

به کارگیری رویکردهای یادگیری ماشین جهت پیش‌بینی انحراف ابعاد کاشی‌های سرامیکی

مرضیه السادات طباطبائی^۱ مهدی یزدیان دهکردی^۲ امیر جهانگرد رفسنجانی^۳

۱- دانشجوی دکتری- دانشکده مهندسی کامپیوتر- دانشگاه یزد- یزد- ایران

ms.tabatabaei@stu.yazd.ac.ir

۲- استادیار- دانشکده مهندسی کامپیوتر- دانشگاه یزد- یزد- ایران

yazdian@yazd.ac.ir

۳- استادیار- دانشکده مهندسی کامپیوتر- دانشگاه یزد- یزد- ایران

jahangard@yazd.ac.ir

چکیده: امروزه، رویکردهای یادگیری ماشین نقش مهمی را در شناسایی عوامل مؤثر بر کیفیت محصولات تولیدی از جمله تولید کاشی و سرامیک ایفا می‌کنند. یکی از چالش‌های موجود در تولید کاشی و سرامیکی، معیوب شدن کاشی‌ها بدلیل ایجاد انحراف در ابعاد کاشی تولیدی است. در صورتیکه بتوان با توجه به پارامترهای فرآیند تولید، امکان ایجاد انحراف در ابعاد کاشی را قبل از تولید پیش‌بینی نمود، می‌توان از تولید کاشی معیوب جلوگیری و نسبت به تنظیم مجدد پارامترهای تولید اقدام نمود. در این پژوهش، یک سیستم خودکار جهت پیش‌بینی دسته‌ی انحراف کاشی و شناسایی عوامل تأثیرگذار بر آن، پیشنهاد شده است. بدین منظور سه طبقه‌بند مختلف شامل رگرسیون منطقی، جنگل تصادفی و ماشین بردار پشتیبان جهت مدل‌سازی پارامترهای مربوطه مورد بررسی قرار گرفته و برترین ساختار معرفی شده است. علاوه بر این با بررسی چند دسته ویژگی و بهره‌گیری از روش انتخاب ویژگی پیش‌رو، متغیرهای مؤثر در تصمیم‌گیری انحراف کاشی نیز شناسایی شده‌اند. نتایج آزمایش‌های انجام‌شده بر روی نمونه‌های واقعی، نشان می‌دهد که رویکرد جنگل تصادفی کارایی بهتری نسبت به رویکردهای دیگر داشته و تأثیرگذارترین پارامترها در ایجاد انحراف کاشی، مقدار نامناسب دماهای کوره بوده است.

واژه‌های کلیدی: دسته‌بندی خودکار، انحراف ابعاد کاشی، یادگیری ماشین، جنگل تصادفی، ماشین بردار پشتیبان، رگرسیون منطقی، انتخاب ویژگی پیش‌رو.

نوع مقاله: پژوهشی

DOI: 10.52547/jiaeee.19.2.199

تاریخ ارسال مقاله: ۱۳۹۸/۰۹/۰۹

تاریخ پذیرش مشروط مقاله: ۱۳۹۹/۰۹/۰۵

تاریخ پذیرش مقاله: ۱۳۹۹/۱۲/۰۲

نام نویسنده‌ی مسئول: دکتر مهدی یزدیان دهکردی

نشانی نویسنده‌ی مسئول: ایران- یزد - دانشگاه یزد - پردیس فنی و مهندسی - دانشکده مهندسی کامپیوتر

۱- مقدمه

در سال‌های اخیر، با توجه به گسترش شهرنشینی و ضرورت پاسخ‌گویی به نیازهای رو به رشد مصرفی در حوزه‌ی مسکن و ساختمان، فن‌آوری‌های تولید انبوه محصولات سرامیکی وارد ایران شده‌اند. با افزایش رقابت در بازار، صنعت کاشی نه تنها با رقابت شدید داخلی بلکه با رقابت در بازار بین‌المللی نیز روبرو شده است. امروزه شرکت‌های کاشی و سرامیک جهت برقراری تعادل و حضور مستمر در بازارهای داخلی و بین‌المللی، سعی در ارتقای کیفیت محصولات تولیدی دارند [۱]. یکی از روش‌های مطرح‌شده در جهت افزایش کیفیت محصولات تولیدی، استفاده از رویکرد الگوبندی پارامترهای کیفی است. کیفیت مطلوب، در نتیجه‌ی روش و فن‌آوری مطلوب صورت می‌گیرد. بدین منظور، روش‌های مختلفی وجود دارند که از میان آنها، روش‌های یادگیری ماشین در دهه‌های اخیر مورد توجه پژوهشگران قرار گرفته‌اند.

الگوریتم‌های یادگیری ماشین، الگوریتم‌های کامپیوتری هستند که از طریق یادگیری و مدل‌سازی بر اساس نمونه داده‌های آموزشی از آن مسئله عمل می‌کنند [۲]. فرآیند یادگیری، با مشاهده‌ی مجموعه داده‌های دارای برچسب مشخص‌شده توسط یک خبره انسانی و ساخت یک مدل برای داده‌ها انجام می‌شود. این مدل با استخراج الگوهای موجود در داده‌ها امکان پیش‌بینی داده‌های بدون برچسب را در آینده فراهم خواهد کرد. تاکنون تحقیقات متعددی بر روی به کارگیری روش‌های یادگیری ماشین در کاربردهایی مختلف صنعتی مانند آشکارسازی نقص در فرآیند بازرسی جوش [۳]، رفع نواقص مربوط به نامشخص بودن پلاک خودروها و آشکارسازی محل پلاک [۴]، دسته‌بندی ارقام برنج [۵] و همچنین تشخیص عیب و دسته‌بندی کاشی‌های سرامیکی انجام شده است. در ادامه برخی از مطالعات انجام شده در حوزه کاشی و سرامیک بررسی شده است.

انتشار صوتی یک روش بسیار کارآمد است که در صنعت برای آزمایش محصولات استفاده می‌شود. در سال ۲۰۰۵ فنگ و همکارانش [6]، یک روش شبکه عصبی مصنوعی^۱ را جهت بررسی تأثیر ارتعاشات صوتی بر روی کاشی‌های دیواری و دسته‌بندی آنها از نظر میزان ارتعاش صوت ارائه کردند. هرچند استفاده از شبکه‌های عصبی امکان تطبیق‌پذیری با داده‌ها را فراهم می‌کند، با این وجود برای رسیدن به یک مدل کارا تعداد نمونه‌های آموزشی زیادی لازم است و زمان یادگیری شبکه نیز طولانی است.

