Journal of Urban Environmental Management

Journal homepage: https://jumee.kgut.ac.ir
Vol.4, No.4; 2027.pp: 93-121

Research paper

(Received 4 Aug. 2026 Accepted 30 Oct. 2026)

Development of Deep Learning Models for Waste Classification and Object Detection in Smart Waste Management Through the Creation of Combined and Multi-Object Datasets

Behnam Aminizadeh [1], Zeynab Jahangiri [1], Bahram Malekmohammadi[*2]

[1] *M.Sc. Student in Environmental Systems Engineering, Graduate Faculty of Environment, University of Tehran*
[2] *Professor, Graduate Faculty of Environment, University of Tehran.*

Abstract

Classification and object detection models were developed for use in automatic waste-sorting systems, support for waste management in construction projects, environmental protection, supporting the health of living organisms, and error and cost reduction. In this article, two independent and traceable datasets, after duplicate and similarity screening, were modeled using EfficientNet-B0 with transfer learning. The models classified waste images into 11 and 8 commonly used classes with accuracies of 97.03% and 91.31%, respectively. Comparison with a custom CNN trained from scratch showed that the EfficientNet-B0-based model achieved faster convergence in the early training epochs, higher validation accuracy, and a lower training time per epoch. After out-of-sample evaluation and the demonstrated importance of image diversity, the two datasets were merged to create the seven-class MWD7 dataset, and the new model achieved an accuracy of 97.02%. Comparison between EfficientNet-B0 and EfficientNet-B2 showed that EfficientNet-B2 improved the evaluation metrics when trained with its standard input size rather than a fixed input size. This result indicates that heavier architectures in this family, when used with their standard input size, can improve model quality. After the limitations of EfficientNet in analyzing images containing multiple waste types became evident, the multi-object MOWD dataset was created, and YOLO26s was trained to simultaneously detect the type and location of each waste object in multi-object images, achieving an mAP50-95 of 0.792. All models were evaluated using unseen images and multiple metrics. The results showed that the proposed methodology improves the accuracy and efficiency of intelligent waste classification and object detection.

Keywords: Waste management, Machine Learning, Deep Learning, Waste Classification, Object Detection, Transfer Learning, Environment, Intelligent Systems.

*Corresponding Author: Bahram Malekmohammadi
DOI: 10.48306/juem.2026.595365.1174
Email: *malekb@ut.ac.ir*
Phone: 02161113185

نشریه علمی مدیریت محیط زیست شهری

آدرس نشریه:https://jumee.kgut.ac.ir

سال چهارم/شماره چهارم/زمستان ۱۴۰۵/صفحه ۹۳-۱۲۱

مقاله پژوهشی

تاریخ دریافت:۱۴۰۵/۵/۱۳ تاریخ پذیرش: ۱۴۰۵/۸/۸

توسعه‌ی مدل‌های یادگیری عمیق برای طبقه‌بندی و تشخیص اشیاء در مدیریت هوشمند پسماند، همراه با ایجاد مجموعه‌داده‌های ترکیبی و چندشیئی

بهنام امینی زاده[۱]، زینب جهانگیری[۱]، بهرام ملک‌محمدی [۲*]

[۱] دانشجوی کارشناسی ارشد مهندسی سیستم‌های محیط‌زیست، دانشکده محیط‌زیست، دانشگاه تهران، تهران، ایران.

[۲] استاد دانشکده محیط‌زیست، دانشگاه تهران، تهران، ایران.

چکیده:

مدل‌های طبقه‌بندی و تشخیص شیء برای کاربرد در سامانه‌های تفکیک خودکار پسماند، پشتیبانی از مدیریت پسماند در پروژه‌های عمرانی، حفاظت از محیط‌زیست، کمک به سلامت جانداران و کاهش خطا و هزینه توسعه یافتند. در این مقاله دو مجموعه‌داده‌ی مستقل و قابل ردیابی، پس از غربالگری تکرار و شباهت، با معماری EfficientNet-B0 و انتقال یادگیری مدل‌سازی شدند. مدل‌ها تصاویر پسماند را به‌ترتیب در ۱۱ و ۸ دسته‌ی پرکاربرد با دقت ۹۷/۰۳٪ و ۹۱/۳۱٪ طبقه‌بندی کردند. مقایسه با CNN سفارشی آموزش‌دیده از صفر نشان داد مدل مبتنی بر EfficientNet-B0 با انتقال یادگیری در دورهای ابتدایی آموزش، همگرایی سریع‌تر، دقت اعتبارسنجی بالاتر و میانگین زمان هر دور آموزش کمتری داشت. پس از ارزیابی برون‌مجموعه‌ای و مشخص شدن اهمیت تنوع تصاویر، با ترکیب دو مجموعه‌داده، مجموعه‌داده‌ی MWD7 در ۷ دسته‌ی پرکاربرد پسماند ایجاد شد و مدل جدید با دقت ۹۷/۰۲٪ آموزش دید. مقایسه‌ی EfficientNet-B0 و EfficientNet-B2، نشان داد EfficientNet-B2 برخلاف اندازه‌ی ورودی ثابت، با اندازه‌ی ورودی استاندارد موجب بهبود معیارهای ارزیابی شد. این نتیجه نشان داد استفاده از معماری‌های سنگین‌تر این خانواده همراه با اندازه‌ی ورودی استاندارد، می‌تواند کیفیت مدل را بهبود دهد. با آشکار شدن محدودیت EfficientNet در تحلیل تصاویر دارای چند نوع پسماند، مجموعه‌داده‌ی چندشیئی MOWD ایجاد شد و مدل YOLO26s برای تشخیص هم‌زمان نوع و موقعیت هر پسماند در تصاویر چندشیئی آموزش یافت و مقدار mAP50-95 برابر ۰/۷۹۲ به دست آمد. تمامی مدل‌ها با تصاویر دیده نشده و چندین معیار معتبر ارزیابی شدند. نتایج نشان داد روش تحقیق، دقت و کارآیی مدل‌ها را برای طبقه‌بندی و تشخیص هوشمند پسماند بهبود می‌دهد.

واژگان کلیدی: مدیریت پسماند، یادگیری ماشین، یادگیری عمیق، طبقه‌بندی پسماند، تشخیص اشیاء، انتقال یادگیری، محیط زیست، هوشمندسازی.

* نویسنده مسئول: بهرام ملک محمدی

ایمیل: malekb@ut.ac.ir

شماره تماس: ۰۲۱۶۱۱۱۳۱۸۵

DOI: 10.48306/juem.2026.595365.1174

۱-مقدمه

پسماندهای با مدیریت نامناسب، فقط یک مشکل ظاهری زندگی شهری نیست؛ بلکه می‌تواند به آلودگی هوا و آب، تخریب زمین، تشدید تغییرات اقلیمی و ایجاد خطرات جدی برای سلامت عمومی منجر شود. یک مرور منتشر شده در سال ۲۰۲۴ نشان می‌دهد که دفع نادرست پسماند شهری، به ویژه دفن و رها سازی غیرقانونی و اشتباه، با تشدید بیماری‌های عفونی و آسیب به پایداری زندگی شهری مرتبط است. بنابراین، مدیریت پسماند شهری عملاً بخشی از مدیریت سلامت انسان و محیط زیست محسوب می‌شود [۱]. مرور جدید دیگری نشان می‌دهد که راهبردهای اقتصاد چرخشی در مدیریت پسماند شهری می‌توانند به کاهش انتشار کربن و تقویت توسعه‌ی پایدار شهری کمک کنند؛ مدیریت پسماند فقط جمع‌آوری زباله نیست بلکه ابزاری برای کاهش اثرات اقلیمی و تقویت تاب‌آوری شهر است [۲]. همچنین مدیریت پسماندهای ساخت و ساز و تخریب، به دلیل سهم بالا در مصرف منابع و انتشار کربن، یکی از محورهای اصلی پایداری در مهندسی عمران است. در این میان تفکیک هوشمند پسماند یک لایه‌ی توانمندساز برای تولید خوراک خالص و یکنواخت جهت مصالح ثانویه محسوب می شود [۳ و ۴]. مطالعات اخیر نشان می‌دهند که هوش مصنوعی، بینایی ماشین و رباتیک در مدیریت پسماند از تشخیص و تفکیک تا بازیابی منابع و بهینه سازی فرآیند، نقش عملیاتی دارند؛ بنابراین کاربرد مدل یادگیری عمیق در جداسازی پسماند صرفاً یک مسئله‌ی پردازش تصویر نیست بلکه یک فناوری توانمندساز برای بازیافت مصالح نیز محسوب می‌گردد [۵ و ۶].

شیشه‌ی بازیافتی تفکیک شده می‌تواند پس از خردایش و کنترل دانه‌بندی به صورت پودر/سنگدانه در بتن استفاده شود [۷ و ۸]. پلاستیک‌های تفکیک شده و تمیز می‌توانند به صورت جایگزین جزئی سنگدانه در بتن به کار روند؛ با این حال اثر آن‌ها بر کارپذیری، مقاومت و دوام باید در در صدهای بهینه ارزیابی شود [۹ و ۱۰]. الیاف حاصل از ضایعات نساجی و پوشاک می‌توانند به عنوان الیاف تقویتی پایدار در بتن استفاده شوند و بخشی از نیاز به الیاف مصنوعی را کاهش دهند. این رویکرد به کنترل ترک خوردگی و افزایش جذب انرژی کمک می‌کند [۱۱ و ۱۲]. کاغذ و کارتن تفکیک شده پس از خردایش و در درصدهای بهینه، می‌تواند به عنوان افزودنی سبک یا بخشی از جایگزینی در ملات و بتن استفاده شود و در برخی مطالعات به بهبود مقاومت‌های فشاری، خمشی و کششی منجر شده است [۱۳ و ۱۴]. مواد آلی و زیستی تفکیک شده پس از تبدیل به زغال زیستی[1]می‌توانند در مواد سیمانی و آسفالتی به عنوان پرکننده یا افزودنی مصرف شوند و هم‌زمان به کاهش انتشار CO_2 و بهبود برخی خواص حرارتی/میکرو ساختاری کمک کنند [۱۵ و ۱۶]. فلزات تفکیک شده به ویژه فولاد می‌توانند دوباره در چرخه‌ی ساخت و ساز و سازه‌های جدید وارد شوند و از این طریق انتشار کربن ناشی از استخراج و فولادسازی کاهش یابد [۱۷]. این سامانه را می‌توان به عنوان ابزار ممیزی پیش از تخریب[2]و برنامه‌ریزی چرخه‌پذیری مصالح[3]در پروژه‌های ساختمانی و تخریبی معرفی کرد؛ به این معنا که پسماند تولید شده توسط پروژه‌های عمرانی پس از تفکیک هوشمند، دوباره به زنجیره‌ی مصالح عمرانی بازمی‌گردد [۴]. تفکیک هوشمند پسماند با افزایش خلوص جریان‌های بازیافتی، وابستگی به دفن را کاهش می‌دهد و با کم کردن آلودگی خاک و آب و کاهش تقاضا برای تولید مواد اولیه‌ی بکر، به کاهش انتشار گازهای گلخانه‌ای کمک می‌کند [۳ و ۴].

