

طراحی یک سیستم خبره برای تشخیص اصالت ارقام برنج با استفاده از ترکیب ویژگی‌های بافتی تصاویر توده‌ی برنج

سید جلال الدین موسوی راد و فردین اخلاقیان طاب

چکیده

برنج یکی از مهمترین مواد غذایی در ایران می‌باشد. ممکن است یک رقم برنج با کیفیت بالا با اهدافی مثل سودجویی با یک رقم برنج دیگر که کیفیت پایین‌تری دارد مخلوط شود. این مقاله به ارائه‌ی یک سیستم خبره جهت تشخیص اصالت ارقام برنج با استفاده از تصاویر گرفته‌شده از توده‌ی برنج پرداخته است. ایده‌ی اصلی جهت تشخیص اصالت بر روی بافت برنج استوار است که ممکن است با مخلوط شدن دو رقم برنج با یکدیگر، بافت توده‌ی آن‌ها نیز تغییر نماید. به این منظور، در ابتدا، با استفاده از یک جعبه سیاه در فواصل مختلف ترکیبی، تصویربرداری انجام شد. سپس ویژگی‌های بافتی مربوط به توده‌ی برنج با استفاده از سه روش هیستوگرام تصویر، ماتریس هم‌رویدادی و الگوی دودویی محلی به دست آمد. جهت پیدا کردن ویژگی‌های برتر استخراجی، از الگوریتم ژنتیک استفاده شد. سپس از یک شبکه عصبی برای رگرسیون استفاده شد. ورودی این شبکه عصبی، ویژگی‌های برتر استخراجی و خروجی آن درصد ترکیب بود. بهترین کارایی با استفاده از الگوی دودویی محلی با مقدار خطای $4/92$ و ضریب همبستگی $0/9153$ به دست آمد. برای بهبود نتایج ارائه شده، نتایج مرحله قبل با هم ترکیب شد که مقدار خطا به $4/21$ و ضریب همبستگی $0/9356$ کاهش پیدا کرد. نتایج این پژوهش می‌تواند در ساخت یک سیستم اصالت‌سنج ارقام برنج مورد استفاده قرار گیرد.

کلید واژه‌ها

اصالت سنجی، برنج، بافت، الگوریتم ژنتیک، هیستوگرام، ماتریس هم‌رویدادی، الگوی دودویی محلی

کشورهای آمریکای شمالی و اروپا روز به روز در حال افزایش است [۱].

روش‌های سنتی ارزیابی حسی در تعیین کیفیت مواد غذایی کاربرد زیادی دارند ولی این روش‌ها زمانبر و پرهزینه هستند. همچنین شرایط فیزیکی انسان مثل خستگی یا حتی شرایط روحی می‌تواند بر نتیجه کار تاثیرگذار باشد. این عوامل سبب ایجاد انگیزه برای توسعه روش‌های جانشین است که در زمان کمتر و با دقت بیشتر خصوصیات کلیدی محصول را ارزیابی کنند. بینایی ماشین یکی از این روش‌ها است. بینایی ماشین، تکنولوژی تهیه و آنالیز تصاویر یک صحنه واقعی به وسیله کامپیوتر در راستای کسب اطلاعات یا کنترل یک پروسه است. می‌توان با کمک بینایی ماشین، خصوصیات تصویر را استخراج نمود و از آن برای تشخیص و شناسایی کیفیت انواع محصولات استفاده کرد. در

۱ مقدمه

محصول برنج یک از مهمترین اقلام غذایی در ایران و جهان می‌باشد. طبق پیش بینی فائو، در سال ۲۰۱۲، ۷۳۲ میلیون تن برنج تولید خواهد شد. امروزه، برنج غذای اصلی نیمی از مردم دنیا است. غذای اصلی مردم آسیا برنج است و مصرف برنج در

این مقاله در مردادماه ۱۳۹۱ دریافت، در آذرماه ۱۳۹۱ بازنگری و در بهمن‌ماه ۱۳۹۱ پذیرفته شد.

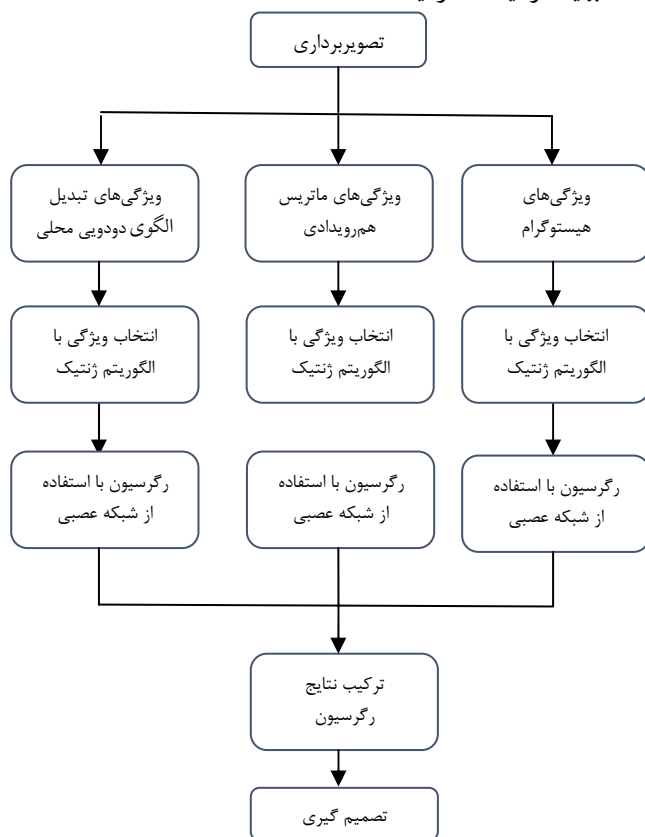
سید جلال الدین موسوی راد، دانشگاه کردستان، گروه مهندسی کامپیوتر و فناوری اطلاعات - دانشگاه کاشان، گروه مهندسی کامپیوتر [jalalamoosavirad@gmail.com](mailto:jalamoosavirad@gmail.com)

فردین اخلاقیان طاب، دانشگاه کردستان، گروه کامپیوتر و فناوری اطلاعات fardin.tab@gmail.com

الگوی دودویی محلی استفاده گشته است. روش ارائه شده نسبت به تغییرات روشنایی تصویر و چرخش تصویر حساس نیست. در ضمن نتایج نشان می‌دهد که این روش با خطای قابل قبولی قادر به تشخیص اصالت ارقام برنج می‌باشد. در این مقاله در بخش ۲، روش پیشنهادی برای تشخیص اصالت ارقام برنج مطرح می‌گردد. بخش ۳ به ارائه نتایج تجربی روش پیشنهادی اختصاص یافته است و در بخش ۴ به بحث و نتیجه‌گیری پرداخته می‌شود.

