرکردن در مینیمم محلی وجود دارد، در این مدل هر جوابی که به دست می‌آید به عنوان مینیمم قطعی یا جهانی می‌باشد، لذا مقادیر بدست آمده از اطمینان بالایی برخوردارند. در این مدل هر چند پارامترهای مختلفی بر عملکرد این مدل تاثیرگذار هستند، ولی همچنان که قبلاً اشاره کردیم اثر پارامتر C و توابع هسته نسبت به سایر پارامترها تاثیر قابل توجهی بر عملکرد این مدل دارند. با بررسی اثر پارامتر C بر عملکرد مدل SVM نتایج نشان داد که با تغییر لگاریتمی این پارامتر، تغییرات قابل توجهی در عملکرد مدل برای داده‌های تست بدست نیامد. همچنان که نتایج شکل ۳ نشان می‌دهد، با بررسی مقدار صحت در C های مختلف، مقدار عملکرد برای C های ۰ تا ۱۰ روند نسبتاً صعودی و برای C های بیشتر از ۱۰ نزولی بود. لذا بهترین عملکرد مدل در مقدار C=10 بدست آمد. در بررسی اثر پارامتر C بر عملکرد مدل SVM مقدار پارامتر گاما برابر با یک ($\gamma = 1$) و از تابع RBF استفاده شد.



شکل ۳. تاثیر مقدار پارامتر هزینه C بر مقدار عملکرد مدل SVM با تابع هسته RBF و مقدار $\gamma = 1$

در بررسی اثر تابع هسته بر عملکرد مدل SVM، نتایج (شکل ۴) نشان داد که تابع RBF به عنوان تابع غیرخطی صحت بهتری نسبت به دو تابع دیگر داشت، به طوری که نسبت به تابع خطی (Linear) ۱۴/۹٪ و نسبت به تابع غیرخطی (Polynomial) ۱۳/۸٪ عملکرد بالاتری را نشان داد. این تابع به دلیل فراهم کردن مرزهای تصمیم‌گیری بسیار منعطف و غیرخطی و بدون ایجاد پیچیدگی محاسباتی بالا عملکرد بهتری نسبت به سایر توابع مخصوصاً در شرایطی که داده ساختار غیرخطی قوی یا نویزدار دارند فراهم می‌کند. این شرایط باعث شده است این مدل از تعمیم‌پذیری بالایی برخوردار باشد و نتایج این پژوهش نیز همسو با قابلیت هسته RBF در افزایش عملکرد مدل SVM است. در چند تحقیق مشابهی که در آن‌ها عملکرد هسته‌ها به شکل ویژه مورد ارزیابی قرار گرفته بود، نتایج مشابهی بدست آمد. در بررسی و مقایسه عملکرد چهار نوع تابع هسته مختلف در روش ماشین بردار پشتیبان برای پیش‌بینی عیوب نرم‌افزاری، با استفاده از ۱۵۲۰ داده آموزش مشخص شد که RBF از توابع هسته خطی و سایر توابع غیرخطی عملکرد بهتری داشت (Azzeh et al., 2023). همچنین در پیش‌بینی تردی سنگ با استفاده از تکنیک‌های تجربی و یادگیری ماشینی مشخص شد RBF مناسب‌ترین هسته برای مدل SVM است (JahedArmaghani et al., 2020).



شکل ۴. تاثیر تابع هسته بر عملکرد مدل SVM

کلاسبند KNN

با توجه به اینکه در کلاسبند KNN، تعداد همسایگی نقش زیادی در عملکرد این کلاسبند دارد، لذا در این مطالعه تعداد k از ۳ تا ۲۱ عدد مورد بررسی قرار گرفت. با بررسی تمامی ترکیب‌های مختلف تعداد همسایگی (شکل ۵-الف)، معیارهای فاصله (شکل ۵-ب) و همچنین با بررسی تمامی معیار وزنی که روابط آن‌ها در شکل ۵-ج و همچنین در رابطه‌های ۶ و ۷ نشان داده شده است، بهترین عملکرد مدل در تعداد $k=13$ ، معیار فاصله City Block و معیار وزنی $1/(c+D^2)$ با مقدار ثابت $c=1$ بدست آمد. نتایج مربوط به اثر معیار فاصله و معیار وزن به ترتیب در شکل ۲-ب و شکل ۲-ج نشان داده شده است. با توجه به نتایج کلی مدل، بیشترین عملکرد مدل KNN برابر با ۹۰/۷۹ درصد حاصل شد.