کنها و همکاران [7] نیز، دو رویکرد متفاوت ماشین بردار پشتیبان^۲ و k-نزدیک‌ترین همسایه^۳ را برای طبقه‌بندی کیفی کاشی‌های سرامیکی بر مبنای انتشار صوت، ارائه داده‌اند. نتایج بدست‌آمده در مرحله‌ی آزمون، دقت بین ۹۵ تا ۹۹/۹ درصد را نشان می‌دهد.

در فرآیند تولید محصولات در مراحل پیش و پس از تولید، ممکن است عیوب متفاوتی در آنها نمایان شود. اطلاع از انواع نقص‌ها در

جهت کنترل پارامترهای مربوط به آن، تأثیر بسزایی در پیش‌گیری از ایجاد نقص و افزایش کیفیت محصول دارد. بدین منظور، رویکردهایی در جهت شناسایی انواع مختلف نقص‌ها معرفی شده است. در ادامه برخی از این رویکردها معرفی می‌گردند.

جهت رفع نواقص مربوط به کاشی‌های سرامیکی، البهیری و همکاران [۸]، یک رویکرد دو مرحله‌ای برای تشخیص نقص بر روی سطوح کاشی‌های سرامیکی ارائه داده‌اند. بدین صورت که در مرحله‌ی اول، از تصاویر کاشی‌های تولیدی به‌عنوان ورودی استفاده کرده و پس از آن، هیستوگرام بلاک‌های تصویر به‌عنوان خروجی نمایش داده می‌شوند. در مرحله‌ی دوم، از خروجی قسمت اول به‌عنوان ورودی استفاده کرده و از تصویر مکمل جهت شناسایی نقص استفاده می‌شود. این روش به دلیل اعمال عملیات پردازش تصویر مکمل، نیازمند انجام عملیات اضافی است و با کاهش سرعت تشخیص همراه است.

شارما و همکاران [۹] نیز، از ترکیب ویژگی‌های مرتبط با رنگ و بافت کاشی‌ها جهت تشخیص نقص استفاده کرده‌اند. در این مطالعه، بمنظور استخراج ویژگی‌های بافت کاشی، ماتریس رخداد همزمان پیشنهاد شده‌است. پژوهش فوق در سه فضای رنگی مجزا صورت گرفته و در هر فضا، طبقه‌بندهای ماشین بردار پشتیبان، k-نزدیک‌ترین همسایه و بیزین بصورت مجزا برای مدل‌سازی، بکار گرفته شده‌اند. طبق نتایج بدست‌آمده، در هر سه فضا، مدل بیزین بیشترین دقت را نسبت به سایر مدل‌ها داشته است.

در پژوهش دیگر پاتیل و همکاران [۱۰] از رویکرد بینایی ماشین برای تشخیص نقص در کاشی‌های سرامیکی استفاده کرده‌اند. در این روش تصاویر کاشی‌های سرامیکی پردازش و طبقه‌بندی شده و نقص احتمالی موجود در آنها تشخیص داده می‌شود. روش پیشنهادی در پژوهش فوق، شامل سه مرحله‌ی اصلی است. اولین مرحله، پیش-پردازش تصاویر است. در مرحله‌ی دوم، از یکسری عملیات پردازشی برای تشخیص نقص در تصاویر استفاده می‌شود. در مرحله‌ی سوم نیز، به طبقه‌بندی انواع مختلف نقص پرداخته می‌شود. براساس آزمایش‌های انجام شده، رویکرد پیشنهادی در تشخیص شش نوع نقص موفق بوده و قادر به شناسایی نقص‌های مربوط به لعاب و خطاهای خراش نیست.

رامادهان و همکاران [۱۱]، از رویکرد کاهش ابعاد به‌همراه رویکرد شبکه عصبی مصنوعی جهت طبقه‌بندی کیفیت سطوح کاشی‌های سرامیکی استفاده کرده‌اند. آنها کاشی‌ها را از نظر سطح به پنج کلاس مجزا شامل: عادی، برش، ترک، لکه‌های خشک و خراش تقسیم کرده‌اند. با انجام آزمایش بر روی مجموعه داده‌ی متشکل از ۴۲۵۶ تصویر، دقت طبقه‌بندی ۹۰/۱۳ درصد حاصل شد.

در مطالعه‌ای تحت عنوان مدل‌هایی برای تشخیص نقص در کاشی‌های سرامیکی، از رویکرد تجزیه و تحلیل تصاویر سه بعدی جهت مکان‌یابی خودکار نقایص موجود در سطح کاشی‌ها استفاده شده است [۱۲]. در این مطالعه، برای ساخت تصاویر سه بعدی از پارامترهای

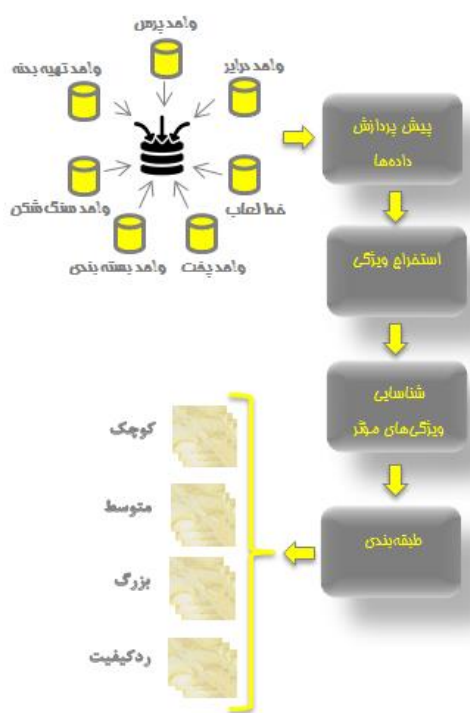


است؛ اما اینکه برخی انحراف کوچک و برخی دیگر انحراف بزرگ داشته باشند، این ناهمگن بودن قابل پذیرش نیست.

بطور معمول در اکثر قریب به اتفاق موارد رد کیفیت، به معنای غیر قابل قبول بودن میزان انحراف کاشی است. به این ترتیب با پیش‌بینی دسته انحراف کاشی، این امکان فراهم می‌شود که در صورت نیاز، کارشناسان نسبت به اصلاح مقادیر ویژگی‌های ورودی اقدام نمایند.