۲ مواد و روش‌ها

در این مقاله، روشی جدید برای تشخیص رقم برنج اصل از ترکیبی و تخمین مقدار ترکیب ارائه شده است. دیاگرام شکل ۱ مراحل کلی انجام کار را برای این منظور نشان می‌دهد. به این منظور در ابتدا، با استفاده از یک جعبه سیاه، تصاویر توده‌ای مربوط به ارقام برنج در فاصله‌های متفاوت ترکیبی به دست آمده است. بعد از گرفتن تصاویر، ویژگی‌های بافتی توده‌ی برنج با استفاده از سه روش هیستوگرام تصویر، ماتریس هم‌رویدادی و الگوی دودویی محلی بدست آمده است. برای پیدا کردن بهترین ویژگی‌ها، الگوریتم ژنتیک مورد استفاده قرار گرفت. با استفاده از هر مجموعه ویژگی استخراجی، رگرسیون انجام شده است. در نهایت جهت کاهش خطا از نتایج هر سه رگرسیون استفاده کرده و یک مدل بهینه‌تر ایجاد گردیده است.



شکل ۱ دیاگرام مراحل کلی روش پیشنهادی

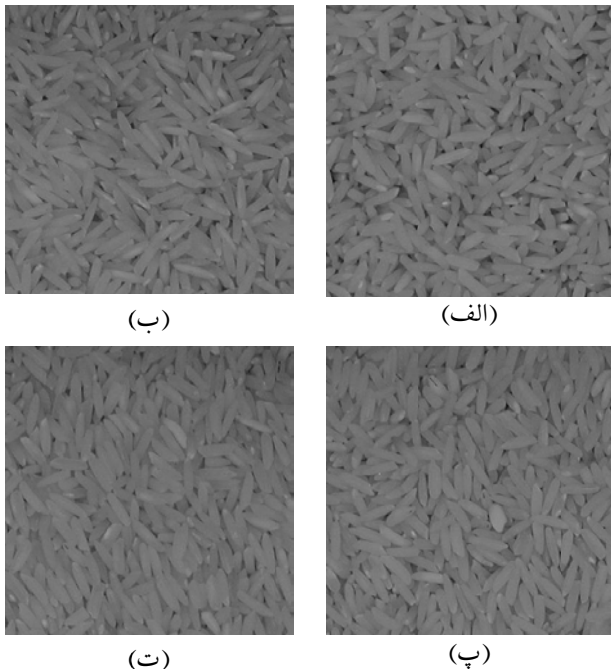
سال‌های اخیر پژوهش‌های محدودی به منظور کیفیت سنجی و درجه بندی برنج با استفاده از فناوری بینایی ماشین انجام شده است. لیو و همکاران [۲] روشی را برای طبقه بندی شش رقم برنج شلتوک دار با استفاده از ویژگی‌های رنگ و شکل ارائه دادند. در این مقاله، هفت ویژگی رنگی و چهل ویژگی شکلی از هر دانه برنج شلتوک دار استخراج شده است و با استفاده از همبستگی بین ویژگی‌ها، ۱۷ ویژگی را به عنوان ویژگی‌های برتر معرفی نموده است. دقت این روش ۸۸٫۳٪ گزارش شده است [۳]. در کاری مشابه، ورما شش ویژگی مربوط به شکل برنج (محیط، مساحت، طول ماکزیمم، پهنای ماکزیمم، فشردگی و درازا) را برای سه رقم برنج استخراج کرد و با استفاده از شبکه عصبی به دقت طبقه بندی بین ۹۰ تا ۹۵٪ رسید. علت بالاتر بودن دقت در این کار نسبت به کار قبلی، استفاده از تعداد کمتری ارقام برنج بوده است [۴]. جینزوس و همکاران (۲۰۱۰) با استفاده از تکنیک‌های بینایی ماشین به بررسی رنگ و ظاهر ارقام برنج پرداختند. [۵]. در پژوهش دیگری دالان (۲۰۰۴) به بررسی توزیع شکستگی در ارقام مختلف برنج با استفاده از تصاویر گرفته شده از توده برنج پرداخت. به این منظور، تحلیل طول و عرض برنج‌های شکسته و سالم، انجام گرفت [۶]. پابامالی و همکاران (۲۰۱۲) ویژگی‌های ماتریس هم‌رویدادی و رنگی را از تصاویر ارقام مختلف برنج استخراج نمودند و با استفاده از یک شبکه عصبی، طبقه بندی را انجام دادند. دقت بین ۶۸٪ تا ۹۴٪ برای ۴ رقم مختلف برنج، حاصل این پژوهش بوده است [۷]. در مقالات ارائه شده، تمامی افراد به تشخیص ارقام به صورت دانه‌ای پرداخته‌اند و تشخیص توده‌ای (تشخیص ارقام از روی کل برنج‌ها و نه تک تک آن‌ها) بررسی نشده است. به این منظور، نویسندگان ارقام برنج را با استفاده از تصاویر گرفته شده از بافت توده‌ی برنج شناسایی کردند. [۸، ۹]. در این دو مقاله، برای تشخیص توده‌ای ارقام برنج، بافت توده برنج مورد بررسی قرار گرفته است. نتایج این دو مقاله، نشان داد که بافت توده‌ی برنج توانایی جداسازی ارقام مختلف برنج از یکدیگر را با کارایی بالا دارد.