در رابطه‌های ۶ و ۷ (Gou et al., 2012) و همچنین رابطه‌های مربوط به شکل ۵-ب مقدار D لیستی از k همسایگی است، $D(1)$ و $D(k)$ نیز بیانگر اولین و k امین همسایه می‌باشد. در رابطه ۶ وزن‌دهی ساده و بر اساس عکس فاصله می‌باشد به طوری که نقاط نزدیک اثر بیشتری نسبت به نقاط دور دارند. به عبارت دیگر، اولین همسایه اثر ۱ و آخرین همسایه از بین k همسایه اثرش به حداقل یا صفر می‌رسد. در رابطه ۷ وزن‌دهی بر اساس توابع خطی یا نمایی بوده و در داده‌های با دامنه فاصله بالا عملکرد بهتری دارد.

$$\text{رابطه ۶)} \quad \frac{(D(K)-D)}{(D(K)-D(1))}$$

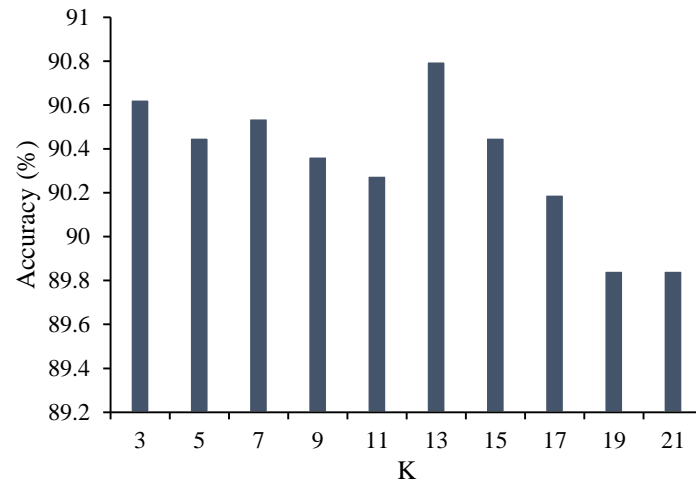
$$\text{رابطه ۷)} \quad \left(\frac{(D(K) - D)}{(D(K) - D(1))} \right) * \left(\frac{(D(K) + D(1))}{(D(K) + D)} \right)$$

در مطالعه‌ای که (Prasath et al., 2017) به بررسی اثر معیارهای فاصله و تعداد همسایگی را بر عملکرد مدل KNN بررسی کردند نتایج مشابهی بدست آمد. در پژوهش آن‌ها، اثر نوع معیار فاصله، تاثیر قابل توجهی بر مقدار عملکرد مدل KNN نسبت به سایر معیارها نشان داد. در تحقیق دیگری که به بررسی اثر روش‌های تنظیم خودکار مقدار k ، انتخاب معیار فاصله مناسب و بررسی حساسیت این پارامترها به ابعاد، نویز و نامتوازنی داده‌ها پرداخته شد، نتایج نشان داد که انتخاب مناسب k به همراه نوع معیار فاصله، اثر زیادی بر صحت و تعمیم‌پذیری مدل دارد (Halder et al., 2024). در تحقیق حاضر نیز که به بررسی اثر پارامترهای تعداد همسایگی و معیار فاصله پرداخته شد، نتایج نشان داد که استفاده از معیار فاصله ساده عملکرد پایین‌تری نسبت به معیارهای فاصله وزن‌دار دارد و این نتیجه یکی از یافته‌های ارزشمند این پژوهش بود.