پسماندهای تولید شده از زندگی شهری و پروژه‌های ساختمانی، تخریبی و صنعتی پس از جمع‌آوری به مرکز پردازش یا بازیابی مواد منتقل می‌شوند و در آنجا برای بازیابی منابع و کاهش دفن، آماده‌سازی اولیه می‌گردند [۳ و ۴]. سامانه‌ی حسگری، داده‌ی تصویری یا طیفی را در حین حرکت روی نوار نقاله یا بستر انتقال، از ماده دریافت می‌کند. داده‌ی دریافت شده به مدل یادگیری عمیق ارسال می‌شود و مدل، نوع پسماند را تعیین می‌کند. در این مرحله، پردازش تصویر به عنوان یک تصمیم عملیاتی برای جداسازی ماده عمل می‌کند. پس از استنتاج مدل، فرمان مربوطه به ربات ارسال می شود و پسماند مدنظر به مخزن متناظر، مسیر جداگانه یا بازوی ربات هدایت می‌گردد؛ بسته به طراحی خط، این مرحله می‌تواند به طرق متعددی انجام شود. در این فرآیند، مدل پردازش تصویر یک جزو کلیدی است که به طور مستقیم کیفیت و کارآیی را تعیین می‌کند. مدل ضعیف به معنای تفکیک و طبقه‌بندی ضعیف و مدل خوب به معنای تفکیک و طبقه‌بندی خوب است.

---

[1] Biochar
[2] Pre-demolition auditing
[3] Circularity Planning

یکی از راهبردها جهت آموزش مدل طبقه‌بندی تصاویر، استفاده از معماری EfficientNet و انتقال یادگیری[1] است. در سال ۲۰۱۹ Tan و Le از گروه تحقیقاتی وابسته به شرکت Google به اسم Brain Team، خانواده‌ای از معماری‌های شبکه‌های عصبی کانولوشنی را با نام EfficientNet معرفی کردند. ایده‌ی اصلی آن‌ها مقیاس‌سازی مرکب بود؛ به این معنا که به جای بزرگ کردن شبکه فقط از یک بعد؛ عمق، عرض و وضوح ورودی به صورت هماهنگ و متعادل افزایش داده می‌شوند. در این خانواده، EfficientNet-B0 به عنوان مدل پایه طراحی شد و سپس نسخه‌های بزرگ‌تر آن از EfficientNet-B1 تا EfficientNet-B7 با همان منطق مقیاس‌سازی توسعه یافتند. نتایج نشان می‌دهد EfficientNet-B0 با ۵/۳ میلیون پارامتر به دقت معادل ۷۷/۱ درصد روی مجموعه‌داده‌ی ImageNet رسیده است و EfficientNet-B7 با ۶۶ میلیون پارامتر به دقت معادل ۸۴/۳ در صد دست یافته است. این خانواده نسبت به بسیاری از معماری‌های قبلی، توازن بهتری میان سرعت، دقت و کارآیی محاسباتی برقرار می‌کند [۱۸]. مجموعه‌داده‌ی ImageNet در سال ۲۰۰۹ توسط Deng و همکاران معرفی شد [۱۹]. در گزارش اولیه، این مجموعه‌داده شامل ۱۲ زیرشاخه و ۳/۲ میلیون تصویر بود. در تحقیقات یادگیری عمیق، عمدتاً از زیرمجموعه‌ی ImageNet-1K / ILSVRC استفاده می‌شود که ۱۰۰۰ کلاس و بخش آموزش آن، ۱،۲۸۱،۱۶۷ تصویر دارد و به عنوان یکی از استانداردترین مجموعه‌داده‌ها در بینایی ماشین شناخته می‌شود. خانواده‌ی مدل‌های EfficientNet روی همین زیرشاخه از ImageNet با بیش از یک میلیون و دویست هزار تصویر آموزش دیده اند [۱۸ و ۲۰]. انتقال یادگیری یک رویکرد در یادگیری ماشین است که در آن دانش آموخته شده از یک دامنه یا مسئله‌ی مبدأ، به دامنه یا مسئله‌ی هدف منتقل می‌شود. این رویکرد به ویژه زمانی مفید است که داده‌های برچسب‌خورده‌ی دامنه‌ی هدف محدود باشند یا توزیع داده‌ها با داده‌های مبدأ تفاوت داشته باشد. در بینایی ماشین، این فرآیند عمدتاً با استفاده از یک مدل از پیش آموزش دیده همانند مدل‌های هشت‌گانه‌ی EfficientNet آغاز می‌شود. در بسیاری از پیاده‌سازی‌ها، لایه‌های ابتدایی به عنوان استخراج‌گر ویژگی، فریز می‌شوند و بدون تغییر باقی می‌مانند و لایه‌های نهایی، متناسب با مسئله‌ی هدف بازآموزی[2] می‌شوند. به همین دلیل به جای آموزش کامل مدل از صفر، می‌توان از دانشی که مدل روی داده‌های بزرگ و عمومی یاد گرفته است استفاده کرد و آن را برای مسئله‌ی خاص، مانند طبقه‌بندی پسماند در تعداد دسته‌های دلخواه، سازگار ساخت [۲۱ و ۲۲].

از طرفی تشخیص اشیاء در پسماند، یک گام پیشرفته‌تر از صرفاً طبقه‌بندی تصاویر حاوی یک نوع پسماند است و یکی از کاربردهای آن در تحقیق حاضر، زمانی اهمیت پیدا می‌کند که در یک تصویر، چندین شیء یا چند نوع پسماند هم‌زمان وجود داشته باشند. در این حالت، شیوه‌ی صحیح هوشمندسازی، نه تنها نوع شیء را تشخیص می‌دهد بلکه محل دقیق آن را نیز مشخص می‌کند؛ بنابراین برای کاربردهای عملی می‌تواند مفید واقع شود. اهمیت این رویکرد در این است که سامانه‌ی خودکار در تصاویر حاوی چند پسماند، فقط تشخیص نمی‌دهد که چه چیزی در تصویر است بلکه تشخیص می‌دهد دقیقاً کجاست و می‌تواند بر همان اساس عمل کند. این قابلیت در محیط‌های واقعی که امکانات و تجهیزات کافی جهت قرارگیری خطی پسماند روی نوار نقاله فراهم نیست و پسماندها درهم‌ریخته هستند، نقش تعیین کننده‌ای در افزایش سرعت پردازش، کاهش خطا در جداسازی و امکان کنترل دقیق‌تر فرآیند دارد. الگوریتم YOLO یکی از مهم‌ترین رویکردها در تشخیص اشیاء است. نام این الگوریتم مخفف عبارت You Only Look Once به معنای تنها یک بار نگاه می‌کنی است. این نام بیانگر مهم‌ترین ایده‌ی این خانواده از مدل‌ها است؛ به این معنا که تصویر تنها یک بار وارد شبکه‌ی عصبی می‌شود و مدل به صورت هم‌زمان مکان اشیاء (جعبه محدود کننده)، کلاس اشیاء و میزان اطمینان را پیش‌بینی می‌کند. برخلاف بسیاری از روش‌های قدیمی تشخیص شیء که فرآیند تشخیص را در چند مرحله انجام می‌دادند، YOLO تمامی مراحل را در قالب یک شبکه عصبی واحد انجام می‌دهد. مدل جدید YOLO26s از خانواده‌ی معماری YOLO از نظر سرعت اجرا و تعداد پارامترها از نسخه‌های بزرگ‌تر، سبک‌تر می‌باشد و عین حال دقت آن برای بسیاری از کاربردهای عملی همچنان مناسب است. همین توازن میان کارآیی و سرعت در قیاس با بسیاری از مدل‌های خانواده‌ی YOLO، مدل YOLO26s را برای خیلی از پژوهش‌ها مناسب می‌گرداند [۲۳ و ۲۴].