ممکن است برنج با اهدافی مثل سودجویی، با ارقام دیگر مخلوط شود. تشخیص درست رقم برنج اصل از ترکیبی، یکی از چالش‌های مطرح می‌باشد. مرور منابع نشان می‌دهد این مساله، تاکنون مورد بررسی قرار نگرفته است و این کار، اولین تلاش در این زمینه می‌باشد؛ با تحقیقاتی که در شمال کشور از کشاورزان و افراد خبره انجام شد مشخص گردید که این افراد با توجه به نحوه قرارگیری ارقام برنج در کنار یکدیگر می‌توانند به تشخیص اصالت بپردازند به عبارت دیگر در واقع این افراد بافت توده‌ی برنج را مورد استفاده قرار می‌دادند؛ همچنین در [۸، ۹] نشان داده شد که با استفاده از بافت توده‌ی ارقام برنج، می‌توان ارقام برنج را شناسایی کرد؛ این دلایل باعث شد تا در این مقاله، بافت توده برنج جهت اصالت سنجی مورد بررسی قرار بگیرد. جهت استخراج ویژگی‌های بافتی از هیستوگرام، ماتریس هم‌رویدادی و

۲-۱ ارقام برنج

در کنار یکدیگر می‌توانند به تشخیص اصالت پردازند به عبارت دیگر در واقع این افراد بافت توده‌ی برنج را مورد استفاده قرار می‌دادند؛ علاوه بر این، در [۸، ۹] نشان داده شد که با استفاده از بافت توده‌ی ارقام برنج، می‌توان ارقام برنج را شناسایی کرد؛ به همین دلیل در این مقاله، بافت توده برنج جهت اصالت سنجی مورد بررسی قرار گرفته است.

هر چند تعریف رسمی برای بافت وجود ندارد اما از نظر شهودی، این توصیفگر معیارهایی از خواص مثل هموار بودن، درشت بودن و منظم بودن را فراهم می‌کند [۱۰]. بافت توده برنج نیز از کنار هم گرفتن دانه‌های برنج در کنار یکدیگر به وجود می‌آید. در واقع شکل ظاهری که یک فرد از برنج‌های روی هم می‌بیند بافت توده برنج می‌باشد. روش‌های مختلفی برای استخراج ویژگی‌های بافت وجود دارد که در این پژوهش از هیستوگرام تصویر، ماتریس هم رویدادی و الگوی دودویی محلی استفاده شده است.



شکل ۳ نمونه‌هایی از تصاویر گرفته شده (الف) رقم برنج هاشمی (ب) رقم برنج باسماتی (پ) برنج هاشمی که ۲۰ درصد آن برنج باسماتی است (ت) برنج هاشمی که ۴۰ درصد آن برنج باسماتی است

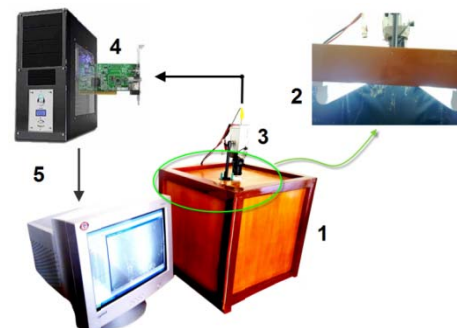
۲-۳-۱ ویژگی‌های استخراجی از هیستوگرام

یکی از روش‌های توصیف بافت، استفاده از ویژگی‌های آماری مربوط به هیستوگرام شدت یک تصویر [۱۱] می‌باشد. از هیستوگرام شدت ۳۱ ویژگی آماری استخراج شده است. برای این کار ابتدا هیستوگرام سطوح خاکستری تصویر به دست آمده و نرمالیزه آن محاسبه شد ($P(i)$). سپس با استفاده از بردار $P(i)$ و روابط مبتنی بر احتمال وقوع سطوح خاکستری، ویژگی‌های آماری استخراج گشته است. ۶ ویژگی میانگین، انحراف معیار، نرمی، ممان سوم، یکنواختی و آنتروپی از هیستوگرام شدت تصاویر استخراج شد. علاوه بر این، ۲۵ ویژگی دیگر مربوط به گروه‌های فراوانی

برای تشخیص اصالت برنج، از دو رقم برنج هاشمی و باسماتی که دارای ظاهری شبیه به هم هستند استفاده شده است. برنج هاشمی که دارای کیفیت مطلوبی است به عنوان برنج اصل انتخاب گشته و برنج باسماتی که دارای کیفیت پایین تری است به عنوان رقم برنجی انتخاب شده که قرار است با برنج هاشمی ترکیب شود. جهت اطمینان از رقم برنج هاشمی، این رقم برنج از شهر آمل تهیه شده است.

۲-۲ تصویر برداری

جهت تصویربرداری، یک جعبه سیاه مجهز به سیستم نورپردازی یکنواخت استفاده شد که در بالای آن یک دوربین (مدل C-2000 Olympus) با دقت ۲ مگاپیکسل نصب شده است. فاصله بین دوربین و نمونه‌های توده‌ای، ۲۰ سانتیمتر در نظر گرفته شده است. شکل ۲، نمایی از جعبه سیاه استفاده شده را نشان می‌دهد. برای تصویربرداری، در ابتدا از برنج هاشمی به عنوان برنج اصل ۱۰۰ تصویر تهیه شد. در مرحله بعد برنج هاشمی ترکیبی طوری به دست آمد که ۵ درصد آن از نوع باسماتی و ۹۵ درصد دیگر از نوع هاشمی باشد. سپس با استفاده از این ترکیب، ۱۰۰ تصویر دیگر تهیه گشت. در هر مرحله، برنج هاشمی طوری به دست می‌آمد که ۵ درصد ترکیب بیشتری از مرحله قبل داشته باشد (یعنی ترکیب ۱۰ درصد برنج باسماتی و ۹۰ درصد برنج هاشمی در مرحله سوم، ترکیب ۱۵ درصد برنج باسماتی و ۸۵ درصد برنج هاشمی در مرحله چهارم و...) و مراحل تصویربرداری دوباره انجام می‌گشت. این کار تا زمانی انجام شد که ۵۰ درصد از برنج هاشمی ترکیبی، از نوع برنج باسماتی باشد. قبل از تصویربرداری، نمونه‌های برنج در یک کیسه نایلونی ریخته شده و کاملاً مخلوط گشته است. سپس نمونه برنج با آرامی در ظرفی مکعبی شکل ریخته شده و سطح آن با استفاده از یک تخته صاف شده است. برای هر تصویر گرفته شده، این کار انجام شده است. نمونه‌هایی از تصاویر گرفته شده، در شکل ۳ نشان داده شده است.