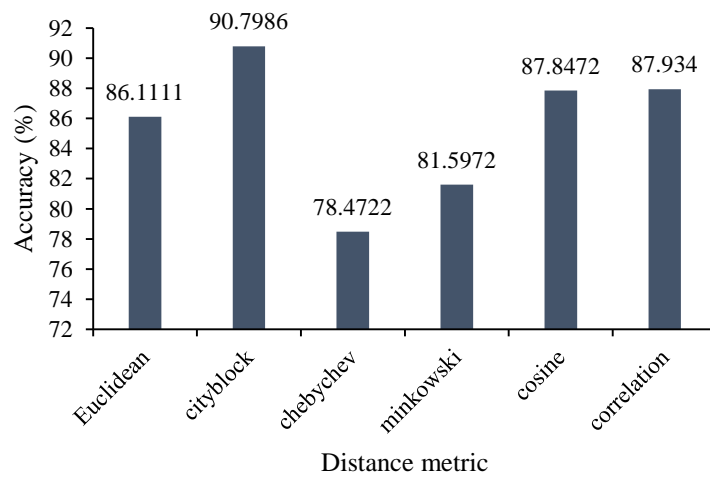
تاثیر اندازه تصاویر بر عملکرد کلاسبندها

با توجه به اینکه در کاربردهای صنعتی و مخصوصاً فرآیندهای جداسازی، سرعت نیز از اهمیت بالایی برخوردار است، اثر اندازه تصاویر ورودی مدل می‌تواند نقش بسزایی در سرعت پردازش اولیه تصاویر داشته باشد. بنابراین در این مطالعه سعی شد اثر اندازه‌های خیلی پایین تصاویر نیز بررسی گردد. هر چند تصاویر نمونه‌های اولیه مربعی و به اندازه 2992×2992 بود ولی برای امکان کاربردی کردن نتایج این تحقیق، تصاویر در سه اندازه 100×100 ، 250×250 و 350×350 مورد بررسی قرار گرفت. با توجه به نتایج بدست آمده در جدول ۲، هر چند با افزایش اندازه تصاویر، عملکرد هر دو مدل از لحاظ صحت افزایش می‌یابد، ولی در رزولوشن‌های بالاتر، عملکرد آنقدر تغییر

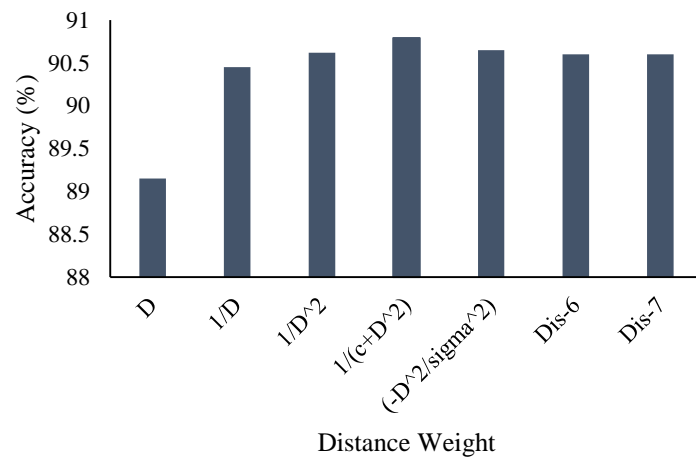
نکرده است که استفاده از تصاویر با اندازه‌های بالاتر را توجیه نماید. لذا با توجه به مقادیر جدول ۱، تصویر ورودی مدل‌ها با اندازه ۲۵۰×۲۵۰ مناسب‌ترین اندازه تصویر انتخاب شد.



الف



ب



ج

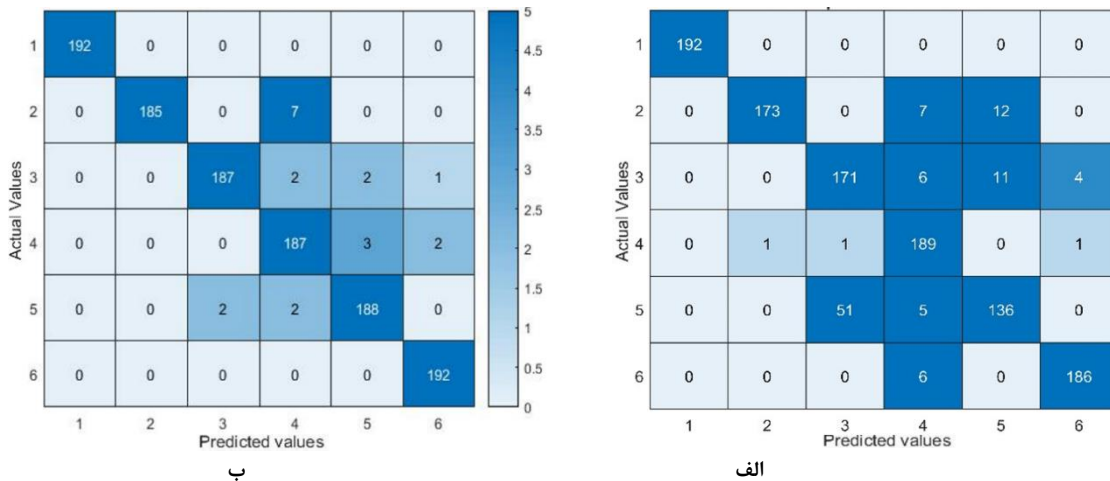
شکل ۵. اثر پارامترهای مختلف بر عملکرد مدل KNN در تشخیص ناخالصی‌های مختلف محصول نخود، (الف) اثر تعداد همسایگی، (ب) اثر معیار وزنی و (ج) اثر شکل یا وزن فاصله