در تشریح اهداف پژوهش حاضر می‌توان گفت مطابق با موارد عنوان شده، تفکیک و پایش پسماند می‌تواند در مهندسی عمران و ساخت و ساز، کاربردهای متعددی داشته باشد. همچنین تفکیک پسماند به عنوان یکی از اصول مهم در مدیریت پسماند می‌تواند تأثیرات مثبت فراوان محیط زیستی به دنبال داشته باشد. تفکیک و پایش پسماند در سکوهای جمع‌آوری، خیابان‌ها و مراکز پردازش،

[1] Transfer Learning
[2] fine tune

یک شاخص عینی برای پاکیزگی شهری ایجاد می‌کند و با کاهش رها شدگی زباله، سرریز مخازن، بوی نامطبوع، جذب جوندگان و انسداد مسیرهای آب‌های سطحی، کیفیت زیست شهری را بهبود می‌دهد و آرامش بیشتر و سلامت مردم را به دنبال خواهد داشت. هوشمند سازی جمع‌آوری و تفکیک پسماند با حذف و یا کاهش حضور نیروی انسانی و حذف بازدیدهای غیر ضروری، موجب کاهش هزینه‌ی عملیاتی، کاهش مصرف سوخت و بهبود کیفیت خدمت‌رسانی می‌شود. تفکیک خودکار، تماس مستقیم نیروی انسانی با پسماندهای آلوده، تیز، خورنده یا خطرناک را کاهش می‌دهد و از این طریق ریسک‌های بهداشتی، شغلی و محیط زیستی را نیز پایین می‌آورد. چنین سامانه‌ای می‌تواند تصمیم‌گیری را سریع‌تر، پایدارتر و مقیاس‌پذیرتر کند و در کاهش هزینه‌ها نقش موثری ایفا کند. همچنین زمانی که تفکیک پسماند توسط انسان انجام شود، مباحثی مثل کاهش دقت به واسطه‌ی خستگی، بی‌دقتی و یا عدم تمرکز اپراتور و همچنین محدودیت زمانی به علت ساعت کاری محدود مطرح می‌گردد که داشتن سامانه‌ای هوشمند، این مشکلات را به شیوه‌ای مناسب مدیریت خواهد کرد. سامانه‌ی هوشمند همچنین می‌تواند پسماندهای خطرناک همچون برخی باتری‌ها را از جریان پسماند جدا کند و منجر به کاهش ریسک آلودگی شود. در این پژوهش یکی از اهداف کلیدی، ساخت مدل پردازش تصویر برای طبقه‌بندی پسماند است. همچنین مدل‌های تشخیص اشیاء می‌توانند در شناسایی انواع خاصی از پسماند در سطح شهر با کمک تصویربرداری هوایی و یا تصاویر دوربین‌های سطح شهر کاربرد داشته باشند؛ هدف کلیدی دیگر این پژوهش ساخت و تحلیل مدل پردازش تصویر برای تشخیص اشیاء در پسماند است.

با بررسی مقالات علمی پژوهشی داخلی پیشین می‌توان دریافت که پژوهش‌های مبتنی بر یادگیری عمیق و تصاویر پسماند هنوز بسیار محدود هستند و این حوزه بیشتر در سطح بین‌المللی بررسی شده است. در پژوهش حاضر از دو مجموعه‌داده‌ی مستقل استفاده شد و سپس با مدل‌سازی روی مجموعه‌داده‌ها و بررسی و تحلیل نتایج و تشخیص نیاز به توسعه‌ی مجموعه‌داده‌های جدید، یک مجموعه‌داده‌ی ترکیبی، با تنوع و تعداد تصاویر بیشتر و یک مجموعه‌داده‌ی چند شیئی با حضور همزمان چندین نوع پسماند در یک تصویر توسعه داده شدند تا بتوان مدل‌های جدیدی را با کمک آن‌ها آموزش داد. در مجموع، تحقیق حاضر با آموزش هفت مدل یادگیری عمیق بر روی چهار مجموعه‌داده، در فاز طبقه‌بندی با مدل‌های خانواده‌ی EfficientNet و انتقال یادگیری و مقایسه با CNN سفارشی آموزش دیده از صفر و در فاز تشخیص نوع و موقعیت شیء با معماری YOLO26s و انتقال یادگیری، انجام آزمایشات متعدد جهت بررسی و ارتقای عملکرد مدل‌ها، ارزیابی برون‌مجموعه‌ای همراه با غربالگری تکرار و شباهت برای اطمینان بیشتر به نتایج ارزیابی، حصول نتایج کاربردی از بررسی و مقایسه‌ی مدل‌های یادگیری عمیق و ارزیابی مدل‌ها و توسعه‌ی مدل‌های کاربردی با معیارهای ارزیابی با مقادیر خوب و با کیفیت، توانسته است در راستای غنای ادبیات موجود در حوزه‌ی مدیریت هوشمند پسماند از طریق پردازش تصویر، گام موثری بردارد و برخی از خلاءهای کلیدی تحقیقات پیشین را پوشش دهد. در تحقیق و در انتهای بخش نتایج، مقایسه با برخی مطالعات پیشین داخلی و بین‌المللی انجام شده است.

۲- معرفی مجموعه‌داده‌های تحقیق

۲-۱- Garbage Classification

مجموعه‌داده‌ی Garbage Classification در Kaggle منتشر شده است، این مجموعه‌داده همچنین در پژوهشی که در سال ۲۰۲۵ انجام گرفته، معرفی شده است [۲۵ و ۲۶].

جدول ۱- معرفی دسته‌های مجموعه‌داده‌ی Garbage Classification [۲۵ و ۲۶]

ردیف	نام اصلی دسته	معادل فارسی نام دسته	تعداد تصویر
۱	battery	باتری	۹۴۵
۲	biological	زیستی	۹۸۵
۳	brown glass	شیشه قهوه‌ای	۶۰۷
۴	cardboard	کارتن	۸۹۱
۵	clothes	پوشاک	۵۳۲۵

۶	green glass	شیشه سبز	۶۲۹
۷	metal	فلز	۷۶۹
۸	paper	کاغذ	۱۰۵۰
۹	plastic	پلاستیک	۸۶۵
۱۰	shoes	کفش	۱۹۷۷
۱۱	white glass	شیشه سفید	۷۷۵
۱۲	Trash	زباله	۶۹۷
	مجموع	۱۵۵۱۵	

مطابق با جدول ۱ این مجموعه‌داده ۱۵۵۱۵ تصویر پسماند را در ۱۲ دسته دربرمی‌گیرد. دسته‌ی trash (زباله)، نیز دربرگیرنده‌ی تصاویر متفرقه از زباله‌ها است.

۲-۲- Realwaste

مجموعه‌داده‌ی Realwaste یک مجموعه‌داده‌ی تصویری از پسماندهای دفن شده در محل‌های بازیافت و انبار زباله است که در اواخر سال ۲۰۲۳ منتشر شده. این مجموعه‌داده، تصاویر واقعی پسماندهای ورودی به محل دفن را شامل می‌شود و در مقاله‌ای که توسط ارائه دهنده‌های آن منتشر شده، معرفی گردیده است [۲۷].

جدول ۲- معرفی دسته‌های مجموعه‌داده‌ی Realwaste [۲۷]

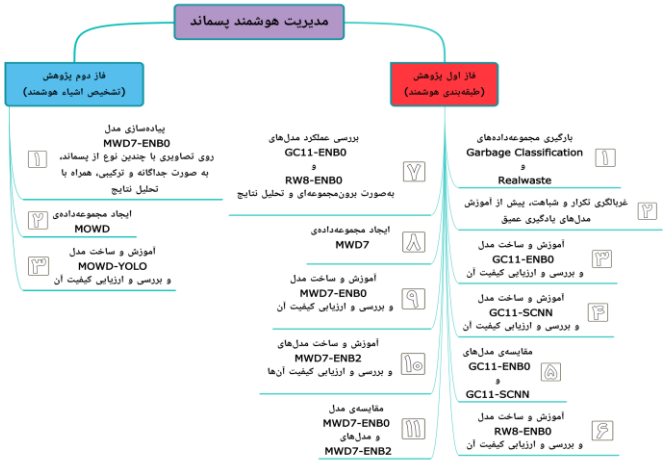
ردیف	نام اصلی دسته	معادل فارسی نام دسته	تعداد تصویر
۱	Cardboard	کارتن	۴۶۱
۲	Food Organics	پسماند آلی غذایی	۴۱۱
۳	Glass	شیشه	۴۲۰
۴	Metal	فلز	۷۹۰
۵	Paper	کاغذ	۵۰۰
۶	Plastic	پلاستیک	۹۲۱
۷	Textile Trash	پسماند نساجی	۳۱۸
۸	Vegetation	پسماند گیاهی	۴۳۶
۹	Miscellaneous Trash	پسماند متفرقه	۴۹۵
	مجموع	۴۷۵۲	

مطابق با جدول ۲ این مجموعه‌داده، ۴۷۵۲ تصویر پسماند را در ۹ دسته دربرمی‌گیرد. دسته‌ی Miscellaneous Trash (پسماندهای متفرقه)، نیز دربرگیرنده‌ی تصاویر متفرقه از زباله‌ها است.

۳- روش‌شناسی پژوهش

۳-۱- مراحل پژوهش

پژوهش حاضر در دو فاز انجام شده است. فاز اول از ۱۱ مرحله‌ی اصلی تشکیل شده؛ این مراحل از بارگیری داده‌های مورد نیاز پژوهش تا ساخت شش مدل یادگیری عمیق براساس دو مجموعه‌داده‌ی مستقل و یک مجموعه‌داده‌ی ترکیبی را دربرمی‌گیرند. همچنین مقایسه‌ی بین مدل‌ها و علت پیشروی در ساخت مدل‌های جدید، بررسی می‌شوند و در هر مرحله، مدل‌ها از نظر کارآیی و کیفیت ارزیابی می‌شوند. در فاز دوم تحقیق، ۳ مرحله‌ی اصلی طی شدند تا مدل آخر و هفتم پژوهش نیز طراحی و ارزیابی شود. شکل ۱، مراحل انجام پژوهش را نمایش می‌دهد.