بطور خلاصه سهم علمی این تحقیق را می‌توان در موارد زیر خلاصه نمود:

- در این پژوهش با مدل‌سازی یک پیش‌بینی‌کننده‌ی انحراف کاشی بر اساس پارامترهای مربوط به تولید مانند درصد آب، رطوبت، ضخامت و دمای کوره، امکان انحراف در ابعاد را قبل از تولید کاشی، پیش‌بینی نموده تا بتوان از تولید کاشی معیوب جلوگیری کرد.
- تا آنجایی که بررسی‌های ما نشان می‌دهد، تاکنون مطالعه‌ای در زمینه‌ی پیش‌بینی انحرافات ابعادی کاشی‌های تولیدی قبل از تولید کاشی بر اساس روش‌های هوشمند، انجام نشده است.
- جهت بهبود فرآیند طبقه‌بندی کیفی کاشی‌ها چندین طبقه‌بندی‌کننده‌ی مختلف ارزیابی و کاراترین طبقه‌بندی‌کننده در بین آن‌ها مشخص شده است. همچنین چندین دسته ویژگی ارزیابی شده و متغیرهای مؤثر در افزایش دقت تصمیم‌گیری انحراف کاشی نیز شناسایی شده‌اند. اعمال این رویکرد به‌گونه‌ای بوده است که دقت تشخیص بر روی داده‌ها از مرز ۹۹ درصد هم فراتر می‌رود.



شکل (۱): بلوک دیاگرام روش پیشنهادی

تصویربرداری سطح کاشی به همراه روش مثلث‌سازی لیزر^۴ استفاده شده است. رویکرد پیشنهادی شامل الگوریتم‌های تجزیه و تحلیل تصاویر سه بعدی مربوط به سطح کاشی‌ها جهت جستجو، مکان‌یابی و ارزیابی نقص سطح کاشی است.

در راستای تلاش‌هایی که جهت افزایش دقت و سرعت تشخیص وجود نقص در کاشی‌های سرامیکی انجام شده است، کاره و همکاران [۱۳]، از یک روش مبتنی بر یافتن مؤلفه‌های همبند جهت طبقه‌بندی کاشی‌های سرامیکی استفاده کرده‌اند. این روال، بر اساس تحلیل رنگ بافت کاشی‌ها عمل کرده و عملیات مورفولوژی^۵ را شامل می‌شود. روش ارائه‌شده امکان شناسایی تعدادی عیوب خاص را دارد.

بطور معمول در تولیدات محصولات سرامیکی، کاشی درجه یک بالاترین کیفیت را داراست و اعداد بزرگ‌تر، کاشی‌ها با کیفیت پایین‌تر را نشان می‌دهند. یکی از نیازهای اساسی در صنعت سرامیک، تعیین نوع درجه کاشی است به همین دلیل طراحی یک سیستم درجه‌بندی خودکار، می‌تواند نقش مهمی را برای ارتقاء این صنعت ایفا کند.

در مطالعه‌ای سنتیکمار و همکاران [۱۴]، یک سیستم درجه‌بندی خودکار را برای کاشی‌های سرامیکی ارائه داده‌اند. در این سیستم، از تبدیل موجک گسسته^۶ جهت استخراج ویژگی‌های مربوط به بافت کاشی و از ماتریس کوواریانس سطح خاکستری^۷ برای استخراج ویژگی‌های آماری استفاده شده است. برای طبقه‌بندی نیز از تلفیق ویژگی‌ها به همراه طبقه‌بند شبکه عصبی پیش‌رو و k-نزدیک‌ترین همسایه، استفاده شده است. نتایج تجربی دقت مطلوب رویکرد پیشنهادی با استفاده از طبقه‌بند شبکه عصبی را نشان می‌دهد.

در مطالعه‌ی دیگری راجا [۱۵]، یک سیستم درجه‌بندی خودکار برای کاشی‌های سرامیکی ارائه کرده است. در این سیستم، از شبکه عصبی کانولوشن مبتنی بر رویکرد یادگیری عمیق جهت درجه‌بندی سطوح کاشی‌ها استفاده شده است. نتایج تجربی دقت طبقه‌بندی ۹۶/۱۷ درصد را نشان می‌دهند.

در این پژوهش قصد داریم با به‌کارگیری روش‌های یادگیری ماشین، پارامترهای مؤثر در ایجاد انحرافات ابعادی کاشی‌های سرامیکی را شناسایی کرده و مدلی برای پیش‌بینی دسته‌ی انحراف کاشی قبل از تولید آن را ارائه دهیم. پارامترهای ورودی مورد استفاده در سیستم پیشنهادی، شامل مقادیر عددی هستند که توسط کارشناسان در قسمت‌های مختلف خطوط تولید اندازه‌گیری شده‌اند. پارامتر خروجی مورد انتظار، ابعاد کاشی‌های تولیدی است. بدین‌صورت کارشناسان، قبل از تولید محصول، انحراف ابعادی کاشی‌های تولیدی را پیش‌بینی کرده و بر اساس نتیجه پیش‌بینی میزان انحراف، تصمیم‌گیری مناسب را با توجه به سیاست‌های مدیریتی آن مجموعه اتخاذ می‌نمایند. بعنوان مثال، در برخی موارد ممکن است انحراف متوسط قابل پذیرش باشد و در برخی مواقع غیر قابل قبول باشد. یا بعنوان مثال دیگر در یک ساختمان اگر تمام کاشی‌ها دارای انحراف کوچک باشند، پذیرفته

۲- روش پیشنهادی

در راستای اجرای پژوهش حاضر، روشی چهار مرحله‌ای پیشنهاد داده شده است. در مرحله اول، از پیش‌پردازش داده‌های جمع‌آوری شده جهت مرتفع‌سازی خطاهای احتمالی موجود در داده‌ها استفاده می‌شود. در مرحله دوم، از ساختار اولیه و ساختارهای اصلاح شده ویژگی‌ها برای شناسایی و استخراج ویژگی‌های مؤثر اولیه استفاده می‌شود. در مرحله سوم نیز، به شناسایی و معرفی ویژگی‌های مؤثر نهایی پرداخته و در مرحله آخر، سه طبقه‌بند مختلف برای پیش‌بینی میزان انحراف کاشی‌های تولیدی معرفی می‌شوند. مراحل فوق در شکل (۱) نمایش داده شده و در ادامه به ترتیب تشریح می‌شوند.

۲-۱- پیش‌پردازش داده‌ها

قبل از استفاده از داده‌های مورد نظر، اقداماتی در جهت آماده‌سازی و رفع چالش‌های موجود در آنها صورت می‌گیرد. آماده‌سازی داده‌ها، یکی از مهمترین مراحل کار با داده‌های حجیم است. از جمله اقداماتی که در این مرحله انجام می‌گیرد، نرمال‌سازی داده‌ها است. هدف از نرمال‌سازی، مرتفع‌سازی تفاوت‌های ناشی از اثرات ناخواسته در مجموعه داده‌ها است. این تفاوت‌ها، ممکن است در مقدار پارامترهای اندازه‌گیری شده از نمونه‌های انتخابی مشاهده گردیده و اثرات نامطلوبی را در دقت پژوهش به همراه آوردند [۱۶].