جدول ۲. اثر اندازه تصویر بر صحت (%) مدل‌های SVM و KNN

مدل	سایز تصاویر (Image size)		
	۳۵۰×۳۵۰	۲۵۰×۲۵۰	۱۰۰×۱۰۰
SVM	۹۸/۱۶	۹۸/۰۹	۹۶/۴۸
KNN	۹۰/۹۱	۹۰/۸۰	۹۰/۰۵

مقایسه و ارزیابی کلاسبندها

نتایج حاصل از ماتریس اغتشاش مربوط به دو کلاسبند KNN و SVM به ترتیب در شکل ۶- الف و ب نشان داده شده است. اعداد روی قطر اصلی مشخص‌کننده تعداد داده‌های نمونه‌های مجموعه تست می‌باشند که به درستی تشخیص داده شده‌اند. در این جدول اعداد ۱، ۲، ۳، ۴، ۵ و ۶ به ترتیب نشان دهنده نمونه نخود سیاه، قهوه‌ای، نارس، سالم، سنگ و لپه هستند و تعداد واقعی داده‌های تست برای هر کلاس ۱۹۲ عدد می‌باشد. با توجه به نتایج هر دو جدول، نخود سیاه به دلیل داشتن ویژگی کاملاً متمایز با سایر کلاس‌ها، در هر دو کلاسبند بیشترین تفکیک‌پذیری را داشته است. در سایر کلاس‌ها مانند نخود رنگی، سبز، سالم و سنگ در هر دو کلاسبند، نمونه‌هایی به اشتباه تشخیص داده شده‌اند که این مورد نیز به دلیل نزدیکی ویژگی‌های مرتبط با آن‌ها می‌باشد. در مقایسه دو کلاسبند، نتایج حاصل از ماتریس اغتشاش و معیارهای ارزیابی نشان داده شده در جدول ۳ نشان دهنده عملکرد بالای مدل SVM نسبت به مدل KNN می‌باشد. با توجه به نتایج جدول ۳، از آنجائیکه در مدل SVM علاوه بر شاخص کلی صحت، شاخص F1-Score همه کلاس‌ها بجز کلاس سنگ بالاتر از ۹۷ درصد می‌باشند، نشان از قابلیت بسیار خوب این مدل در شناسایی ناخالصی‌ها می‌باشد. البته پایین بودن شاخص ارزیابی کلاس سنگ به دلیل شباهت رنگی بعضی از سنگ‌ها یا کلوخ‌های سفت با نخودهای مختلف می‌باشد. از طرفی از آنجائیکه اصولاً تفاوت زیادی بین چگالی سنگ با نخودها وجود دارد، اگر سیستم بوجاری بخوبی تنظیم شود این ناخالصی در مرحله اول جدا می‌گردد و در صنعت خیلی نگران‌کننده نیست.



شکل ۶. ماتریس اغتشاش مربوط به کلاسبندهای (الف) KNN و (ب) SVM

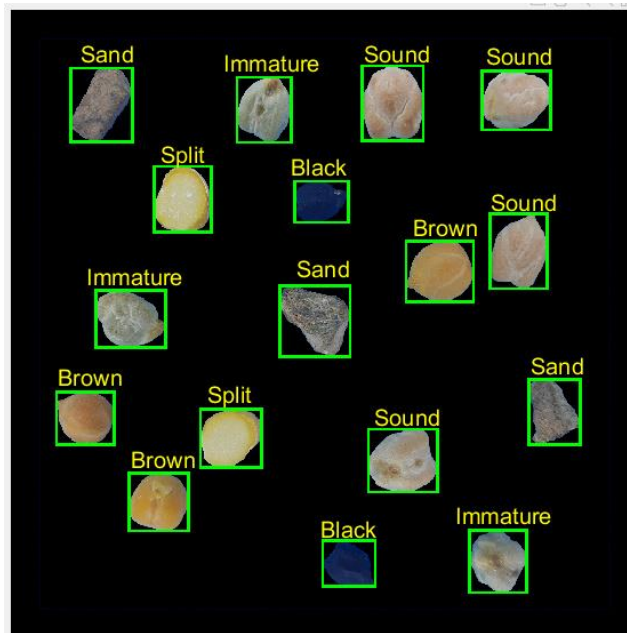
جدول ۳. پارامترهای ارزیابی مدل‌های SVM و KNN در تصاویر با رزولوشن ۲۵۰×۲۵۰