شکل ۱- مراحل انجام تحقیق

۳-۲- غربالگری تکرار و شباهت

پیش از آغاز آموزش مدل‌ها، برای جلوگیری از نشت داده[1] دو مجموعه‌داده‌ی معرفی شده در بخش ۲-۱ و ۲-۲ به صورت فقط خواندنی بررسی شدند؛ یعنی هیچ فایل تصویری تغییر نام داده نشد و همچنین جابه‌جایی و حذف صورت نگرفت. هدف از این مرحله، کسب اطمینان مناسب از استقلال دو مخزن تصویری و شناسایی هرگونه هم‌پوشانی یا شباهت زیاد میان تصاویر آن‌ها بود. این مرحله برای افزایش اعتبار ارزیابی نهایی ضروری است چراکه در این تحقیق، ارزیابی برون‌مجموعه‌ای یک مدل روی مجموعه‌داده‌ی دیگر صورت گرفته است [۲۸]. در گام نخست برای هر فایل تصویر یک اثر انگشت دیجیتال[2] بر پایه‌ی SHA-256 محاسبه شد. SHA-256 یک تابع درهم‌ساز رمزنگارانه است که برای هر فایل، یک خروجی ثابت تولید می‌کند؛ بنابراین اگر دو فایل اثر انگشت دیجیتال یکسان داشته باشند، از دید این غربالگری، تکرار دقیق تلقی می‌شوند. این روش برای تشخیص تکرار دقیق در سطح فایل[3] مناسب است، زیرا بر محتوای باینری فایل متکی است و به نام فایل وابسته نیست [۲۸]. از آن‌جا که دو تصویر ممکن است از نظر ظاهری بسیار شبیه باشند اما از نظر بایت به بایت یکسان نباشند، در گام دوم از هش ادراکی استفاده شد. هش ادراکی بر خلاف هش رمزنگارانه، هدفش تشخیص شباهت بصری است، نه فقط برابری دقیق فایل. در این پژوهش pHash مبنای اصلی غربالگری شبه‌تکرارها بود. pHash با تکیه بر تبدیل کسینوسی گسسته[4] ضرایب متناظر با فرکانس‌های پایین تصویر را استخراج کرده و بر اساس آن‌ها یک کد دودویی فشرده تولید می‌کند [۲۹]. برای عملی نگه‌داشتن محاسبات، pHash هر تصویر با یک سطل‌بندی بر اساس ۱۶ بیت ابتدایی گروه‌بندی شد و مقایسه‌ی جفت‌تصاویر فقط درون هر سطل انجام گرفت. سپس فاصله‌ی همینگ[5] بین pHashها محاسبه شد و جفت‌هایی که فاصله‌ی همینگ آن‌ها ۵ یا کمتر بود به عنوان نامزد شبه‌تکرار ثبت شدند. این آستانه به گونه‌ای انتخاب شد تا از حساسیت بیش از حد و هشدارهای کاذب جلوگیری کند [۲۹]. نتیجه‌ی اجرای این غربالگری نشان داد که از میان ۱۵۵۱۵ تصویر مجموعه‌داده‌ی نخست و ۴۷۵۲ تصویر مجموعه‌داده‌ی دوم، هیچ گروه تکرار دقیق مبتنی بر SHA-256 مشاهده نشد. همچنین هیچ برخورد دقیق pHash نیز ثبت نشد. در سطح شبه‌تکرارها، تنها چهار جفت، به صورت دو به دو با فاصله‌ی همینگ کمتر یا مساوی ۵ شناسایی شد. نکته‌ی مهم آن است که حداقل یکی از تصاویر هر چهار جفت، به دسته‌های trash و miscellaneous trash مربوط بودند و مطابق با الگویی که در ادامه شرح داده خواهد شد، در دسته‌های نهایی حفظ شده برای آموزش مدل‌ها وارد نمی‌شدند. بنابراین در دسته‌هایی که در عمل برای ساخت مدل استفاده شدند، نشانه‌ای از هم‌پوشانی معنادار میان دو مجموعه‌داده مشاهده نشد.

---

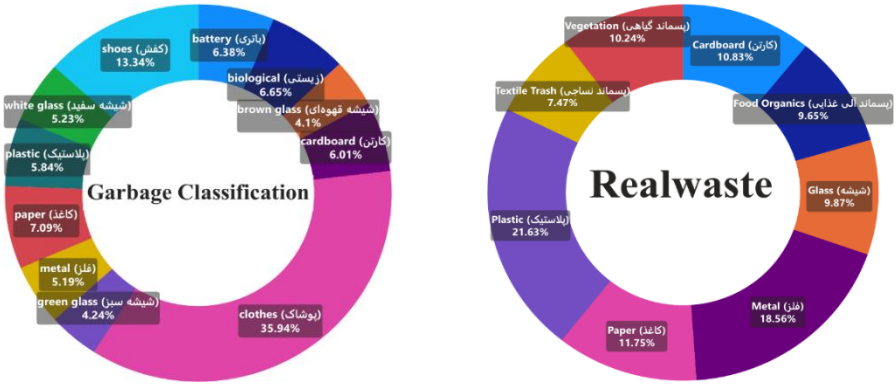
[1] Data Leakage
[2] Digest
[3] duplicate file-level
[4] Discrete Cosine Transform
[5] Hamming

۳-۳- ساختار مدل‌های پردازش تصویر

در این پژوهش، چند مدل مختلف برای طبقه‌بندی تصاویر پسماند توسعه داده شد. با وجود تفاوت در معماری پایه[1]یا مجموعه‌داده‌ی مورد استفاده در برخی آزمایش‌ها، تمامی مدل‌های طبقه‌بندی از یک چارچوب آموزشی مشترک پیروی کردند. استفاده از یک چارچوب آموزشی یکسان باعث می‌شود مقایسه میان مدل‌ها منصفانه بوده و تفاوت نتایج عمدتاً ناشی از تفاوت معماری یا داده‌های ورودی باشد.

داده‌ها پیش از آغاز آموزش مدل‌ها به سه زیرمجموعه‌ی آموزش، اعتبارسنجی و آزمون تقسیم شدند. مجموعه‌ی آموزش برای یادگیری پارامترهای مدل، مجموعه‌ی اعتبارسنجی برای پایش روند آموزش، تنظیم فرآیند یادگیری و مدیریت بیش‌برازش[2]و مجموعه‌ی آزمون برای ارزیابی نهایی عملکرد مدل استفاده شد. نسبت تقسیم داده‌ها به ترتیب ۸۰ درصد برای آموزش، ۱۰ درصد برای اعتبارسنجی و ۱۰ درصد برای آزمون در نظر گرفته شد. این بدان معنا است که به جهت سنجش کیفیت مدل، مدل ساخته شده، روی ۱۰ درصد از داده‌هایی که درحین آموزش هرگز آن‌ها را ندیده است تست نهایی می‌شود. به منظور حفظ توزیع کلاس‌ها در هر سه زیرمجموعه، از تقسیم‌بندی طبقه‌بندی شده[3]استفاده شد. در این روش، نسبت نمونه‌های هر کلاس تا حد امکان در مجموعه‌های آموزش، اعتبارسنجی و آزمون حفظ می‌شود. این موضوع به ویژه در مسائل طبقه‌بندی چندکلاسه اهمیت دارد، زیرا از ایجاد سوگیری ناشی از توزیع نامتوازن کلاس‌ها در زیرمجموعه‌های مختلف جلوگیری می‌کند و ارزیابی و مقایسه‌ی مدل‌ها را منصفانه‌تر می‌سازد [۳۰].

در مسائل بینایی ماشین، تعداد محدود تصاویر آموزشی می‌تواند باعث شود مدل به جای یادگیری الگوهای کلی، جزئیات خاص تصاویر آموزشی را حفظ کند. این پدیده که به بیش‌برازش معروف است، موجب کاهش توان تعمیم مدل روی تصاویر جدید می‌شود. به منظور افزایش تنوع داده‌های آموزشی و کاهش خطر بیش‌برازش، از افزایش مصنوعی داده‌ها[4]استفاده شد. در این روش، بدون تولید تصویر جدید توسط انسان و بدون تغییر برچسب کلاس، نسخه‌های متفاوتی از تصاویر موجود به صورت تصادفی در حین آموزش تولید می‌شوند. در این پژوهش چهار نوع تبدیل روی تصاویر آموزشی اعمال شد که عبارت اند از وارون‌سازی افقی، چرخش تصادفی، بزرگ‌نمایی و کوچک‌نمایی تصادفی و تغییر تصادفی کنتراست. هدف از این تبدیلات، شبیه‌سازی شرایط مختلف تصویربرداری و افزایش تنوع نمونه‌های آموزشی بود تا مدل بتواند ویژگی‌های اصلی اشیاء را مستقل از تغییرات ظاهری جزئی یاد بگیرد. نکته‌ی مهم آن است که افزایش مصنوعی داده‌ها فقط در مرحله‌ی آموزش فعال بود و روی تصاویر اعتبارسنجی و آزمون اعمال نشد؛ بنابراین معیارهای اعتبارسنجی و آزمون همواره روی تصاویر اصلی و دست‌نخورده محاسبه شدند. این موضوع سبب می‌شود ارزیابی عملکرد مدل منعکس کننده‌ی توان واقعی آن در مواجهه با داده‌های جدید باشد. در عمل، استفاده از افزایش مصنوعی داده‌ها معمولا باعث می‌شود دقت مجموعه‌ی آموزش اندکی کاهش یابد، زیرا مدل با نسخه‌های متنوع‌تر و دشوارتری از تصاویر مواجه می‌شود؛ اما در مقابل، توان تعمیم مدل افزایش یافته و عملکرد آن روی داده‌های اعتبارسنجی و آزمون پایدارتر می‌شود [۳۱].