از دیگر مشکلات موجود در بانک‌های اطلاعاتی که شبکه‌ها را دچار خطا می‌کنند؛ عدم وجود برخی داده‌ها و همچنین وجود اغتشاش در آنها است [۱۷، ۱۸]. جهت جلوگیری از وقوع چنین خطاهایی، رکوردهای حاوی مقادیر فوق از مجموعه بانک اطلاعاتی موجود حذف می‌شوند. همچنین برخی ویژگی‌های نامرتب در تعیین ابعاد کاشی‌های تولیدی، پس از مشورت با افراد خبره و برخی نیز توسط روش همبستگی^۸، از مجموعه داده‌ها حذف می‌گردند.

۲-۲- استخراج ویژگی

در این مرحله، از تعیین تمامی ویژگی‌های دخیل در فرآیند تولید جهت برقراری ارتباط بین مقادیر ویژگی‌های ثبت‌شده از یک نمونه کاشی و برچسب خروجی همان نمونه، استفاده می‌شود. این ویژگی‌ها با ورود نمونه‌های جدید به‌عنوان داده‌های ورودی، باید قادر به تشخیص درست برچسب مناسب برای خروجی آن نمونه باشند.

۲-۳- شناسایی ویژگی‌های مؤثر

پس از استخراج ویژگی‌ها، به تشخیص ویژگی‌های مؤثر تعیین‌کننده خروجی کیفی مدنظر پرداخته می‌شود. بدین منظور برخی ویژگی‌های اولیه اصلاح و حذف می‌شوند. همچنین به تحلیل و ارزیابی نتایج بر روی ویژگی‌های اولیه و ویژگی‌های اصلاح‌شده پرداخته می‌شود.

۲-۴- طبقه‌بندی

در این پژوهش، از سه رویکرد جنگل تصادفی^۹، ماشین بردار پشتیبان و رگرسیون منطقی^{۱۰} جهت مدل‌سازی و تخمین مقادیر پارامتر ابعاد کاشی‌های تولیدی استفاده می‌شود. در ادامه به تشریح این سه رویکرد پرداخته می‌شود.

۲-۴-۱- جنگل تصادفی

جنگل تصادفی یکی از روش‌های یادگیری ماشین، با ساختاری حاوی تعدادی درخت تصمیم است. در این روش، درخت‌های تصمیم به‌صورت مجزا از یکدیگر و با استفاده از مجموعه آموزش تولیدشده و با روش‌های مختلف ترکیب و ارزیابی می‌شوند [۱۹-۲۱]. از مهم‌ترین دلایل استفاده از رویکرد جنگل تصادفی بعنوان جایگزین درخت تصمیم، رفع مشکل بیش‌برازش است [۲۲، ۲۳].

۲-۴-۲- ماشین بردار پشتیبان

ماشین بردار پشتیبان یک طبقه‌بند دودویی است که با استفاده از بردارهای پشتیبان یک ابر صفحه بهینه را یافته و داده‌های یک کلاس را از کلاس دیگر جدا می‌کند [۲۴]. این طبقه‌بند سعی می‌کند داده‌های دو کلاس را به گونه‌ای از یکدیگر جدا کند که بیشترین حاشیه را از خط جدا کننده داشته باشند. به همین علت ماشین بردار پشتیبان معروف به داشتن عمومیت در داده‌ها است. هنگامی که این کلاس‌ها بصورت غیرخطی قابل تفکیک باشند، می‌توان از توابع هسته متفاوتی برای نگاشت بردارهای ورودی در فضای با ابعاد بالاتر استفاده کرد.

۲-۴-۳- رگرسیون منطقی

طبقه‌بند رگرسیون منطقی یک روش پرکاربرد در زمینه‌های مختلف است که سعی می‌کند تابع احتمال پسین^{۱۱} هر کلاس را تخمین زده و کلاس با بیشترین احتمال را انتخاب کند. در مدل رگرسیون منطقی احتمال وقوع یک کلاس در محدوده ۰ تا ۱ برآورد می‌شود [۲۵].

۳- ارزیابی نتایج

در پژوهش حاضر، از زبان برنامه‌نویسی پایتون بمنظور پیاده‌سازی روش پیشنهادی، استفاده شده است. همچنین، جهت بررسی صحت مدل پیشنهادی، رویکرد اعتبارسنجی 10-Times-10-Fold به‌کار گرفته شده است. انتخاب داده‌ها نیز، به‌صورت تصادفی و از هر چهار برچسب خروجی سیستم صورت می‌گیرد. انتخاب داده‌ها به‌صورت تصادفی دارای عملکرد توزیع یکنواخت بوده و تأثیرات نامطلوب نحوه توزیع داده‌ها برای فرآیند مدل‌سازی را مرتفع می‌نماید [۲۶]. در این قسمت به پیاده‌سازی مراحل روش پیشنهادی و ارزیابی نتایج پیاده‌سازی پرداخته می‌شود.



۳-۱- توصیف و آماده‌سازی داده‌ها

ساختار بانک اطلاعات مورد استفاده در این تحقیق، شامل یک ماتریس ورودی با ابعاد ۷۱×۴۷۸ و یک بردار خروجی با ابعاد ۱×۴۷۸ است. ماتریس ورودی، داده‌های مربوط به خط تولید کاشی کف یک کارخانه تولیدی هستند. بر طبق نظرات افراد خبره، مجموعه داده فوق، جامع و همراه با ویژگی‌های تعیین‌کننده درزمینه‌ی تحقیق، تشخیص و بررسی کیفیت کاشی‌های تولیدی است. بردار خروجی نیز، حاوی اطلاعات مربوط به ابعاد کاشی‌های تولیدی با ۴ برجسب مختلف برای تعداد ۴۷۸ نمونه کاشی است. هریک از این برجسب‌ها نشان‌دهنده‌ی این است که کاشی تولیدی مربوطه برحسب تشخیص کارشناسان، به کدام قسمت باید انتقال یابد. اسامی کلاس‌ها و محدوده ابعادی که در این دسته‌ها قرار می‌گیرند در جدول (۱) نشان داده شده است.