کلاسبند	KNN			SVM		
	F1_score(%)	Recall(%)	Precision(%)	F1_score(%)	Recall(%)	Precision(%)
سیاه	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰
قهوه‌ای	۹۴/۵۳	۹۹/۴۲	۹۰/۱۰	۹۸/۱۴	۱۰۰	۹۶/۳۵
نارس	۸۲/۴۰	۷۶/۶۸	۸۹/۰۶	۹۸/۱۶	۹۸/۹۴	۹۷/۳۹
سالم	۹۳/۳۳	۸۸/۷۳	۹۸/۴۳	۹۵/۸۹	۹۴/۴۴	۹۷/۳۹
سنگ	۷۷/۴۹	۸۵/۵۳	۷۰/۸۳	۹۷/۶۶	۹۷/۴۰	۹۷/۹۱
شکسته	۹۷/۱۲	۹۷/۳۸	۹۶/۸۷	۹۹/۲۲	۹۸/۴۶	۱۰۰
صحت کل (%)		۹۰/۸۸			۹۸/۰۹	

ارقام ناخالصی

با توجه به مطالعات مشابه انجام شده، نتایج مطالعه حاضر کاملاً رضایت‌بخش بود. در مطالعه مشابهی که به کمک مدل شبکه عصبی تفکیک و جداسازی ۷ کلاس مختلف شامل پنج کلاس نخود (سالم، چروکیده، سبز رنگ، قهوه‌ای رنگ و لپه) و دو کلاس ناخالصی (سنگ و ساقه) به انجام شد، متوسط صحت ۹۱/۹ درصد گزارش شد (سلام و خیرعلی‌پور، ۱۴۰۰). یک از دلایل احتمالی در اختلاف نتایج این دو تحقیق می‌تواند به تعداد نمونه‌ها و یا تعداد ویژگی‌های بکار رفته برای مدلسازی باشد. در مدلسازی آنها تنها از ۱۴ ویژگی در حالیکه در تحقیق حاضر ۱۸۵ ویژگی استفاده شد. البته عوامل دیگر مانند نوع مدل بکار رفته، شرایط آموزش و کیفیت و محیط عکسبرداری هم می‌تواند از دلایل بهتر شدن نتایج تحقیق حاضر باشد. از طرفی در تحقیق حاضر از تعداد نمونه‌ها بیشتری استفاده شد. مقایسه نتایج تحقیق حاضر نشان داد که مدل پیشنهادی SVM توانست صحت تشخیص را به اندازه ۶/۸ درصد افزایش دهد. در مطالعه دیگر تمایز و کلاسبندی ۷ نوع حبوبات خشک با استفاده از مدل‌های یادگیری ماشین روی مجموعه‌ای با بیش از ۱۲۹۰۰ نمونه انجام شد. نتایج تحقیق آنها نشان داد که الگوریتم SVM عملکرد بهتری (صحت ۹۴/۴ درصد) نسبت به سایر مدل‌ها داشت و همسو با نتایج تحقیق حاضر بود (Ardeshirifar, 2024).