شکل ۲- درصد تصاویر هر دسته در هر مجموعه‌داده، به غیر از دسته‌ی متفرقه

[1] Backbone
[2] Overfitting
[3] Stratified Split
[4] Data Augmentation

مطابق با جدول ۱ و ۲ و شکل ۲، تعداد تصاویر موجود در کلاس‌های مختلف یکسان نبود. در چنین شرایطی مدل ممکن است در طول آموزش بیشتر به سمت کلاس‌های پرتعداد متمایل شود و کلاس‌های کم‌نمونه را با دقت پایین‌تری شناسایی کند. برای کاهش این مشکل، از وزن‌دهی خودکار کلاس‌ها استفاده شد. در این روش، وزن هر کلاس بر اساس تعداد نمونه‌های موجود در مجموعه‌ی آموزش محاسبه می‌شود، به گونه‌ای که کلاس‌های کم‌نمونه وزن بیشتری در تابع زیان دریافت کنند. هرچه تعداد نمونه‌های یک کلاس کمتر باشد، سهم خطاهای آن کلاس در تابع زیان بیشتر خواهد بود. در نتیجه مدل هنگام آموزش توجه بیشتری به کلاس‌های کم‌نمونه خواهد داشت. در این پژوهش، از معماری EfficientNet-B0 به عنوان استخراج کننده‌ی ویژگی[1] استفاده شد. وزن‌های اولیه‌ی این بخش از مدل از نسخه‌ی پیش آموزش دیده روی مجموعه‌داده‌ی ImageNet دریافت شد و در طول آموزش ثابت نگه داشته شد. پس از بخش استخراج ویژگی، چند لایه‌ی اضافی برای انطباق مدل با مسئله‌ی طبقه‌بندی پسماند افزوده شد. تمامی تصاویر پیش از ورود به مدل به ابعاد ثابت ۲۲۴×۲۲۴ پیکسل تغییر اندازه داده شدند. استفاده از اندازه‌ی ورودی یکسان، امکان پردازش دسته‌ای داده‌ها و سازگاری کامل با معماری EfficientNet-B0 را فراهم می‌کند. پس از تغییر اندازه، مقادیر پیکسل‌ها مطابق رویه‌ی استاندارد خانواده‌ی EfficientNet نرمال‌سازی شدند تا توزیع داده‌های ورودی با داده‌هایی که مدل پایه در مرحله‌ی پیش‌آموزش مشاهده کرده است، همخوانی داشته باشد. این مرحله موجب پایداری بیشتر فرآیند آموزش و همگرایی مناسب‌تر مدل می‌شود [۱۸]. ساختار بخش طبقه‌بندی مدل در این پژوهش شامل لایه‌ی میانگین‌گیری سراسری دوبعدی، لایه‌ی نرمال‌سازی دسته‌ای، لایه‌ی حذف تصادفی نورون‌ها با نرخ ۰/۴۰، لایه‌ی کاملا متصل با ۵۱۲ نورون و تابع فعال‌سازی ReLU، لایه‌ی نرمال‌سازی دسته‌ای دوم، لایه‌ی Dropout دوم با نرخ ۰/۴۰، لایه‌ی خروجی Softmax با تعداد نورون برابر با تعداد کلاس‌های مسئله است. استفاده از Global Average Pooling باعث کاهش تعداد پارامترهای قابل آموزش و کاهش احتمال بیش‌برازش می‌شود. همچنین Dropout و نرمال‌سازی دسته‌ای به افزایش پایداری آموزش و بهبود تعمیم‌پذیری مدل کمک می‌کنند [۳۲].

در این پژوهش، مسئله‌ی طبقه‌بندی از نوع چندکلاسه بود؛ بنابراین از تابع زیان آنتروپی متقاطع چندکلاسه استفاده شد[2][۳۳].

$$L = -\sum_{i=1}^{c} y_i \log(\hat{y}_i) \quad (۱)$$

رابطه‌ی ۱ [۳۳]، فرم کلی تابع زیان را مشخص می‌کند که در آن، c تعداد کلاس‌ها، y_i برچسب حقیقی و $\hat{y}_i$ احتمال پیش‌بینی شده توسط مدل است. برای به‌روزرسانی وزن‌های شبکه از الگوریتم Adam Optimizer با نرخ یادگیری اولیه برابر با ۰/۰۰۰۱ استفاده شد. الگوریتم Adam ترکیبی از مزایای روش‌های AdaGrad و RMSProp است و به دلیل همگرایی سریع و پایداری مناسب، یکی از رایج‌ترین بهینه‌سازهای مورد استفاده در یادگیری عمیق محسوب می‌شود [۳۴ و ۳۵]. برای جلوگیری از ادامه‌ی آموزش در شرایط عدم بهبود عملکرد مدل و پیشگیری از بیش‌برازش، از راهبرد توقف زودهنگام استفاده شد. در این تحقیق اگر دقت مجموعه‌ی اعتبارسنجی طی پانزده دوره‌ی آموزشی متوالی[3] بهبود نداشته باشد، فرآیند آموزش به طور خودکار متوقف می‌شد. همچنین از راهبرد کاهش خودکار نرخ یادگیری استفاده شد. در این تحقیق اگر مقدار تابع زیان اعتبارسنجی طی پنج دوره‌ی متوالی کاهش پیدا نکند، نرخ یادگیری به نصف مقدار قبلی کاهش می‌یابد. این راهبرد به مدل کمک می‌کند در مراحل پایانی آموزش به نقاط بهینه‌ی دقیق‌تری همگرا شود [۳۶]. همچنین در طول آموزش، سه نسخه از مدل ذخیره می‌شود که عبارت اند از بهترین مدل بر اساس بیشترین دقت اعتبارسنجی، بهترین مدل بر اساس کمترین زیان اعتبارسنجی و مدل نهایی حاصل از آخرین دور آموزش. پس از پایان آموزش، هر سه نسخه به صورت مستقل روی مجموعه‌ی آزمون ارزیابی می‌شوند.

در آموزش مدل YOLO، وزن‌های مربوط به بهترین و آخرین دور آموزش ذخیره شدند. در چارچوب پیاده‌سازی Ultralytics، وزن بهترین مدل بر اساس معیار داخلی انتخاب مدل در فرآیند اعتبارسنجی تعیین می‌شود که mAP50-95 در آن نقش کلیدی دارد [۲۴]. در ادامه، این معیار و سایر شاخص‌های ارزیابی تشخیص شیء معرفی می‌شوند. پس از پایان آموزش، وزن‌های بهترین مدل برای ارزیابی نهایی روی مجموعه‌ی آزمون مورد استفاده قرار گرفتند. در این پژوهش تصاویر ورودی مدل YOLO26s با اندازه‌ی ۶۴۰ × ۶۴۰ پیکسل به مدل ارائه شدند.

---

[1] Feature Extractor
[2] Sparse Categorical Cross-Entropy
[3] Epoch

۳-۴- معیارهای ارزیابی مدل

برای سنجش کارآیی و کیفیت مدل‌ها از چند معیار ارزیابی استفاده شده است که در ادامه، شرح داده خواهند شد. این معیارها در مقالات پیشین، معرفی شده اند و کاربرد آن‌ها در پژوهش‌هایی همچون پژوهش حاضر، مشخص شده است [۳۷].

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (۲)$$

$$Precision = \frac{TP}{TP+FP} \quad (۳)$$

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

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

در روابط این بخش، TP بیانگر تعداد نمونه‌های مثبت درست طبقه‌بندی شده، TN بیانگر تعداد نمونه‌های منفی درست طبقه‌بندی شده، FP بیانگر تعداد نمونه‌های منفی که به اشتباه مثبت تشخیص داده شده اند و FN بیانگر تعداد نمونه‌های مثبت که به اشتباه منفی تشخیص داده شده اند است. رابطه ۲ یا دقت، نسبت کل پیش‌بینی‌های صحیح به تعداد کل نمونه‌ها را نشان می‌دهد و تصویری کلی از عملکرد مدل ارائه می‌کند. با این حال در مسائل چندکلاسه و به ویژه در مجموعه‌داده‌های نامتوازن، اتکا به دقت به تنهایی ممکن است گمراه کننده باشد؛ زیرا عملکرد ضعیف مدل در کلاس‌های کم‌نمونه می‌تواند در مقدار کلی دقت پنهان شود. مطابق با رابطه ۳، معیار صحت نشان می‌دهد که از میان نمونه‌هایی که مدل به یک کلاس خاص نسبت داده است، چه نسبتی واقعا متعلق به همان کلاس بوده اند. بنابراین این معیار میزان قابل اعتماد بودن پیش‌بینی‌های مدل را ارزیابی می‌کند. مطابق با رابطه ۴، معیار یادآوری نشان می‌دهد که از میان تمام نمونه‌های واقعی یک کلاس، چه نسبتی به درستی توسط مدل شناسایی شده اند. این معیار توانایی مدل در کشف نمونه‌های متعلق به هر کلاس را نشان می‌دهد. مطابق با رابطه ۵، امتیاز F1 میانگین هارمونیک Precision و Recall است و زمانی مقدار بالایی خواهد داشت که هر دو معیار به صورت هم‌زمان عملکرد مناسبی داشته باشند. از این رو، F1 یکی از متداول‌ترین معیارهای ارزیابی در مسائل طبقه‌بندی محسوب می‌شود. از آنجا که مسئله‌ی حاضر یک مسئله‌ی طبقه‌بندی چند کلاسه با توزیع نامتوازن داده‌ها است، از میانگین کلان[1]و میانگین وزنی[2]استفاده شد. در میانگین کلان، تمامی کلاس‌ها دارای وزن یکسان هستند و عملکرد مدل در کلاس‌های کم‌نمونه و پرنمونه به یک اندازه در نتیجه‌ی نهایی تاثیر می‌گذارد. در مقابل، میانگین وزنی با در نظر گرفتن تعداد نمونه‌های هر کلاس محاسبه می‌شود و بنابراین سهم کلاس‌های پرتعداد در مقدار نهایی بیشتر خواهد بود. استفاده‌ی هم‌زمان از چند معیار، امکان ارزیابی جامع‌تری از عملکرد مدل را فراهم می‌کند. هر یک از این معیارها جنبه متفاوتی از رفتار مدل را توصیف می‌کنند و گزارش مجموعه‌ای از آن‌ها باعث می‌شود نقاط قوت و ضعف مدل با دقت بیشتری مشخص شود و از نتیجه‌گیری‌های گمراه‌کننده ناشی از اتکا به یک معیار منفرد جلوگیری گردد [۳۸].

۳-۴-۲- معیارهای ارزیابی مدل YOLO

در این پژوهش به دلیل ماهیت مسئله که از نوع تشخیص شیء[3]است، ارزیابی مدل فقط بر پایه‌ی درستی برچسب کلاس انجام نشد؛ زیرا در تشخیص شیء علاوه بر کلاس، مکان جعبه‌ی پیش‌بینی شده[4]نیز باید با شیء واقعی هم‌پوشانی مناسب داشته باشد. از این رو، از شاخص‌های ویژه‌ی تشخیص شیء شامل هم‌پوشانی اشتراک و یا Intersection over Union: IoU، میانگین دقت و یا Average Precision: AP و میانگین دقت میان‌کلاسه و یا mean Average Precision: mAP و صحت و یادآوری استفاده شد.