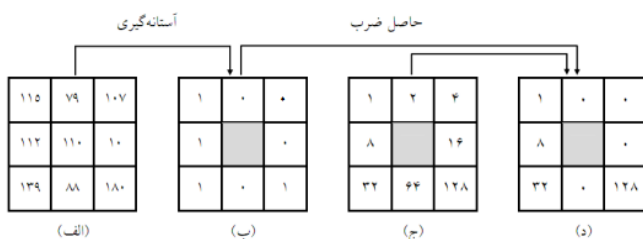
شکل ۲ نمایی از جعبه سیاه و تجهیزات استفاده شده

۲-۳-۲ استخراج ویژگی‌های بافتی

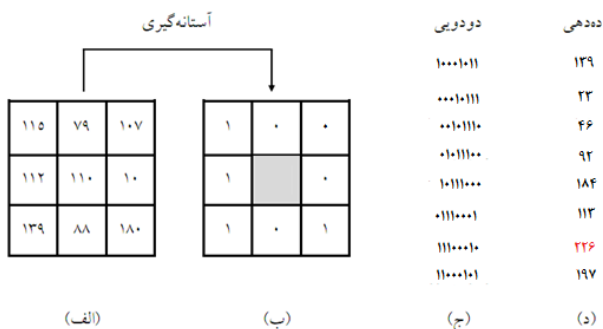
با تحقیقاتی که در شمال کشور از کشاورزان و افراد خیره انجام شد مشخص گردید که این افراد با توجه به نحوه قرارگیری ارقام برنج

الگوی باینری محلی دیده می‌شود که خروجی این عملگر یک عدد باینری p بیتی است که دارای 2^p مقدار متفاوت است. ضمناً دیده می‌شود که مقدار $LBP_{P,R}$ به نحوه اندیس‌گذاری پیکسل‌های موجود در همسایگی کاملاً وابسته است. لذا برای اینکه بتوان مقدار یکتایی را به هر کدام از الگوهای محلی اختصاص داد با چرخش ساعتگرد عدد باینری به دست آمده و با انتخاب بیشینه مقدار ممکن، مقاوم نمودن ماتریس LBP به چرخش انجام می‌شود. همانطور که در شکل ۵ دیده می‌شود مقدار 2^{34} بیشینه مقدار ممکن است که در چرخش هفتم به دست می‌آید.

از ماتریس الگوهای دودویی محلی در مجموع ۳۱ ویژگی شامل میانگین، کنتراست، همواری، ممان سوم، یکنواختی، آنتروپی و ۲۵ گروه فراوانی هیستوگرام سطح خاکستری استخراج گردید.



شکل ۴ مراحل به دست آوردن الگوهای دودویی محلی پایه



شکل ۵ روش مقاوم نمودن ماتریس LBP به چرخش

۲-۴ انتخاب ویژگی‌های برتر با استفاده از الگوریتم ژنتیک

ممکن است بعضی از ویژگی‌های استخراجی زائد باشند و نقشی در بالا بردن دقت طبقه‌بندی نداشته باشند؛ حتی ممکن است این ویژگی‌ها باعث شوند تا دقت طبقه‌بندی پایین بیاید. بنابراین نیاز به فرایندی برای یافتن کاراترین مجموعه ویژگی‌های طبقه‌بندی وجود دارد. برای انتخاب ویژگی‌های کارا، از الگوریتم ژنتیک استفاده شده است.

الگوریتم ژنتیک یک روش بهینه‌سازی است که از تکامل طبیعت الهام گرفته است. الگوریتم ژنتیک راه حل برای یک مسئله مشخص را به صورت یک کروموزوم کد می‌کند و یک جمعیت از کروموزوم‌ها ایجاد می‌کند. هر کروموزوم شامل تعدادی ژن است که تعداد و نحوه نمایش آن‌ها بستگی به تعداد پارامترهای مسئله بهینه‌سازی دارد. الگوریتم ژنتیک در بار اول اجرا، جمعیتی از

هیستوگرام سطح خاکستری است. برای محاسبه ویژگی‌های گروه‌های فراوانی هیستوگرام سطح خاکستری، ۲۵۶ مقدار ممکن هیستوگرام سطوح خاکستری در بازه بین صفر تا ۲۵۵ به ۲۵ بازه تقسیم شد و تعداد پیکسل‌هایی از تصویر که مقدار سطح خاکستری‌شان در هر بازه قرار می‌گرفت به عنوان ویژگی معرفی شد. بدین ترتیب ۲۵ عدد برای هر تصویر به دست می‌آید که هر کدام یک ویژگی تلقی می‌شود.

۲-۳-۲ ویژگی‌های استخراجی از ماتریس هم‌رویدادی

اگر Q موقعیت مکانی دو پیکسل، نسبت به یکدیگر را مشخص نماید و f تصویری با L سطح شدت باشد؛ آنگاه ماتریس هم‌رویدادی^۱ ماتریسی است که عنصر g_{ij} آن تعداد دفعاتی می‌باشد که جفت‌های پیکسل با شدت‌های z_i و z_j در موقعیتی از تصویر f قرار دارند که Q مشخص می‌کند. تعداد سطوح شدت ممکن در تصویر، اندازه ماتریس G را تعیین می‌کند. برای تصویر ۸ بیتی، اندازه G برابر با 256×256 خواهد بود. با استفاده از این ماتریس و انتخاب عمگرهای موقعیتی مناسب، می‌توان الگوهای بافت شدت را تشخیص داد. به این منظور ۱۱ ویژگی از ماتریس هم‌رویدادی استخراج شده است [۱۲]. این ویژگی‌ها عبارتند از ماکزیمم احتمال، همبستگی، کنتراست، یکنواختی، آنتروپی، همگنی، عدم شباهت، میانگین در جهت x و y ، سایه خوشه‌۲ و برجستگی خوشه‌۳.

برای هر تصویر، ماتریس هم‌رویدادی آن در چهار جهت (۰، ۴۵، ۹۰، ۱۳۵) استخراج شده و ویژگی‌های بالا از آن استخراج گردید یعنی در مجموع ۴۴ ویژگی استخراج شده است (۴ جهت $11 \times$ ویژگی).

۲-۳-۳ ویژگی‌های استخراجی از الگوی دودویی محلی

یکی از روش‌هایی که برای طبقه‌بندی بافت تصویر استفاده می‌شود، روش الگوی دودویی محلی [۱۳] است. در این روش ابتدا یک همسایگی از تصویر در نظر گرفته شده و شدت روشنایی نقاط موجود در این همسایگی با شدت روشنایی نقطه موجود در مرکز همسایگی مقایسه می‌شود. در شکل ۴ الف نمونه‌ای از همسایگی با شعاع ۱ و تعداد نقاط ۸ نشان داده شده است.