هر چند اخیراً در حوزه سورتینگ خیلی از محصولات کشاورزی از سورت‌های صنعتی ویژه‌ای طراحی شده‌اند که بیشتر بر اساس رنگ تفکیک و جداسازی ناخالصی‌ها را انجام می‌دهند ولی این سورت‌ها بیشتر در تفکیک ناخالصی‌هایی که اختلاف فاحش رنگی با محصول اصلی داشته باشند مفید هستند (Cujbescu et al., 2023; Fan et al., 2024). در محصول نخود به دلیل شرایط محیطی یا تاخیر در زمان رسیدگی محصول، بعضی از نخودها به رنگ قهوه‌ای روشن دیده می‌شوند که تفکیک آنها همیشه چالش‌برانگیز است یا در داخل محصول ممکن است سنگ‌ها یا کلوخ‌هایی باشند که اختلاف خیلی کمی از نظر رنگی با محصول دارند و بهتر است از ویژگی‌های بافتی برای جداسازی آنها استفاده شود. در پژوهش حاضر با توجه به اینکه کلاس‌های با ویژگی‌های نزدیک به هم به صورت جداگانه‌ای به عنوان کلاس مجزایی تفکیک شدند، و نتایج رضایت‌بخش از این تفکیک حاصل شد، لذا خروجی این تحقیق می‌تواند جایگزین مناسبی برای استفاده در دستگاه‌های موجود باشد. البته نکته‌ای که باید مورد توجه قرار گیرد آن است که در الگوریتم توسعه‌یافته، بهترین زمان تشخیص برای هر تصویر با استفاده از پردازنده CPU حدود ۲۵/۶ میلی‌ثانیه به دست آمده است؛ زمانی که ممکن است برای برخی کاربردهای صنعتی بهینه‌سازی بیشتری نیاز داشته باشد. با این حال، می‌توان با بهره‌گیری از سخت‌افزارهای قدرتمندتر یا استفاده از پردازنده‌های GPU این زمان پردازش را به‌طور قابل توجهی کاهش داد. در شکل ۷ یک نمونه از خروجی مدل تشخیص کلاس‌های مختلف موجود در نمونه نخود نشان داده شده است. با توجه به اینکه یکی از اهداف این تحقیق تعیین مکان کلاس‌های مختلف بود، بعد از تعیین نوع کلاس هر شی، تعیین مکان هر یک از نمونه‌ها، در محیط متلب از طریق مشخص شدن مرکز دانه (از لحاظ مختصات) و ایجاد کادر محدود کننده بر اساس اندازه قطر دانه انجام شد.



شکل ۷. نمونه تصویر تست شده توسط الگوریتم SVM شامل برچسب نوع کلاس و کادرهای مربوط به موقعیت مکانی هر کلاس

نتیجه‌گیری

در این تحقیق به منظور شناسایی و تفکیک نخودهای سالم از انواع ناخالصی‌ها از دو کلاسیک معروف SVM و KNN استفاده شد. کلاسیک SVM نسبت به کلاسیک KNN عملکرد بهتری از خود نشان داد. نتایج بررسی‌های بیشتر در این تحقیق نشان داد که پارامترهایی مانند تعداد همسایگی، معیار فاصله و همچنین اثر فاصله‌های وزن‌دار تاثیر قابل توجهی بر عملکرد مدل داشتند. از طرفی در مدل SVM نیز نوع تابع هسته تاثیر قابل توجهی بر عملکرد مدل داشت. یکی از ویژگی‌های قابل تمایز این تحقیق با پژوهش‌های مشابه، استفاده از کلاس‌هایی با ویژگی‌های خیلی مشابه می‌باشد که با تفکیک‌پذیری آن‌ها امکان بدست آوردن محصول نهایی با خلوص خیلی خوب را میسر می‌کند. هر چند سعی شد از ارقام مرسوم و متداول ایران و همچنین با در نظر گرفتن ناخالصی‌های متداول در محصول نخود استفاده شود، ولی تفاوت در ویژگی‌های رنگی، بافتی یا شکلی بعضی از ارقام دیگر می‌تواند بر قدرت تعمیم‌پذیری مدل توسعه داده شده تاثیرگذار باشد. لذا توصیه می‌شود در مطالعات تکمیلی به منظور افزایش تعمیم‌پذیری، ناخالصی‌ها و یا رقم‌های زیادی در آموزش استفاده گردد. با توجه به گسترش کارایی ربات‌ها و ماشین‌های خودکار در انواع زمینه‌ها، از خروجی این مطالعه و الگوریتم بدست آمده می‌توان در شناسایی موقعیت مکانی و نوع ناخالصی‌های محصول نخود در درجه‌بندهای اتوماتیک استفاده کرد. در انتها پیشنهاد می‌گردد که به منظور ارزیابی دقیق این الگوریتم، این مدل در ماشین‌های اتوماتیک لیزری درجه‌بندی محصولات کشاورزی تست و ارزیابی گردد.