جدول (۱): محدوده اندازه و کلاس کاشی‌های تولیدی

کلاس	محدوده ابعاد (برحسب میلی‌متر)
Small	۴۹۷/۹ - ۴۹۹/۰
Medium	۴۹۹/۱ - ۵۰۰/۱
Large	۵۰۰/۲ - ۵۰۱/۱
Reject	< ۴۹۷/۹ و > ۵۰۱/۱

همان‌طور که در قسمت پیش‌پردازش داده‌ها اشاره شد، پس از حذف داده‌های مشکل‌ساز و وابسته، جهت به‌کارگیری داده‌ها در مدل پیشنهادی، باید آنها را نرمال نمود. پس از نرمال‌سازی مقادیر ستون‌های داده‌ای و پیاده‌سازی مراحل بیان‌شده در قسمت پیش‌پردازش داده‌ها، اندازه بانک اطلاعات نهایی به ابعاد ۳۲۴×۵۶ کاهش پیدا کرده است. پارامترهای ورودی حذف‌شده عبارتند از: اندازه قالب پرس، فشار پرس، سیکل، کد شناسه، کد طرح و تعداد ده عدد از پارامترهای مربوط به برخی دماهای کوره.

پارامتر محدود ابعاد که خروجی تولید را شامل می‌شود، نیاز به پیش‌بینی دارد. بعنوان مثال، در صورتیکه در انتهای فاز مدل‌سازی، پیش‌بینی شود که کاشی بر اساس اندازه‌های موجود در جدول (۱) به دسته‌ی reject می‌رود، بایستی مراحل تولید توسط کارشناسان متوقف شده و مقادیر پارامترهای فرآیند تولید مجدداً تنظیم و اصلاح گردند. بدین‌صورت، خروجی قبل از تولید تغییر پیدا کرده و بهینه می‌شود.

۳-۲- معیارهای ارزیابی

جهت تحلیل و ارزیابی نتایج اعتبارسنجی مدل‌های پیشنهادی، نیاز به تعیین معیار ارزیابی است. معیارهای ارزیابی مورد استفاده در این پژوهش، عبارتند از: دقت (Accuracy)، صحت (Precision)، بازخوانی (Recall) و معیار F (F-Measure).

رابطه (۱) جهت بدست آوردن مقدار میانگین دقت پیش‌بینی، به ازای تمامی کلاس‌های استفاده‌شده در مدل پیشنهادی بکار می‌رود [۲۷].

$$Accuracy = \frac{1}{l} \sum_{i=1}^l \frac{tp_i + tn_i}{tp_i + fp_i + tn_i + fn_i} \quad (1)$$

که در آن tp_i تعداد رکوردهایی است که در کلاس i به‌صورت صحیح، مثبت تشخیص داده می‌شوند؛ tn_i تعداد رکوردهایی است که در کلاس i به‌صورت صحیح، منفی تشخیص داده می‌شوند؛ fp_i تعداد رکوردهایی است که در کلاس i اشتبهاً مثبت تشخیص داده می‌شوند؛ fn_i تعداد رکوردهایی است که در کلاس i اشتبهاً منفی تشخیص داده می‌شوند و l نیز تعداد کل نمونه‌ها است.

رابطه (۲)، رابطه‌ی بدست آوردن مقدار میانگین بازخوانی مدل پیشنهادی به ازای هر یک از کلاس‌ها است [۲۷]. از این رابطه جهت بدست آوردن نسبت موارد طبقه‌بندی‌شده به‌صورت صحیح از یک کلاس به تعداد موارد حاضر در کلاس مذکور استفاده می‌گردد.

$$Recall = \frac{1}{l} \sum_{i=1}^l \frac{tp_i}{tp_i + fn_i} \quad (2)$$

مقدار میانگین بازخوانی ممکن است، به دلیل ضعیف بودن مدل پیشنهادی، زیاد شود. جهت رفع این مشکل، در کنار معیار بازخوانی از معیار دیگری به نام صحت نیز استفاده می‌شود. بدین منظور، از رابطه (۳) جهت بدست آوردن میزان صحت مدل پیشنهادی به ازای هریک از کلاس‌های طبقه‌بندی‌شده، استفاده می‌شود [۲۷].

$$Precision = \frac{1}{l} \sum_{i=1}^l \frac{tp_i}{tp_i + fp_i} \quad (3)$$

در برخی موارد، معیارهای بازخوانی و صحت باهم رشد نکرده و قادر به افزایش دقت کلاس‌بندی نیستند [۲۹، ۲۸]. جهت حل مشکل فوق، از معیار F در کنار معیارهای پیشین استفاده می‌شود. این معیار، توصیف‌کننده میانگین وزن دار مابین دو کمیت بازخوانی و صحت است؛ بنابراین، استفاده از آن سبب افزایش دقت ارزیابی کلاس‌بندی می‌گردد [۲۹]. معیار F از طریق رابطه (۴) محاسبه می‌گردد [۲۸، ۲۷].

$$F - Measure = \frac{2 * Precision * Recall}{Precision + Recall} \quad (4)$$

که در آن Prc_M میانگین ماکروی صحت و Rec_M میانگین ماکروی بازخوانی است.

مقادیر معیارهای فوق، بر اساس ماتریس درهم‌ریختگی محاسبه می‌شوند. از ماتریس درهم‌ریختگی، جهت رفع مشکل عدم توازن برجسب‌های خروجی استفاده می‌شود [۳۰]. در مطالعه‌ی حاضر، تحلیل و مقایسه مدل‌های پیاده‌سازی شده بر اساس معیارهای ماتریس درهم‌ریختگی با تعداد چهار کلاس مجزا صورت می‌گیرد.

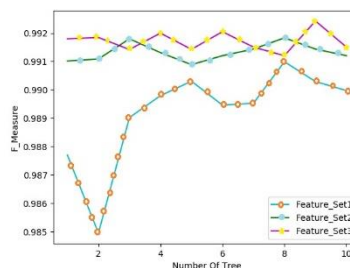
۳-۳- تحلیل مجموعه ویژگی‌های مؤثر اولیه و طبقه‌بندی کننده

در این تحقیق، جهت بررسی ویژگی‌های مؤثر در تعیین ابعاد کاشی‌های تولیدی، از طراحی سه سناریوی متفاوت استفاده می‌شود. در سناریوی اول، از تمامی متغیرهای مستخرج در مرحله آماده‌سازی داده‌ها جهت آموزش مدل‌ها استفاده می‌گردد. در این راستا معیارهای ارزیابی با تعداد ۳۲۴ رکورد آموزش و ۵۶ ویژگی در چهار سطح مربوط به ابعاد کاشی‌های تولیدی، مورد بررسی قرار گرفته است. متخصصان، ارتباط تغییر ابعاد کاشی‌های تولیدی با ویژگی‌های اندازه‌گیری شده‌ی پس از مرحله قالب‌گیری از نمونه اولیه در قسمت پرس و نیز پارامترهای مربوط به دماهای مختلف واحد پخت و کوره را محتمل‌تر می‌دانند. بدین جهت در سناریوهای دوم، معیارهای ارزیابی با تعداد ۳۲۴ رکورد آموزش و ۳۳ ویژگی در چهار سطح مربوط به ابعاد کاشی‌های تولیدی، مورد بررسی قرار گرفته است. این پارامترها عبارتند از: دمای سطح، درصد آب، دانسیته، ویسکوزیته و وزن مربوط به بخش‌های پس از مرحله پرس و در نهایت دمای قسمت‌های مختلف کوره.