در شکل پایه‌ی این روش، الگوی باینری محلی در یک همسایگی از تصویر به صورت زیر تعریف می‌شود

$$LBP_{P,R} = \sum_{i=0}^{P-1} s(g_i - g_c) 2^i S(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases} \quad (1)$$

که در آن P نشان‌دهنده‌ی تعداد نقاط همسایه نقطه مرکزی است که شدت روشنایی آن‌ها با g_c و شدت روشنایی نقطه مرکزی آن‌ها با g_i نشان داده شده است (شکل ۴). با توجه به شکل تعریف

¹ Co-occurrence matrix

² Cluster Shade

³ Cluster prominence

۲-۴-۳ عملگرهای انتخاب

در فرایند انتخاب، افراد جمعیت بر طبق تابع برازندگی که دارند انتخاب می‌شوند. افراد جمعیت با تابع برازندگی بالاتر، با احتمال بیشتری انتخاب می‌شوند. پس از انتخاب افراد، عملگرهای همگذری و جهش برای تولید فرزندان جدید اعمال می‌شود. در این پژوهش از عملگر چرخ رولت [۱۴]، استفاده شده است.

۲-۴-۴ عملگر همگذری

در الگوریتم ژنتیک، عملگر همگذری، برای ایجاد تغییر در کروموزومها از یک نسل به نسل بعد استفاده می‌شود. عملگر همگذری، زیرمجموعه‌ای از ژنهای دو کروموزوم را با هم تعویض می‌کند و دو فرزند جدید ایجاد می‌کند. فرایند کلی همگذری با یک احتمال انجام می‌شود که این احتمال به هر جفت از والدین تعلق می‌گیرد تا فرزندان جدیدی را تولید کند. عملگر همگذری دونقطه‌ای، برای تولید جواب‌های جدید در این پژوهش به کار رفته است. در این عملگر، دو نقطه به صورت تصادفی از محل ژن‌ها انتخاب می‌گردد و زیرمجموعه‌ای از ژنهای این دو والد با یکدیگر ترکیب می‌شود.

۲-۴-۵ عملگر جهش

در الگوریتم ژنتیک، عملگر جهش برای ایجاد تغییرات کوچک به کار برده می‌شود. در اینجا از عملگر جهش یکنواخت استفاده شده است. این عملگر به صورت تصادفی برخی از ژنهای کروموزوم را با یک احتمال ثابت تغییر می‌دهد. این پژوهش، با توجه به احتمال جهش، هر یک از ژن‌ها از یک به صفر یا از صفر به یک، تغییر می‌کنند.

۲-۴-۶ تشکیل جمعیت جدید

فرایند انتخاب ویژگی توسط الگوریتم ژنتیک بدین صورت است که در ابتدا جمعیتی از کروموزومها به صورت تصادفی ایجاد می‌شود و افراد این جمعیت مورد ارزیابی قرار می‌گیرند. بعد از انجام این کار با استفاده از چرخ رولت و عملگرهای ژنتیکی، والدینی انتخاب و فرزندان جدیدی تولید می‌شوند. فرزندان به دست آمده دوباره مورد ارزیابی قرار می‌گیرند و به همراه جمعیت فعلی در استخر انتخاب برای نسل بعدی قرار می‌گیرند. تعدادی کروموزوم از بهترین افراد (نخه‌ها) این استخر به نسل بعد منتقل می‌شوند و مابقی افراد نسل بعد با استفاده از چرخ رولت گزینش می‌شوند. همین روال تا زمانی که شرط خاتمی برآورده شود ادامه می‌یابد.

۲-۵ نرمال کردن ویژگی‌ها

ویژگی‌های استخراج شده، دارای واحدهای مختلفی هستند به همین دلیل هر ویژگی به عددی بین صفر تا یک نرمال می‌شود. اگر کوچکترین عدد مربوط به یک ویژگی با \min و بزرگترین عدد با \max نشان داده شود عدد نرمال شده x به صورت زیر به دست می‌آید:

جواب‌ها را ایجاد می‌کند و سپس با استفاده از عملگرهای الگوریتم ژنتیک مانند همگذری و جهش و انتخاب، جمعیت جدیدی را شکل می‌دهد. در هر نسل از الگوریتم ژنتیک جواب‌هایی که بهترین تابع برازندگی را دارند با احتمال بیشتری برای تولید مثل انتخاب می‌شوند. روند تولید نسل‌ها تا زمانی که شرط خاتمه برآورده شود ادامه می‌یابد.

الگوریتم ژنتیک را می‌توان برای مسئله انتخاب ویژگی به کار برد زیرا آن‌ها در فضاهای جستجوی بسیار موثر عمل می‌کنند. وظیفه الگوریتم ژنتیک این است که زیرمجموعه بهینه m را از فضای جستجو با ابعاد d بیابد که $m < d$ است. برای استفاده از الگوریتم ژنتیک برای انتخاب ویژگی، در ابتدا باید ساختار مناسب برای آن تعریف شود. اجزای این ساختار عبارتند از نحوه نمایش کروموزوم، تابع برازندگی و عملگرهای همگذری، جهش و انتخاب.

۲-۴-۱ نحوه نمایش کروموزومها

در این مسئله، کروموزومها به صورت بردار باینری ژن‌ها انتخاب شده است که هر ژن نشان‌دهنده‌ی یک ویژگی است. زمانی که ژن زام برابر با یک باشد، نشان‌دهنده این است که ویژگی زام انتخاب شده است و هرگاه ژن زام برابر با صفر باشد بدین معنی است که ویژگی زام انتخاب نشده است. در شکل ۶ این نحوه نمایش نشان داده شده است.