منابع

سلام، سمیه؛ خیرعلی پور، کامران (۱۴۰۰). توسعه و ارزیابی سامانه طبقه بندی نخود بر اساس فناوری پردازش تصویر مرئی و شبکه عصبی مصنوعی. فناوری‌های جدید در صنعت غذا. ۹(۲)، ۱۸۱-۱۹۳. <https://doi.org/10.22104/jift.2021.5173.2063>

REFERENCES

- Aggarwal, A. K., & Mohan, R. (2010). Aspect ratio analysis using image processing for rice grain quality. *International Journal of Food Engineering*, 6(5). DOI:[10.2202/1556-3758.1788](https://doi.org/10.2202/1556-3758.1788)
- Ardeshirifar, R. (2024). Automated Classification of Dry Bean Varieties Using XGBoost and SVM Models. *arXiv preprint arXiv:2408.01244*. <https://doi.org/10.48550/arXiv.2408.01244>
- Azzeh, M., Elsheikh, Y., Nassif, A. B., & Angelis, L. (2023). Examining the performance of kernel methods for software defect prediction based on support vector machine. *Science of Computer Programming*, 226, 102916. <https://doi.org/10.1016/j.scico.2022.102916>
- Bazrafkan, A., Navasca, H., Kim, J. H., Morales, M., Johnson, J. P., Delavarpour, N., ... & Flores, P. (2023). Predicting Dry Pea Maturity Using Machine Learning and Advanced Sensor Fusion with Unmanned Aerial Systems (UASs). *Remote Sensing*, 15(11), 2758. <https://doi.org/10.3390/rs15112758>
- Brosnan, T., & Sun, D. W. (2002). Inspection and grading of agricultural and food products by computer vision systems—a review. *Computers and electronics in agriculture*, 36(2-3), 193-213. [https://doi.org/10.1016/S0168-1699\(02\)00101-1](https://doi.org/10.1016/S0168-1699(02)00101-1)
- Cubero, S., Aleixos, N., Moltó, E., Gómez-Sanchis, J., & Blasco, J. (2011). Advances in machine vision applications for automatic inspection and quality evaluation of fruits and vegetables. *Food and bioprocess technology*, 4, 487-504. <https://doi.org/10.1007/s11947-010-0411-8>
- Cujbescu, D., Nenciu, F., Persu, C., Găgeanu, I., Gabriel, G., Vlăduț, N. V., ... & Boruz, S. P. (2023). Evaluation of an Optical Sorter Effectiveness in Separating Maize Seeds Intended for Sowing. *Applied Sciences*, 13(15), 8892. DOI:[10.3390/app13158892](https://doi.org/10.3390/app13158892)
- Dheer, P., & Singh, V. (2019). Classifying wheat varieties using machine learning model. *Journal of Pharmacognosy and Phytochemistry*, 8(3), 47-49. DOI: 10.13140/RG.2.2.16338.81600
- Fan, F., Chen, H., Gao, Y., & Mou, T. (2024). Quantitative detection and sorting of broken kernels and chalky grains in milled rice using computer vision algorithms. *Journal of Food Engineering*, 383, 112225. <https://doi.org/10.1016/j.jfoodeng.2024.112225>
- Geng, J., Min, H., & Rao, X. (2021). Separation of clods and stones from harvested potatoes using laser backscattering imaging technique. *Journal of Food Measurement and Characterization*, 15, 3262-3273. <https://doi.org/10.1007/s11694-021-00896-9>
- Gou, J., Du, L., Zhang, Y., & Xiong, T. (2012). A new distance-weighted k-nearest neighbor classifier. *J. Inf. Comput. Sci*, 9(6), 1429-1436.
- Halder, R. K., Uddin, M. N., Uddin, M. A., Aryal, S., & Khraisat, A. (2024). Enhancing K-nearest neighbor algorithm: a comprehensive review and performance analysis of modifications. *Journal of Big Data*, 11(1), 113. <https://doi.org/10.1186/s40537-024-00973-y>
- Haralick, R. M., Shanmugam, K., & Dinstein, I. H. (2007). Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics*, (6), 610-621. [10.1109/TSMC.1973.4309314](https://doi.org/10.1109/TSMC.1973.4309314)