[1] Macro Average
[2] Weighted Average
[3] Object Detection
[4] Bounding Box

IoU میزان هم‌پوشانی بین جعبه‌ی پیش‌بینی شده و جعبه‌ی واقعی را نشان می‌دهد و پایه‌ی اصلی ارزیابی مکان‌یابی در تشخیص شیء است [۳۷].

$$IoU = \frac{Area(B \cap B^*)}{\text{Area}(B \cup B^*)} \quad (۶)$$

در این رابطه [۳۷]، B جعبه‌ی پیش‌بینی شده توسط مدل و $\hat{B}$ جعبه‌ی واقعی است. IoU نشان می‌دهد جعبه‌ی پیش‌بینی شده تا چه حد با موقعیت واقعی شیء هم‌پوشانی دارد و پایه‌ی محاسبه‌ی AP و mAP است. **صحت** نشان می‌دهد از میان پیش‌بینی‌های مدل، چه سهمی درست بوده است و یادآوری نشان می‌دهد از میان اشیاء واقعی، چه سهمی توسط مدل پیدا شده است. برای هر کلاس، ابتدا منحنی صحت-یادآوری رسم می‌شود و سپس مساحت زیر این منحنی به عنوان AP محاسبه می‌شود:

$$AP_c = \int_0^1 P_c\,(R)dR \quad (۷)$$

در اینجا، Pc(R) منحنی صحت-یادآوری برای کلاس c است. AP مساحت زیر این منحنی را نشان می‌دهد.
سپس میانگین AP روی همه‌ی کلاس‌ها به صورت زیر محاسبه می‌شود [۳۸]:

$$mAP50 = \frac{1}{C} \sum_{c=1}^{C} AP_c \mid_{IoU=0.50} \quad (۸)$$

$$mAP50-95 = \frac{1}{10C} * \sum_{c=1}^{C} \sum_{t=0}^{9} AP_c \Big|_{IoU=0.50+0.05t} \quad (۹)$$

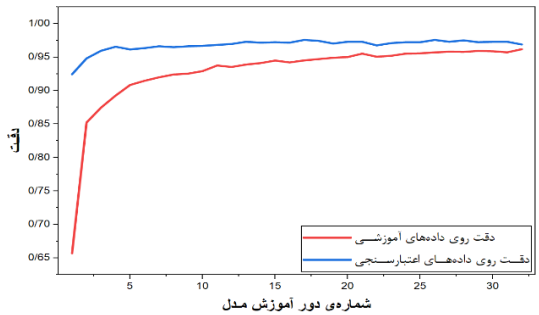
در این فرمول‌ها، C تعداد کلاس‌ها است. mAP50 میانگین AP در آستانه‌ی IoU برابر ۰/۵۰ را نشان می‌دهد، و mAP50-95 میانگین AP را در ده آستانه‌ی ۰/۵۰ تا ۰/۹۵با گام ۰/۰۵ گزارش می‌کند؛ بنابراین mAP50-95 از mAP50 سخت‌گیرانه‌تر و جامع‌تر است.

۴- نتایج

۴-۱- آموزش مدل روی مجموعه‌داده‌ی Garbage Classification

در این پژوهش، مطابق با مرحله‌ی ۱ از شکل ۱، در ابتدا مجموعه‌داده‌های معرفی شده در بخش ۲ تحقیق، بارگیری شدند. سپس مطابق با مرحله‌ی ۲ از شکل ۱، و با استفاده از الگوی معرفی شده در بخش ۳-۲ تحقیق، غربالگری تکرار و شباهت انجام گرفت. در گام بعدی مطابق با مرحله‌ی ۳ از شکل ۱، مدل GC11-ENB0[1] طراحی و آموزش داده شد. این مدل روی مجموعه‌داده‌ی Garbage Classification که در بخش ۲-۱ تحقیق معرفی شد آموزش داده شد. به جهت تمرکز مدل روی تصاویر با برچسب دقیق، دسته‌ی trash کنار گذاشته شد و مدل ساخته شده، پتانسیل طبقه‌بندی تصاویر در ۱۱ دسته‌ی باقی مانده را خواهد داشت. البته این امکان وجود دارد که درهنگام استفاده از مدل، شرطی تعیین کرد که اگر مدل تصویر خاصی را با اطمینان مشخصی در هیچکدام از ۱۱ دسته قرار نداد، دسته‌ی متفرقه به تصویر اطلاق شود. برای ساخت مدل، از معماری EfficientNet-B0 و روش انتقال یادگیری که در بخش مقدمه تحقیق معرفی شدند استفاده شد و مدل طبق ساختار معرفی شده در بخش ۳-۳ تحقیق، ایجاد شد. شکل ۳، نمودار دقت روی داده‌های آموزشی و اعتبارسنجی را در حین آموزش مدل نشان می‌دهد.

[1] Garbage Classification, 11 classes, EfficientNetB0



شکل ۳- نمودار تغییرات دقت در حین آموزش مدل GC11-ENB0

در شکل ۳ خط قرمز رنگ، دقت روی داده‌های آموزشی و خط آبی رنگ، دقت روی داده‌های اعتبارسنجی را نشان می‌دهد. همچنین محور عمودی، دقت را در بازه‌ی ۰ تا ۱ و محور افقی شماره‌ی دورهای آموزش مدل را نمایش داده است. در حین آموزش مدل، بهترین دقت اعتبارسنجی در دور شماره ۱۷ رخ داد و سپس با بهتر نشدن دقت در ۱۵ دور بعدی، آموزش مدل در پایان دور ۳۲ به صورت خودکار متوقف شد. در این پژوهش، دقت به عنوان معیار کلی عملکرد و نیز سه معیار صحت، یادآوری و امتیاز F1 در دو شکل میانگین کلان و میانگین وزنی گزارش شدند تا هم عملکرد برابر برای همه کلاس‌ها و هم اثر نامتوازن بودن تعداد نمونه‌ها بررسی شود. همانطور که در بخش ۳-۳ تحقیق ذکر شد، ۱۰ درصد از داده‌ها در هر دسته برای آزمون نهایی کنار گذاشته شدند و در حین آموزش به مدل داده نشدند. جدول ۳ نتایج سه نسخه‌ی ذخیره شده از مدل را روی مجموعه‌ی آزمون نشان می‌دهد شامل بهترین مدل بر اساس دقت اعتبارسنجی، بهترین مدل بر اساس کمترین خطای اعتبارسنجی و مدل نهایی.

جدول ۳- نتایج ارزیابی نهایی مدل GC11-ENB0

معیار ارزیابی / مدل	بهترین/دقت اعتبارسنجی	بهترین/خطای اعتبارسنجی	مدل پایان آموزش
دقت	۰/۹۶۸۹۶۱	۰/۹۷۰۳۱۰	۰/۹۷۰۳۱۰
صحت میانگین کلان	۰/۹۵۳۹۵۶	۰/۹۵۶۷۲۷	۰/۹۵۶۷۲۷
یادآوری میانگین کلان	۰/۹۵۶۷۲۷	۰/۹۵۷۲۷۹	۰/۹۵۷۲۷۹
امتیاز 1 میانگین کلان	۰/۹۵۴۸۷۳	۰/۹۵۶۸۵۴	۰/۹۵۶۸۵۴
صحت میانگین وزنی	۰/۹۶۹۷۸۰	۰/۹۷۰۶۹۷	۰/۹۷۰۶۹۷
یادآوری میانگین وزنی	۰/۹۶۸۹۶۱	۰/۹۷۰۳۱۰	۰/۹۷۰۳۱۰
امتیاز 1 میانگین وزنی	۰/۹۶۹۰۹۹	۰/۹۷۰۳۹۰	۰/۹۷۰۳۹۰

معیارهای ارزیابی مدل در بخش ۳-۴ معرفی شدند. جدول ۳ انواع معیارهای ارزیابی مدل GC11-ENB0 را نمایش می‌دهد؛ مقدار بیشینه‌ی هر کدام ۱ و کمینه‌ی آن ۰ است. مطابق با جدول ۳ مدل از کیفیت خیلی خوبی برخوردار است. مشاهده می‌گردد که مدل ذخیره شده براساس بهترین و یا کم‌ترین خطای اعتبارسنجی که همان مدل پایان آموزش بوده است، عملکرد بهتری روی داده‌های آزمون داشته است. دقت کارآمدترین نسخه‌ی مدل، ۰/۹۷۰۳۱۰ است که یعنی مدل توانسته در ۹۷/۰۳ درصد تصاویر، برچسب صحیح را به درستی تشخیص دهد. از جدول ۳ می‌توان مشاهده کرد که سایر معیارها نیز مقادیر بالا و نزدیک بیشینه دارند.