ویژگی ۱	ویژگی ۲	ویژگی ۳	-	ویژگی n
۰	۱	۰		۰

n: تعداد تمام ویژگی‌ها ۰: ویژگی انتخاب نشده ۱: ویژگی انتخاب شده

شکل ۶ نحوه نمایش باینری برای انتخاب ویژگی توسط الگوریتم ژنتیک

۲-۴-۲ تابع برازندگی

انتخاب یک تابع برازندگی مناسب، یکی از قدم‌های اصلی در استفاده از الگوریتم ژنتیک می‌باشد. اگر یک کروموزوم دارای ویژگی‌های مناسب باشد دقت طبقه بندی با این ویژگی‌ها بالا می‌رود. به همین دلیل دقت طبقه بندی به عنوان تابع برازندگی انتخاب شده است. طبقه‌بندی‌کننده‌های زیادی می‌تواند برای طبقه‌بندی استفاده شود. دو نمونه از محبوب‌ترین این طبقه‌بندی‌کننده‌ها، KNN و SVM می‌باشد. هرچند که SVM یک طبقه‌بندی‌کننده‌ی قدرتمند است اما آموزش آن، زمان زیادی را می‌برد. به همین دلیل از KNN استفاده شده است که نسبت به SVM ساده تر و سریع تر می‌باشد. تابع شایستگی به صورت زیر تعریف می‌شود.

$$\text{تایم شایستگی} = \frac{\text{نمونه‌هایی که درست طبقه‌بندی شده‌اند}}{\text{کل نمونه‌ها}}$$

که با استفاده از آن می‌توان مشخص نمود که ۲ متغیر تا چه میزان به یکدیگر وابسته هستند. اگر هیچ ارتباطی بین دو متغیر وجود نداشته باشد (دو متغیر مستقل از هم باشند) ضریب همبستگی برابر صفر خواهد بود. اگر یک همبستگی کامل بین دو متغیر وجود داشته باشد به طوری که افزایش یک متغیر همراه با افزایش متناسب متغیر دیگر باشد مقدار ضریب معادل +۱ خواهد بود. اگر افزایش در یک متغیر، همواره با کاهش متناسب در متغیر دیگر باشد ضریب معادل با -۱ خواهد بود. ضریب همبستگی با استفاده از رابطه (۴) محاسبه می‌شود.

$$r = \frac{Cov(x, y)}{S_x \cdot S_y} \quad (4)$$

که S_x و S_y معادل با انحراف معیار برای x و y است و $Cov(x, y)$ ، کواریانس دو متغیر را نشان می‌دهد.

۳ نتایج و بحث

برای انجام آزمایش‌ها، داده‌ها به سه قسمت تقسیم شده است. ۵۰ درصد اول داده‌ها برای آموزش شبکه عصبی در مرحله اول، ۲۰ درصد دوم داده‌ها برای آموزش شبکه عصبی در قسمت ترکیب نتایج و ۳۰ درصد سوم برای داده‌های تست انتخاب شد. ارقام برنج اصل و ترکیبی، به ترتیب دو رقم هاشمی و باسماتی در نظر گرفته شد که دارای شباهت ظاهری بالایی هستند. بعد از تصویربرداری، سه مجموعه ویژگی بافتی از تصاویر استخراج شد. جهت انتخاب ویژگی‌های کارای استخراج شده، از الگوریتم ژنتیک استفاده شد. انتخاب ویژگی مبتنی بر الگوریتم ژنتیک چندین بار با استفاده از پارامترهای مختلف اجرا شد. بهترین تنظیمات برای این مجموعه داده‌ها در جدول ۱ نشان داده شده است.

جدول ۱ بهترین تنظیمات برای الگوریتم ژنتیک جهت انتخاب ویژگی

پارامتر	مقدار
جمعیت اولیه	۱۰۰
تعداد نسل‌ها	۲۰۰
نوع جهش	دو نقطه‌ای
نرخ جهش	۰,۰۵
نوع هم‌گذری	یکنواخت
تعداد نخه‌ها	۱۵

در شکل ۷، نمودار مقدار تابع برازندگی را بر حسب تعداد نسل‌ها برای انتخاب ویژگی‌های برتر الگوی دودویی محلی نشان می‌دهد. محور افقی تعداد نسل‌ها و محور عمودی مقدار تابع برازندگی می‌باشد. با توجه به نمودار، بعد از نسل ۱۲۰، بهبودی در برازندگی دیده نمی‌شود و مقدار برازندگی ثابت شده است. برای انتخاب ویژگی‌های برتر از ماتریس هیستوگرام و هم‌رویدادی نیز نمودار مشابهی دیده می‌شود. ویژگی‌های انتخاب شده توسط الگوریتم ژنتیک برای مجموعه ویژگی‌های بافتی مختلف در زیر آورده شده است.

$$y = \frac{x - \min}{\max - \min} \quad (2)$$

۲-۶ رگرسیون

همانطور که در بخش قبلی دیده شد در بازه‌های مختلف ۵ درصدی، از ارقام برنج، تصویربرداری شد و سپس بر روی تصاویر گرفته شده، ویژگی‌های بافتی استخراج گردید. در این مرحله با استفاده از ویژگی‌های استخراجی و درصد ترکیب عمل رگرسیون انجام می‌شود. به عبارت دیگر رگرسیون مدلی را ارائه خواهد کرد که با ورودی ویژگی‌های استخراجی به آن، خروجی درصد ترکیب را مشخص خواهد کرد. جهت رگرسیون از شبکه عصبی پس انتشار استفاده شده است.

۲-۷ ترکیب نتایج

با استفاده از ویژگی‌های استفاده شده از مرحله قبل، ۳ مدل به دست آمد که هر کدام دارای مقدار خطا و ضریب همبستگی مخصوص به خود بودند. به نظر می‌رسد می‌توان جهت کاهش خطا و افزایش ضریب همبستگی، نتایج حاصل از مرحله قبل با هم ترکیب شوند. به این منظور، نتایج حاصل از مرحله قبل با استفاده از شبکه عصبی با هم ترکیب شده‌اند. در واقع دو سطح شبکه عصبی در روش پیشنهادی وجود دارد. ورودی سطح اول، ویژگی‌های انتخاب شده می‌باشد و ورودی سطح دوم، خروجی سطح اول است. شمای کلی از روش پیشنهادی در شکل ۱ ارائه شده است. در روش پیشنهادی، داده‌ها به سه قسمت تقسیم شده‌اند:

قسمت اول داده‌ها، برای آموزش شبکه عصبی در سطح اول انتخاب شدند. بعد از اینکه با استفاده از قسمت اول داده‌ها، شبکه آموزش داده شد از قسمت دوم داده‌ها، جهت آموزش شبکه عصبی سطح دوم استفاده شد به این صورت که در هر مرحله، داده‌ها وارد شبکه عصبی آموزش دیده‌ی سطح اول شده و خروجی آن جهت آموزش وارد شبکه عصبی سطح دوم داده، شبکه عصبی آموزش داده شد، از داده‌های مجموعه سوم برای تست روش پیشنهادی استفاده می‌شود.