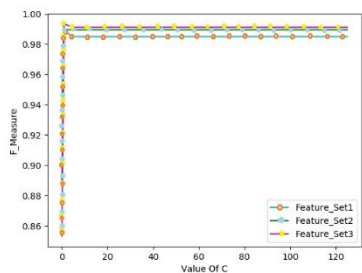
در سناریوی سوم نیز، معیارهای ارزیابی با تعداد ۳۲۴ رکورد آموزش و ۲۰ ویژگی در چهار سطح مربوط به ابعاد کاشی‌های تولیدی، مورد بررسی قرار گرفته است. هریک از این ویژگی‌ها، دمای قسمت‌های مختلف کوره را نشان می‌دهند. بدین‌صورت که تمامی پارامترهای مربوط به قبل از ورود به مرحله پخت از مجموعه داده حذف می‌گردند. نتایج محاسبه معیارهای فوق در جدول (۲) قابل مشاهده است.

جدول (۲): نتایج ارزیابی سناریوهای انتخابی

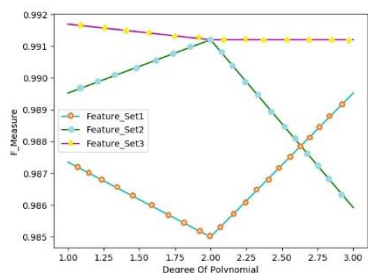
سناریو	مدل	دقت	بازخوانی	صحت	معیار F
اول	جنگل تصادفی	۰/۹۸۹۹	۰/۹۸۸۵	۰/۹۹۱۷	۰/۹۸۹۱
	ماشین بردار پشتیبان	۰/۹۸۲۰	۰/۹۸۲۴	۰/۹۷۸۹	۰/۹۷۷۹
	رگرسیون منطقی	۰/۹۸۸۲	۰/۹۸۷۰	۰/۹۸۹۷	۰/۹۸۷۲
دوم	جنگل تصادفی	۰/۹۹۱۴	۰/۹۸۹۶	۰/۹۹۴۳	۰/۹۹۱۳
	ماشین بردار پشتیبان	۰/۹۸۲۵	۰/۹۷۵۸	۰/۹۷۱۳	۰/۹۶۹۶
	رگرسیون منطقی	۰/۹۸۹۸	۰/۹۸۶۷	۰/۹۹۳۱	۰/۹۸۸۸
سوم	جنگل تصادفی	۰/۹۹۱۸	۰/۹۹۰۲	۰/۹۹۴۵	۰/۹۹۱۷
	ماشین بردار پشتیبان	۰/۹۸۰۵	۰/۹۶۸۹	۰/۹۶۶۳	۰/۹۶۴۰
	رگرسیون منطقی	۰/۹۹۱۹	۰/۹۸۹۷	۰/۹۹۴۵	۰/۹۹۱۳



شکل (۲): نمودار کارایی مدل جنگل تصادفی بر روی سناریوها



شکل (۳): نمودار کارایی مدل ماشین بردار پشتیبان بر روی سناریوها



شکل (۴): نمودار کارایی مدل رگرسیون منطقی بر روی سناریوها

نمودارهای ترسیم‌شده در شکل‌های (۲)، (۳) و (۴) به ترتیب، معیار F بدست‌آمده از مدل‌های جنگل تصادفی، ماشین بردار پشتیبان و رگرسیون منطقی بر روی تمامی سناریوها را نمایش می‌دهند.

۳-۴- تحلیل و انتخاب مجموعه ویژگی‌های مؤثر نهایی

در پژوهش حاضر، از مدل‌های جنگل تصادفی، ماشین بردار پشتیبان و رگرسیون منطقی جهت دسته‌بندی خودکار مقدار پارامتر ابعاد کاشی تولیدی، استفاده شده است. نتایج مقایسه معیارهای ارزیابی در هر سه مدل، دقت مطلوب در پیش‌بینی انحرافات ابعادی را نمایش می‌دهد.

به‌منظور بررسی پارامترهای مؤثر اولیه در تخمین خروجی مدل‌های پیش‌بینی، طی سه سناریو، سه دسته از پارامترها بعنوان ورودی به مدل‌ها تغذیه شدند. در سناریوی اول هرچند خروجی تمامی مدل‌های پیاده‌سازی شده از دقت مطلوبی برخوردار است ولی تعداد پارامترهای ورودی، نسبتاً زیاد است. هدف، رسیدن به دقت مطلوب با تعداد ورودی‌های کمتر است. متخصصان، دو احتمال را در رابطه با انحرافات کاشی‌های سرامیکی ارائه دادند. این احتمالات عبارتند از:

۱. انحرافات ابعادی کاشی‌های تولیدی، به پارامترهای اندازه‌گیری شده پس از مرحله پرس بستگی دارند.

۲. انحرافات ابعادی کاشی‌های تولیدی، به دماهای قسمت‌های مختلف واحد پخت و کوره بستگی دارند.

به‌منظور برآورده سازی احتمالات فوق، از ساختارهای پیشنهادی در سناریوهای دوم و سوم استفاده شده است. نتایج ارزیابی مدل‌های پیاده‌سازی شده بر روی سناریوهای انتخابی، در نمودارهای موجود در شکل (۵) نمایش داده شده است.

۴- نتیجه گیری

در سال‌های اخیر، استفاده از رویکردهای یادگیری ماشین در حوزه‌های مختلف بسیار مورد توجه قرار گرفته است. یکی از موارد استفاده از این روش‌ها، بکار بردن آنها در بازرسی کاشی‌های سرامیکی و کشف انحرافات ابعادی کاشی‌های تولیدی است. در این پژوهش، از مقایسه سه رویکرد جنگل تصادفی، ماشین بردار پشتیبان و رگرسیون منطقی جهت رسیدن به هدف فوق استفاده شده است.