۴-۲- بررسی اهمیت استفاده از معماری EfficientNet همراه با انتقال یادگیری

در گام بعدی، مطابق مرحله‌ی ۴ از شکل ۱، مدل GC11-SCNN[1] طراحی و آموزش داده شد. هدف از توسعه‌ی این مدل، مقایسه‌ی تجربی یک مدل آموزش دیده از صفر با مدل GC11-ENB0 بود که بر پایه‌ی معماری EfficientNet-B0 و روش انتقال یادگیری توسعه یافته است. در مدل GC11-SCNN از معماری EfficientNet و وزن‌های از پیش آموزش دیده استفاده نشد و تمامی لایه‌های مدل از ابتدا آموزش داده شدند. با این حال، سایر اجزای اصلی فرآیند آموزش شامل مجموعه‌داده، تقسیم‌بندی داده‌ها، پیش‌پردازش،

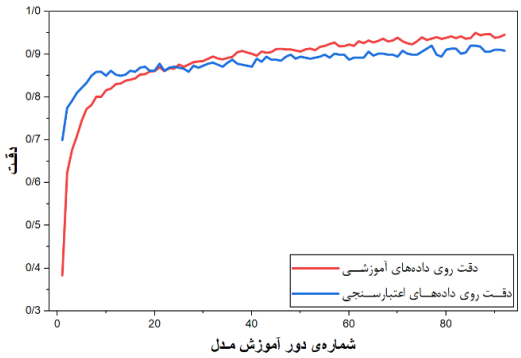
[1] Garbage Classification, 11 classes, Scratch CNN

افزایش مصنوعی داده، معیارهای ارزیابی، تنظیمات آموزش و شرایط اجرای آزمایش تا حد امکان یکسان نگه داشته شد تا مقایسه‌ی نتایج دو رویکرد با حداقل عوامل مداخله‌گر انجام شود. در طول آموزش و در ۱۰ دور نخست، مدل GC11-ENB0 توانست به بیشترین دقت اعتبارسنجی برابر با ۹۶/۶۲۶ درصد در دور دهم دست یابد، در حالی که مدل GC11-SCNN پس از ۱۰ دور آموزش، بیشترین دقت اعتبارسنجی برابر با ۶۱/۰۶۶ درصد را در دور نهم کسب کرد. این اختلاف نشان می‌دهد که در مسئله‌ی طبقه‌بندی تصاویر پسماند و با حجم داده‌ی مورد استفاده در این پژوهش، بهره‌گیری از معماری از پیش آموزش دیده‌ی EfficientNet-B0 همراه با انتقال یادگیری، موجب همگرایی سریع‌تر مدل و دست‌یابی به دقت بالاتر در مراحل ابتدایی آموزش شده است. میانگین زمان هر دور آموزش در مدل GC11-SCNN برابر با ۳۴۵۷/۹۴ ثانیه و در مدل GC11-ENB0 برابر با ۵۰۲/۶۹ ثانیه اندازه‌گیری شد؛ بنابراین زمان مورد نیاز برای هر دور آموزش در مدل GC11-SCNN حدود ۶/۸۷ برابر بیشتر از مدل GC11-ENB0 بود. در مدل GC11-ENB0، از معماری EfficientNet-B0 با وزن‌های از پیش آموزش دیده مورد استفاده قرار گرفت و مطابق تنظیمات آموزش، وزن‌های این بخش در مرحله‌ی اولیه‌ی آموزش ثابت نگه داشته شدند؛ در نتیجه در این قسمت نیازی به محاسبه‌ی گرادیان و انجام پس‌انتشار خطا نبود و تنها لایه‌های انتهایی طبقه‌بندی بازآموزی شدند. در مقابل، در مدل GC11-SCNN تمامی لایه‌ها از ابتدا قابل آموزش بودند و در هر دور آموزش، عملیات به‌روزرسانی پارامترها برای کل شبکه انجام شد. بخش قابل آموزش مدل GC11-ENB0 تنها شامل لایه‌های انتهایی طبقه‌بندی بود و تعداد پارامترهای قابل آموزش آن به مراتب کمتر از مدل GC11-SCNN است. در نتیجه، حجم محاسبات مورد نیاز در هر دور آموزش در مدل GC11-SCNN بیشتر است. بنابراین کاهش زمان هر دور آموزش در مدل GC11-ENB0 را باید حاصل استفاده از انتقال یادگیری و کاهش بار محاسباتی ناشی از ثابت بودن بخش استخراج کننده‌ی ویژگی دانست. همان‌گونه که در بخش مقدمه بیان شد، وزن‌های اولیه‌ی EfficientNet-B0 از آموزش بر روی بیش از ۱/۲ میلیون تصویر از مجموعه‌داده‌ی ImageNet حاصل شده اند و توانایی استخراج ویژگی‌های عمومی تصاویر، مانند لبه‌ها، بافت‌ها، الگوهای هندسی و ساختارهای پایه را کسب کرده اند. در مدل GC11-ENB0 این دانش پیشین برای حل مسئله‌ی جدید مورد استفاده قرار گرفت و تنها لایه‌های نهایی متناسب با طبقه‌بندی پسماند بازآموزی شدند. در مقابل، مدل GC11-SCNN تمامی این ویژگی‌ها را تنها بر اساس تصاویر موجود در مجموعه‌داده‌ی Garbage Classification فرا گرفت. از آنجا که حجم و تنوع مجموعه‌داده‌ی ImageNet به مراتب بیشتر از تصاویر آموزشی این پژوهش است، انتظار می‌رود مدل مبتنی بر انتقال یادگیری با سرعت بیشتری همگرا شود و در مراحل اولیه‌ی آموزش به عملکرد بهتری دست یابد. نتایج تجربی این پژوهش نیز این انتظار را تایید می‌کند؛ به طوری که مدل GC11-ENB0 در همان ۱۰ دور نخست، به دقت اعتبارسنجی بسیار بالاتری نسبت به مدل آموزش دیده از صفر دست یافت. در پژوهش‌های یادگیری عمیق و بینایی ماشین، علاوه بر دقت، هزینه‌ی محاسباتی و زمان آموزش نیز از شاخص‌های مهم ارزیابی محسوب می‌شوند؛ زیرا در بسیاری از کاربردها مدل‌ها باید روی هزاران، ده‌ها هزار، صدها هزار و حتی میلیون‌ها تصویر آموزش ببینند. در چنین شرایطی، دست‌یابی هم‌زمان به همگرایی سریع‌تر، دقت مناسب و کاهش هزینه‌ی محاسباتی می‌تواند فرآیند توسعه و به‌کارگیری عملی مدل را تسهیل کند. نتایج این بخش نشان داد که در مسئله‌ی مورد بررسی، استفاده از معماری EfficientNet-B0 همراه با انتقال یادگیری، راهکاری موثر برای دست‌یابی به این اهداف بوده است.

۴-۳- آموزش مدل روی مجموعه‌داده‌ی Realwaste

در گام بعدی تحقیق، مطابق با مرحله‌ی ۶ از شکل ۱، مدل با به‌کارگیری معماری EfficientNet-B0 و روش انتقال یادگیری، این بار بر روی مجموعه‌داده‌ی Realwaste ساخته شد. در ساخت مدل RW8-ENB0[1]، دسته‌ی Miscellaneous Trash حذف گردید تا تمرکز مدل بر روی دسته‌هایی با برچسب دقیق باشد. بنابراین مدل تصاویر را در ۸ دسته‌ی باقی مانده که در جدول ۲ از بخش ۲-۲ تحقیق معرفی شدند، طبقه‌بندی می‌کند. آموزش مدل RW8-ENB0، تا پایان دور شماره‌ی ۹۲، ادامه داشت و به دلیل بهبود پیدا نکردن دقت روی داده‌های اعتبارسنجی در ۱۵ دور متوالی، آموزش به صورت خودکار متوقف شد تا از احتمال بیش‌برازش و کاهش کارآیی مدل جلوگیری شود. شکل ۴، نمودار دقت روی داده‌های آموزشی و اعتبارسنجی را در حین آموزش مدل نشان می‌دهد.

[1] RealWaste, 8 classes, EfficientNetB0



شکل ۴- نمودار تغییرات دقت در حین آموزش مدل RW8-ENB0

جدول ۴ نتایج حاصل از اعمال مدل RW8-ENB0 بر روی ۱۰ درصد از داده‌های کنار گذاشته شده برای آزمون نهایی را نشان می‌دهد.

جدول ۴- نتایج ارزیابی نهایی مدل RW8-ENB0

معیار ارزیابی / مدل	بهترین/دقت اعتبارسنجی	بهترین/خطای اعتبارسنجی	مدل پایان آموزش
دقت	۰/۸۹۹۰۶۱	۰/۹۰۳۷۵۶	۰/۹۱۳۱۴۶
صحت میانگین کلان	۰/۹۰۴۵۱۰	۰/۹۱۰۷۵۸	۰/۹۱۶۷۴۴
یادآوری میانگین کلان	۰/۹۱۲۶۶۷	۰/۹۱۵۲۸۶	۰/۹۲۴۰۰۰
امتیاز F1 میانگین کلان	۰/۹۰۷۵۱۱	۰/۹۱۱۳۰۴	۰/۹۱۹۱۲۷
صحت میانگین وزنی	۰/۹۰۰۵۸۱	۰/۹۰۶۵۹۷	۰/۹۱۴۹۵۰
یادآوری میانگین وزنی	۰/۸۹۹۰۶۱	۰/۹۰۳۷۵۶	۰/۹۱۳۱۴۶
امتیاز F1 میانگین وزنی	۰/۸۹۸۷۴۲	۰/۹۰۳۶۱۸	۰/۹۱۲۹۴۱

از جدول ۴ می‌توان نتیجه گرفت که کارآمدترین نسخه‌ی مدل RW8-ENB0 دارای دقت ۹۱/۳۱ درصدی بر روی داده‌های آزمون بوده است و سایر معیارها نیز همگی بالای ۰/۹ و نزدیک به مقدار بیشینه‌ی ۱ هستند. بنابراین مدل از کیفیت خیلی خوبی برخوردار است.