۲-۸ معیارهای ارزیابی

جهت ارزیابی کارایی الگوریتم پیشنهادی، دو معیار میانگین مقدار خطا می‌باشد که از تفاوت مقدار اصلی و مقدار پیش‌بینی شده به دست می‌آید. فرمول محاسبه مقدار خطا در زیر آمده است.

$$E = \frac{\sum_{i=1}^N |P_i - R_i|}{N} \quad (3)$$

در رابطه بالا، N تعداد نمونه‌ها، P مقدار پیش‌بینی شده، R مقدار واقعی و E مقدار خطا می‌باشد.

معیار دیگری که مورد استفاده واقع شده است ضریب همبستگی است. ضریب همبستگی شاخصی بین -۱ تا +۱ است

جدول ۲ نتایج رگرسیون با استفاده از معیارهای کارایی معرفی شده

ساختار بهینه شبکه عصبی	ضریب همبستگی	میانگین خطا	ویژگی‌های ماتریس سطح خاکستری
۳-۵-۶	۰/۸۱۵۱	۶/۳۷	ویژگی‌های ماتریس سطح خاکستری
۷-۹-۴	۰/۶۷۷۴	۸/۱۵	ویژگی‌های ماتریس هم‌رویدادی
۵-۲-۶	۰/۹۱۵۳	۴/۹۲	ویژگی‌های ماتریس الگوی دودویی باینری

در قسمت نهایی پژوهش، برای کم کردن خطا و بالا بردن ضریب همبستگی، از نتایج سه مدل قبل استفاده کرده و مدل جدیدی ارائه شده است. به این منظور از ۲۰ درصد دوم داده‌ها استفاده شد. این داده‌ها به طور مستقیم وارد شبکه عصبی سه مدل قبلی شد و خروجی آن به دست آمد یعنی در واقع این مجموعه داده‌ها، داده‌های تست برای سه مدل اول بوده‌اند. نتایج حاصل از این سه مدل ورودی شبکه عصبی بعدی (شکل ۱) بود. به عبارت دیگر، ورودی شبکه عصبی سطح دوم، خروجی شبکه‌های عصبی سطح اول و خروجی شبکه عصبی سطح دوم همان درصد ترکیب می‌باشد. نتایج ترکیب این سه مجموعه ویژگی با بهترین نتیجه در مرحله قبل در جدول ۳ آمده است. با توجه به جدول با استفاده از ترکیب نتایج، مقدار خطا ۰/۷۱ کاهش پیدا کرده است به عبارت دیگر حدود ۱۵ درصد نسبت به کمترین مقدار خطای مرحله کاهش وجود داشته است. همچنین ضریب همبستگی به مقدار ۰/۹۳۵۶ رسید.

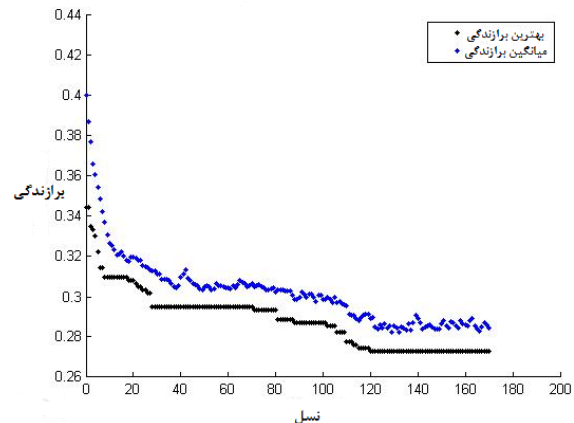
جدول ۳ مقایسه نتایج رگرسیون با استفاده از ترکیب نتایج و بهترین نتیجه از مرحله قبل

ضریب همبستگی	میانگین خطا	ترکیب نتایج
۰/۹۳۵۶	۴/۲۱	ترکیب نتایج
۰/۹۱۵۳	۴/۹۲	ویژگی‌های ماتریس الگوی دودویی باینری

۴ نتیجه گیری

محصول برنج، یکی از مهمترین محصولات غذایی در چرخه غذایی مردم ایران می‌باشد. ممکن است ارقام برنج به دلایلی مثل سودجویی با ارقام دیگر مخلوط شود. در این پژوهش، به تشخیص اصالت ارقام برنج پرداخته شد. به این منظور، با استفاده از یک جعبه سیاه در فواصل مختلف ترکیبی، تصویربرداری انجام شد. ایده اصلی تشخیص اصالت بر روی بافت توده‌ی برنج استوار بود. به این منظور با استفاده از سه روش استخراج بافت، ویژگی‌های بافتی از توده‌ی برنج استخراج شد. بعد از استخراج ویژگی‌های بافتی، برترین ویژگی‌ها با استفاده از الگوریتم ژنتیک پیدا شد. از یک شبکه عصبی جهت رگرسیون استفاده شد که ورودی آن ویژگی‌های بافتی انتخاب شده و خروجی آن درصد ترکیب می‌باشد. الگوی دودویی محلی کمترین مقدار خطا را ارائه نمود. در مرحله بعد برای کم کردن خطا، نتایج مرحله قبل با استفاده از یک شبکه عصبی ترکیب شد. مقدار خطای نهایی ۴/۲۱ و ضریب

- ویژگی‌های انتخابی از هیستوگرام: انحراف معیار، نرمی، ممان سوم، یکنواختی، آنتروپی و گروه فراوانی ۴، ۷، ۹، ۱۱، ۱۳، ۱۴، ۱۶
- ویژگی‌های انتخابی از ماتریس هم‌رویدادی: سایه خوشه در ۰، ۴۵ و ۹۰ درجه، همبستگی در ۰ و ۴۵ درجه، ماکزیمم احتمال، میانگین i و برجستگی خوشه در ۰ درجه
- ویژگی‌های انتخابی از الگوی دودویی محلی: کنتراست، همواری، یکنواختی، آنتروپی و گروه‌های فراوانی ۱، ۴، ۷، ۸، ۹، ۱۰، ۱۱، ۱۲، ۱۳، ۱۴، ۱۵، ۱۶، ۱۷ از ماتریس هیستوگرام.