در مسیر طراحی روش پیشنهادی، سه گام مختلف معرفی شده است. در گام اول، از استخراج ویژگی‌های اولیه برای ایجاد مدل پایه و در گام‌های دوم و سوم نیز، از ویژگی‌های اصلاح‌شده جهت ایجاد مدل‌های کارا تر استفاده شده است. همچنین، پارامتر ابعاد کاشی‌های تولیدی بعنوان خروجی کیفی انتخاب شده است.

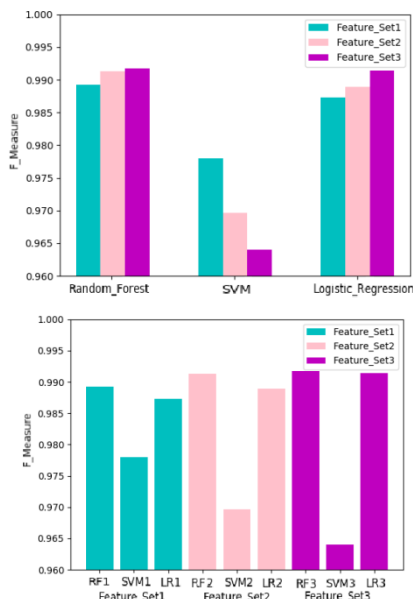
آزمایش‌های انجام‌شده نشان‌دهنده برتری عملکرد مدل جنگل تصادفی در سناریوی پیشنهادی گام سوم است. در این راستا، جهت کاهش زمان و افزایش دقت در کشف انحرافات ابعادی کاشی‌های تولیدی، استفاده از متغیرهای معرفی‌شده در سناریوی سوم، پیشنهاد می‌گردد. این متغیرها، حاوی فاکتورهای قابل اندازه‌گیری در فرآیند تولید کاشی و سرامیک، در مرحله پخت و کوره هستند.

در نهایت، جهت بررسی پارامترهای نهایی تأثیرگذار بر کیفیت، از روش انتخاب ویژگی پیش‌رو استفاده شده است. نتایج پیاده‌سازی روش انتخاب ویژگی پیش‌رو بر روی متغیرهای مربوط به سناریوی برتر، پارامترهای مربوط به دماهای بالاترین نقاط و پایین‌ترین نقاط واحد پخت و کوره را شامل می‌شوند. بدین جهت، در پژوهش حاضر از پیاده‌سازی مدل جنگل تصادفی بر روی پنج پارامتر اندازه‌گیری شده‌ی ابتدایی واحد پخت و کوره که کران‌های بالا و پایین کوره را شامل می‌شوند، جهت دسته‌بندی خودکار مقدار پارامتر ابعاد کاشی‌های تولیدی و کشف انحرافات ابعادی آنها استفاده می‌شود.

مراجع

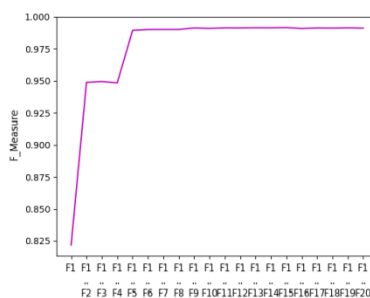
- [1] Zhao, X., "Evaluation and Improvement of the Service Quality of Tile Industry in China A Case Study of Company G", RSU International Research Conference, pp. 1864-1872, 2020.
- [2] Butt, U. A., Mehmood, M., Shah, S. B. H., Amin, R., Shaukat, M. W., Raza, S. M., Piran, M., "A Review of Machine Learning Algorithms for Cloud Computing Security", Electronics, Vol. 9, No. 9, pp. 1379, 2020.

- [۳] بکرانی، مهدی، حمیتی واقف، وحید، "ارائه روش بهبودیافته تصویب‌برداری فراصوتی در بازرسی جوش"، مجله مهندسی برق و الکترونیک ایران، دوره ۱۷، شماره ۱، ص. ۴۵-۵۹، ۱۳۹۹.
- [۴] ابوالقاسمی، وحید، فردوسی، سعیده، "یادگیری واژه نامه برای آشکارسازی محل پلاک خودرو"، مجله مهندسی برق و الکترونیک ایران، دوره ۱۴، شماره ۲، ص. ۱۲۵-۱۱۵، ۱۳۹۶.
- [۵] مودتی، سمیرا، مودتی، سینا، "دسته‌بندی و تعیین اصالت ارقام برنج به کمک الگوریتم فاکتورگیری ماتریس غیرمنفی تنک"، مجله



شکل (۵): نتایج ارزیابی مدل‌های پیاده‌سازی شده بر روی سناریوها

نتایج بدست‌آمده از سناریوها، برتری دقت و عملکرد در پیاده‌سازی ساختار موجود در سناریوی سوم را نشان می‌دهد. در این سناریو، تمامی دماهای واحد پخت و کوره در مدل‌سازی دخالت داده می‌شوند. تا این مرحله، همراه با انتخاب مدل مناسب برای پیش‌بینی، پارامترهای مؤثر اولیه در پیش‌بینی نتایج خروجی نیز به صورت کلی، معرفی گردیدند. این پارامترها، دماهای نقاط مختلف کوره را شامل می‌شوند. باین وجود، هدف نهایی یافتن پارامترهای نهایی دخیل در تخمین خروجی کیفی است. بدین منظور، از روش انتخاب ویژگی پیش‌رو^{۱۲} استفاده شده است. شکل (۶)، نتایج ارزیابی مدل جنگل تصادفی را بر روی پارامترهای موجود در سناریوی سوم، با استفاده از روش انتخاب ویژگی پیش‌رو نشان می‌دهد. در این شکل، کارایی مدل تا زمانی که به پارامتر پنجم برسد بالا رفته و سپس ثابت می‌شود. در واقع، پنج ویژگی اول اندازه‌گیری‌شده در واحد پخت و کوره بعنوان پارامترهای مؤثر نهایی شناخته می‌شوند.



شکل (۶): نمودار کارایی مدل جنگل تصادفی بر روی پارامترهای دماهای کوره

بر این اساس در پژوهش حاضر، از پیاده‌سازی مدل جنگل تصادفی که حاوی بیشترین مقدار برای معیار F است، در پنج پارامتر ابتدایی سناریوی سوم، استفاده می‌شود. این پنج پارامتر، دماهای بالاترین و پایین‌ترین نقاط کوره را شامل می‌شوند.