- Jahed Armaghani, D., Asteris, P. G., Askarian, B., Hasanipناه, M., Tarinejad, R., & Huynh, V. V. (2020). Examining hybrid and single SVM models with different kernels to predict rock brittleness. *Sustainability*, 12(6), 2229. <https://doi.org/10.3390/su12062229>
- Kanouni, H.; Ahari, D.S.; Khoshroo, H.H. Chickpea Research and Production in Iran. In Proceedings of the 7th International Food Legume Research Conference, Marrakech, Morocco, 6–8 May 2018. (In Persian).
- Kılıç, K., Boyacı, I. H., Köksel, H., & Küsmenoğlu, İ. (2007). A classification system for beans using computer vision system and artificial neural networks. *Journal of Food Engineering*, 78(3), 897-904. <https://doi.org/10.1016/j.jfoodeng.2005.11.030>
- Lefebvre, M., Zimmerman, T., Baur, C., Guegerli, P., & Pun, T. (1995, January). Potato operation: automatic detection of potato diseases. In *Optics in Agriculture, Forestry, and Biological Processing* (Vol. 2345, pp. 2-9). SPIE. <https://doi.org/10.1117/12.198858>
- Liu, D., Ning, X., Li, Z., Yang, D., Li, H., & Gao, L. (2015). Discriminating and elimination of damaged soybean seeds based on image characteristics. *Journal of Stored Products Research*, 60, 67-74. <https://doi.org/10.1016/j.jspr.2014.10.001>
- Ozan, A. K. I., Güllü, A., & Uçar, E. (2015, November). Classification of rice grains using image processing and machine learning techniques. In *International scientific conference* (pp. 20-21).
- Prasath, V. B., Alfeilat, H. A. A., Hassanat, A., Lasassmeh, O., Tarawneh, A. S., Alhasanat, M. B., & Salman, H. S. E. (2017). Distance and similarity measures effect on the performance of K-nearest neighbor classifier—a review. *arXiv preprint arXiv:1708.04321*. DOI: [10.48550/arXiv.1708.04321](https://doi.org/10.48550/arXiv.1708.04321)
- Salam, S., Kheiralipour, K. (2022). Development and evaluation of chickpea classification system based on visible image processing technology and artificial neural network. *Innovative Food Technologies*, 9(2), 181-193. <https://doi.org/10.22104/jift.2021.5173.2063>. (In Persian)
- Salam, S., Kheiralipour, K., & Jian, F. (2022). Detection of unripe kernels and foreign materials in chickpea mixtures using image processing. *Agriculture*, 12(7), 995. <https://doi.org/10.22104/jift.2021.5173.2063>
- Stegmayer, G., Milone, D. H., Garran, S., & Burdyn, L. (2013). Automatic recognition of quarantine citrus diseases. *Expert systems with applications*, 40(9), 3512-3517. <https://doi.org/10.1016/j.eswa.2012.12.059>
- Venkataraman, D., & Mangayarkarasi, N. (2017, September). Support vector machine based classification of medicinal plants using leaf features. In *2017 International conference on advances in computing, communications and informatics (ICACCI)* (pp. 793-798). IEEE. DOI: [10.1109/ICACCI.2017.8125939](https://doi.org/10.1109/ICACCI.2017.8125939)
- 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(1), 161-166. <https://doi.org/10.1016/j.foodres.2006.09.001>
- Voisin, A. S., Guéguen, J., Huyghe, C., Jeuffroy, M. H., Magrini, M. B., Meynard, J. M., ... & Pelzer, E. (2014). Legumes for feed, food, biomaterials and bioenergy in Europe: a review. *Agronomy for Sustainable Development*, 34, 361-380. <https://doi.org/10.1007/s13593-013-0189-y>
- Yang, H., Ni, J., Gao, J., Han, Z., & Luan, T. (2021). A novel method for peanut variety identification and classification by Improved VGG16. *Scientific Reports*, 11(1), 15756. <https://doi.org/10.1038/s41598-021-95240-y>
- Zareiforush, H., Minaei, S., Alizadeh, M. R., & Banakar, A. (2016). Qualitative classification of milled rice grains using computer vision and metaheuristic techniques. *Journal of food science and technology*, 53, 118-131. doi: [10.1007/s13197-015-1947-4](https://doi.org/10.1007/s13197-015-1947-4)
- Zhu, B., Jiang, L., Jin, F., Qin, L., Vogel, A., & Tao, Y. (2007). Walnut shell and meat differentiation using fluorescence hyperspectral imagery with ICA-kNN optimal wavelength selection. *Sensing and Instrumentation for Food Quality and Safety*, 1, 123-131. <https://doi.org/10.1007/s11694-007-9015-z>