شکل ۷ نمودار مقدار تابع برازندگی را بر حسب تعداد نسل‌ها

بعد از انتخاب ویژگی‌های برتر، از یک شبکه عصبی جهت رگرسیون استفاده شد. همانطور که در قسمت تصویربرداری مشاهده شد با درصد‌های مختلفی از برنج ترکیبی، تصویربرداری انجام شد. ورودی شبکه عصبی ویژگی‌های برتر انتخاب‌شده‌ی هر مجموعه ویژگی و خروجی آن درصد ترکیب می‌باشد؛ به عبارت دیگر، شبکه عصبی مدلی را به خواهد ساخت که با دادن ویژگی، درصد ترکیب را مشخص می‌نماید. در نهایت شبکه عصبی با استفاده از سه مجموعه ویژگی، سه مدل را متناظر با مجموعه ویژگی‌های هیستوگرام، هم‌رویدادی و الگوی دودویی محلی ارائه نمود. نتایج رگرسیون با استفاده از این سه مدل بر روی داده‌های تست (۳۰ درصد سوم) در جدول ۲ نشان داده شده است.

با توجه به جدول ۲، ماتریس الگوی دودویی باینری بهترین نتایج را با خطای ۴/۹۲ و ضریب همبستگی ۰/۹۱۵۳ ارائه نموده است. ماتریس هم‌رویدادی و الگوی دودویی محلی، خطاهای به مراتب بیشتری را داشتند. برای یافتن ساختار مناسب شبکه عصبی، به صورت سعی و خطا عمل شد. تعداد گره‌ها در لایه اول برابر تعداد ویژگی‌ها در مجموعه ویژگی و تعداد گره‌ها در لایه آخر یکی است که بیان‌کننده درصد ترکیب می‌باشد. برای لایه‌های میانی، ساختارهای یک، دو و سه لایه میانی با تعداد نرون‌های ۱ تا ۴۰ در هر لایه مورد آزمایش قرار گرفت. ساختار بهینه به دست آمده برای هر مجموعه ویژگی در جدول ۲ آمده است.

- [10] R. C. Gonzalez and E. Richard, "Woods, digital image processing", ed: Prentice Hall Press, ISBN 0-201-18075-8, 2002.
- [11] R. C. Gonzalez, Digital Image Processing, 2008.
- [12] R. M. Haralick, K. Shanmugam, and I. H. Dinstein, "Textural features for image classification," Systems, Man and Cybernetics, IEEE Transactions on, vol. 3, pp. 610-621, 1973.
- [13] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 24, pp. 971-987, 2002.
- [14] T. Bäck and I. o. Physics, Evolutionary computation: Institute of Physics, 2000.

سید جلال الدین موسوی راد در سال ۱۳۶۵ در نیشابور متولد شد. او تحصیلات کارشناسی و کارشناسی ارشد خود را در دانشگاه آزاد مشهد و کردستان گذراند. او هم اکنون دانشجوی دکتری مهندسی کامپیوتر در دانشگاه کاشان می‌باشد. علاقمندی تحقیقاتی ایشان پردازش تصویر، شناسایی آماری الگو و الگوریتم‌های فراتکاملی است.



فردین اخلاقیان طاب در سال ۱۳۴۴ در سنندج متولد شد. او مقاطع کارشناسی، کارشناسی ارشد، و دکترای خود را در دانشگاه صنعتی اصفهان، تربیت مدرس تهران، و دانشگاه ولونگنگ استرالیا گذراند. او هم اکنون استادیار دانشگاه کردستان می‌باشد. منطقه تحقیقاتی مورد علاقه ایشان پردازش تصویر، بینایی ماشین، شناسایی آماری الگو و پردازش سیگنال می‌باشد. از ایشان بیش از هفتاد مقاله در مجلات و کنفرانس‌های معتبر بین‌المللی و داخلی منتشر شده است.



همبستگی ۰/۹۳۵۶ حاصل این ترکیب می‌باشد. نتایج این پژوهش می‌تواند در ساخت یک سیستم اصالت‌سنجی ارقام برنج مورد استفاده قرار بگیرد. با توجه به تجربیاتی که در این پژوهش به دست آمد استخراج ویژگی‌های بافتی دیگر مثل تبدیل موجک و ویژگی‌های رنگی جهت ادامه کار پیشنهاد می‌گردد.

مراجع

- [1] Fao, "Rice Market Monitor," vol. XV, 2012.
- [2] Z.-y. Liu, F. Cheng, Y.-b. Ying, and X.-q. Rao, "Identification of rice seed varieties using neural network," Journal of Zhejiang University. Science. B, vol. 6, p. 1095, 2005.
- [3] Z. Liu, F. Cheng, Y. Ying, and X. Rao, "Identification of rice seed varieties using neural network," Journal of Zhejiang University. Science. B, vol. 6, p. 1095, 2005.
- [4] B. Verma, "Image processing techniques for grading & classification of rice," in International conferences on computer and communication technology, India, 2010, pp. 220-223.
- [5] M. Jinorose, S. Prachayawarakorn, and S. Soponronnarit, "Development of a computer vision system and novel evaluation criteria to characterize color and appearance of rice," Drying Technology, vol. 28, pp. 1118-1124, 2010.
- [6] G. Van Dalen, "Determination of the size distribution and percentage of broken kernels of rice using flatbed scanning and image analysis," Food research international, vol. 37, pp. 51-58, 2004.
- [7] L. Pabamalie and H. Premaratne, "An intelligent rice quality classifier," International Journal of Internet Technology and Secured Transactions, vol. 3, pp. 386-406, 2011.
- [8] F. A. T. S.J. Mousavirad, K. Mollazade, "Real Time Identification of Rice Varieties Using Texture Features Based on Classifier Fusion Method," in the 16th CSI International Symposium on Artificial Intelligence and Signal Processing, Shiraz University, Iran, 2012.
- [9] F. A. T. S.J. Mousavirad, K. Mollazade, "Identification of Rice Varieties Using Co-occurrence Matrix Features of Bulk Samples and Support Vector Machine," presented at the 20th Iranian Conference on Electrical Engineering, University of Tehra, Teharn, Iran, 2012.