- [20] Fratello, M., Tagliaferri, R., "Encyclopedia of bioinformatics and computational biology: ABC of bioinformatics", Decision Trees and Random Forest, Vol. 1, pp. 374-384, 2018.
- [21] Chen, W., Zhang, S., Li, R., Shahabi, H., "Performance evaluation of the GIS-based data mining techniques of best-first decision tree, random forest, and Naïve Bayes tree for landslide susceptibility modeling", Science of the total environment, Vol. 644, pp. 1006-1018, 2018.
- [22] Sun, Y., Zhang, H., Zhao, T., Zou, Z., Shen, B., Yang, L., "A New Convolutional Neural Network With Random Forest Method for Hydrogen Sensor Fault Diagnosis", IEEE Access, Vol. 8, pp. 85421-85430, 2020.
- [23] Denisko, D., Hoffman, M. M., "Classification and interaction in random forests", Proceedings of the National Academy of Sciences, Vol. 115, No. 8, pp. 1690-1692, 2018.
- [24] Chagas, P., Souza, L., Araujo, I., Aldeman, N., Duarte, A., Angelo, M., Oliveira, L., "Classification of glomerular hypercellularity using convolutional features and support vector machine", Artificial Intelligence in Medicine, Vol. 103, 2020.
- [25] Nguyen, P. T., Ha, D. H., Avand, M., Jaafari, A., Nguyen, H. D., Al-Ansari, N., Ho, L. S., "Soft Computing Ensemble Models Based on Logistic Regression for Groundwater Potential Mapping", Applied Sciences, Vol. 10, No. 7, pp. 2469, 2020.
- [26] Abd Elgawad, M. A., Barakat, H. M., Yan, T., "Bivariate limit theorems for record values based on random sample sizes", Sankhya A: The Indian Journal of Statistics, Vol. 82, No. 1, pp. 50-67, 2020.
- [27] Hossin, M., Sulaiman, M. N., "A review on evaluation metrics for data classification evaluations", International Journal of Data Mining & Knowledge Management Process, Vol. 5, No. 2, 2015.
- [28] Khan, B., Naseem, R., Muhammad, F., Abbas, G., Kim, S., "An Empirical Evaluation of Machine Learning Techniques for Chronic Kidney Disease Prophecy", IEEE Access, Vol. 8, pp. 55012-55022, 2020.
- [29] Paradowski, M., Szytkowski, M., Kwasnicka, H., "A new F-score gradient-based training rule for the linear model", Pattern Analysis and Applications, Vol. 22, No. 2, pp. 537-548, 2019.
- [30] Luque, A., Carrasco, A., Martin, A., de las Heras, A., "The impact of class imbalance in classification performance metrics based on the binary confusion matrix", Pattern Recognition, Vol. 91, pp. 216-231, 2019.
- مهندسی برق و الکترونیک ایران، دوره ۱۷، شماره ۱، ص. ۲۱-۱۱، ۱۳۹۹
- [6] Feng, T., Xiao-Mei, X., Tso, S. K., Liu, K. P., "Application of evolutionary neural network in impact acoustics based nondestructive inspection of tile-wall", In Proceedings of Int. Conference on Communications, Circuits and Systems, pp. 974-978, 2005.
- [7] Cunha, R., Maciel, R., Nandi, G., Daros, M., Cardoso, J., Francis, L., Ramos, V., Marcelino, R., Frohlich, A., De Araujo, G. M., "Applying Non-destructive Testing and Machine Learning to Ceramic Tile Quality Control", In 2018 VIII Brazilian Symposium on Computing Systems Engineering (SBESC), pp. 54-61, 2018.
- [8] Elbehieri, H., Hefnawy, A., Elewa, M., "Surface defects detection for ceramic tiles using image processing and morphological techniques", World Academy of Science, Engineering and Technology, pp. 158-162, 2005.
- [9] Sharma, M., Kaur, G., "Integrated approach for defect detection in ceramic tiles", Int. journal of computers & technology, Vol. 3, No. 2b, pp. 259-262, 2012.
- [10] Patil, A. G., Phakade, S. V., Shirsath, D. O., "Automatic Defect Detection and Classification of Ceramic Tiles", International Journal of Advance Research in Science and Engineering, Vol. 5, No. 1, pp. 127-133, 2016.
- [11] Ramadhan, M. H., Rachmat, H., Atmaja, D. S. E., Ibrahim, R., "Multi-class Classification of Ceramic Tile Surface Quality using Artificial Neural Network and Principal Component Analysis", International Conference on Industrial Enterprise and System Engineering (ICoIESE 2018). Vol. 2, pp. 334-338, 2019.
- [12] Sioma, A., "Automated Control of Surface Defects on Ceramic Tiles Using 3D Image Analysis", Materials, Vol. 13, No. 5, 1250, 2020.
- [13] Karhe, R. R., Nagare, M. N. N., "Automatic Defect Detection and Classification Technique from Image Processing", International Journal on Recent and Innovation Trends in Computing and Communication, Vol. 6, No. 1, pp. 136-141, 2018.
- [14] Senthilkumar, M., Palanisamy, V., Jaya, J., "An Efficient Feature Fusion Technique for Surface Grading of Ceramic Tiles", Middle-East Journal of Scientific Research, Vol. 23, No. 1, pp. 59-65, 2015.
- [15] Raja, V. L., "Surface Grading Using Deep Learning Approach for Ceramic Tiles", International Journal of Advances in Signal and Image Sciences, Vol. 4, No. 1, pp. 1-7, 2018.
- [16] Cole, M. B., Risso, D., Wagner, A., DeTomaso, D., Ngai, J., Purdom, E., Dudoit, S., Yosef, N., "Performance assessment and selection of normalization procedures for single-cell RNA-seq", Cell systems, Vol. 8, No. 4, pp. 315-328, 2019.
- [17] Krause, R. W., Huisman, M., Steglich, C., Sniiders, T. A., "Missing network data a comparison of different imputation methods", In 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), pp. 159-163, 2018.
- [18] Newman, M.E.J., "Network structure from rich but noisy data", Nature Physics, Vol. 14, No. 6, pp. 542-545, 2018.
- [19] Onar, S. C., Oztaysi, B., Kahraman, C., Ozturk, E., "A Decision Support System to Optimize Debt Collection Assignments", In International Conference on Intelligent and Fuzzy Systems, pp. 178-187, 2019.

زیر نویس ها

- ¹ Neural Network Artificial
- ² Support Vector Machine (SVM)
- ³ K-Nearest Neighbor (KNN)
- ⁴ Laser Triangulation Method (LTM)
- ⁵ Morphological Operations
- ⁶ Discrete Wavelet Transform (DWT)
- ⁷ Gray-Level Co-Occurrence Matrix (GLCM)
- ⁸ Correlation Method
- ⁹ Random Forest
- ¹⁰ Logistic Regression
- ¹¹ Posterior Probability
- ¹² Forward Feature Selection