۴-۴- ارزیابی برون‌مجموعه‌ای مدل‌ها و تحلیل نتایج

مشاهده شد که مدل GC11-ENB0 در طبقه‌بندی تصاویر مجموعه‌داده‌ی Garbage Classification در ۱۱ دسته‌ی معرفی شده در جدول ۱ به غیر از دسته‌ی متفرقه، و مدل RW8-ENB0 در طبقه‌بندی تصاویر مجموعه‌داده‌ی Realwaste در ۸ دسته‌ی معرفی شده در جدول ۲ به غیر از دسته‌ی متفرقه، کارآیی خیلی خوبی داشته اند. در گام بعدی تحقیق، مطابق مرحله‌ی ۷ از شکل ۱، عملکرد مدل‌ها به صورت برون‌مجموعه‌ای نیز بررسی شد. مدل GC11-ENB0 بر روی مجموعه‌داده‌ی Realwaste و مدل RW8-ENB0 بر روی مجموعه‌داده‌ی Garbage Classification اعمال شدند. برای این کار، تنها دسته‌هایی از مجموعه‌داده‌ی دیگر بررسی شدند که از نظر ماهیت با کلاس‌های آموزشی مدل متناظر بودند. به عنوان مثال، مجموعه‌داده‌ی Realwaste دسته‌ای شامل تصاویر باتری ندارد، بنابراین مدل ساخته شده با تصاویر Realwaste یعنی RW8-ENB0 بر روی دسته‌ی شامل تصاویر باتری در Garbage Classification یعنی دسته‌ی battery ارزیابی نمی‌گردد. در گام نخست، مدل GC11-ENB0 بر روی دسته‌های مناسب در Realwaste اعمال شد. لازم به ذکر است که مدل GC11-ENB0، تصاویر شیشه را به سه دسته‌ی متفاوت طبقه‌بندی می‌کند و چون مجموعه‌داده‌ی Realwaste یک دسته شیشه دارد، هر کدام از دسته‌های شیشه‌ی پیش‌بینی شده توسط مدل، در دسته‌ی شیشه و یا Glass لحاظ شدند. جدول ۵ نتایج حاصل از اعمال مدل GC11-ENB0 بر روی برخی دسته‌های مجموعه‌داده‌ی Realwaste را نشان می‌دهد.

جدول ۵- نتایج اعمال مدل GC11-ENB0 بر روی مجموعه‌داده‌ی Realwaste

نام دسته در Realwaste	سه دسته با بیشترین پیش‌بینی توسط مدل GC11-ENB0		
Metal (فلز)	Metal ۷۱/۲۷% (فلز)	Plastic ۱۴/۰۵% (پلاستیک)	Cardboard ۳/۹۲% (کارتن)
Plastic (پلاستیک)	Plastic ۶۴/۸۲% (پلاستیک)	Metal ۱۶/۸۳% (فلز)	Glass ۸/۷۹% (شیشه)
Glass (شیشه)	Glass ۵۲/۸۶% (شیشه)	Plastic ۳۰/۷۱% (پلاستیک)	Metal ۹/۵۲% (فلز)
Paper (کاغذ)	Paper ۴۶/۶۰% (کاغذ)	Plastic ۱۹/۰۰% (پلاستیک)	Cardboard ۱۸/۰۰% (کارتن)
Cardboard (کارتن)	Plastic ۳۵/۳۶% (پلاستیک)	Cardboard ۳۳/۱۹% (کارتن)	Paper ۱۴/۹۷% (کاغذ)

مجموعه‌داده‌های Garbage Classification و Realwaste، دو مجموعه‌داده‌ی کاملاً مستقل هستند. قبل‌تر مشخص شد که مدل GC11-ENB0 بر روی تصاویر جدید در مجموعه‌داده‌ی Garbage Classification عملکرد خیلی خوبی داشت و توانست با دقت ۹۷/۰۳ درصدی، دسته‌ی تصاویر را تشخیص دهد. همچنین مدل RW8-ENB0 توانست تصاویر جدید در مجموعه‌داده‌ی Realwaste را با دقت ۹۱/۳۱ درصدی طبقه‌بندی کند که نشان می‌دهد هر دو مدل واقعاً روی تصاویری متناظر با تصاویر موجود در مجموعه‌داده‌ای که با کمک آن ساخته شدند، به خوبی آموزش دیده اند. در ادامه و پس از تحلیل برون‌مجموعه‌ای، از جدول ۳ مشخص است که مدل GC11-ENB0، که با تصاویر مجموعه‌داده‌ی Garbage Classification توسعه داده شد، بر روی دسته‌ی Metal از مجموعه‌داده‌ی Realwaste بهترین عملکرد را داشت و ۷۱/۲۷ درصد تصاویر این دسته در مجموعه‌داده‌ی خارجی را به درستی طبقه‌بندی کرده است. بدترین عملکرد مدل GC11-ENB0 بر روی یک دسته از مجموعه‌داده‌ی Realwaste متعلق به دسته‌ی Cardboard می‌باشد که مدل با اختلاف اندکی، حتی یک دسته‌ی متفاوت را بیشتر از دسته‌ی Cardboard پیش‌بینی کرده است و ۳۳/۱۹ درصد تصاویر را در دسته‌ی Cardboard طبقه‌بندی کرده است. به جهت ایجاد امکان مقایسه، در این مرحله، مدل RW8-ENB0 روی چند دسته از مجموعه‌داده‌ی Garbage Classification اعمال شد که جدول ۶ نتایج را نمایش می‌دهد.

جدول ۶- نتایج اعمال مدل RW8-ENB0 بر روی مجموعه‌داده‌ی Garbage Classification

نام دسته در Garbage Classification	سه دسته با بیشترین پیش‌بینی توسط مدل RW8-ENB0		
cardboard (کارتن)	Cardboard ۸۹/۷۹% (کارتن)	Paper ۷/۳۰% (کاغذ)	Metal ۱/۲۳% (فلز)
metal (فلز)	Metal ۷۴/۷۷% (فلز)	Cardboard ۱۶/۶۴% (کارتن)	Paper ۲/۹۹% (کاغذ)
paper (کاغذ)	Paper ۶۳/۰۵% (کاغذ)	Cardboard ۳۲/۴۸% (کارتن)	Textile Trash ۲/۶۷% (پسماند نساجی)
white-glass (شیشه سفید)	Glass ۵۸/۸۴% (شیشه)	Metal ۱۲/۱۳% (فلز)	Cardboard ۱۰/۴۵% (کارتن)
brown-glass (شیشه قهوه‌ای)	Glass ۵۴/۳۷% (شیشه)	Cardboard ۱۷/۷۹% (کارتن)	Metal ۱۰/۰۵% (فلز)
green-glass (شیشه سبز)	Glass ۴۹/۱۳% (شیشه)	Plastic ۱۳/۹۹% (پلاستیک)	Cardboard ۱۲/۵۶% (کارتن)
plastic (پلاستیک)	Plastic ۴۶/۴۷% (پلاستیک)	Cardboard ۱۷/۹۲% (کارتن)	Glass ۱۶/۵۳% (شیشه)

از جدول ۶ می‌توان برداشت کرد که مدل RW8-ENB0 بر روی مجموعه‌داده‌ی Garbage Classification عملکرد بهتری داشته است. بهترین عملکرد مربوط به دسته‌ی Cardboard با دقت ۸۹/۷۹ درصدی است و بدترین عملکرد روی دسته‌ی Plastic با

دقت ۴۶/۴۷ درصدی بوده است. نکته‌ی قابل توجه در مقایسه‌ی جدول‌های ۵ و ۶، عملکرد بسیار متفاوت مدل‌ها روی دسته‌ی Cardboard می‌باشد. مدلی که با مجموعه‌داده‌ی Garbage Classification ساخته شده است، روی دسته‌ی Cardboard در مجموعه‌داده‌ی Realwaste، دقت نه چندان خوب ۳۳/۱۹ درصدی داشته است؛ از طرفی مدلی که با مجموعه‌داده‌ی Realwaste ساخته شده است، روی دسته‌ی Cardboard از مجموعه‌داده‌ی Garbage Classification، دقت ۸۹/۷۹ درصدی داشته است. شکل ۵، چند مورد از تصاویر دسته‌ی Cardboard در این دو مجموعه‌داده‌ی متفاوت را نشان می‌دهد.



شکل ۵- مقایسه‌ی تصاویر دسته‌ی Cardboard در دو مجموعه‌داده

چهار تصویر سمت راست در شکل ۵، مربوط به مجموعه‌داده‌ی Realwaste و چهار تصویر سمت چپ، مربوط به مجموعه‌داده‌ی Garbage Classification می‌باشد. همانطور که از شکل ۵ نیز مشخص است، تصاویر مجموعه‌داده‌ی Realwaste در دسته‌ی Cardboard، نسبت به تصاویر همین دسته در مجموعه‌ی Garbage Classification، از تنوع بالاتری برخوردار هستند. تصاویر مجموعه‌داده‌ی Realwaste، انواع کارتن‌ها و پاکت‌های محصولات غذایی و سایر محصولات، به شکل‌های مختلف و در رنگ‌های متنوع را دربرمی‌گیرند، درحالی که تصاویر مجموعه‌داده‌ی Garbage Classification عمدتاً کارتن‌های بسته‌بندی وسایل را دربرمی‌گیرند؛ این موضوع را می‌توان یکی از مهم‌ترین و کلیدی‌ترین علل عملکرد بهتر مدل RW8-ENB0 در تشخیص دسته‌ی Cardboard در Garbage Classification دانست. مدل GC11-ENB0 روی تصاویر با تنوع بسیار کمتری آموزش دیده است، به همین دلیل در تشخیص تصاویر متنوع دسته‌ی Cardboard در Realwaste عملکرد ضعیف‌تری داشته است. بنابراین یکی از علت‌های اصلی عملکرد بهتر مدل RW8-ENB0 روی مجموعه‌داده‌ی خارجی در قیاس با مدل GC11-ENB0، تنوع بیشتر تصاویر مجموعه‌داده‌ی Realwaste است. در اینجا یک مفهوم مهم در پردازش تصویر مطرح می‌شود و آن تلاش برای انطباق حداکثری تصاویری که قرار است توسط مدل بررسی شوند با تصاویری است که مدل، روی آن‌ها آموزش دیده است. این انطباق به معنای یکسان بودن تصاویر نیست بلکه به معنای انطباق در ماهیت تصاویر است؛ به عنوان مثال اگر مدلی روی تصاویر پسماند آموزش ببیند، نمی‌تواند در طبقه‌بندی تصاویر حیوانات خوب عمل کند. در نتیجه وقتی قرار است از مدلی در پردازش تصویر استفاده شود، باید مدل با انواع متفاوت از تصاویری که قرار است بعداً روی آن‌ها اعمال شود آموزش داده شود. می‌توان گفت هرچه نمونه‌های آموزشی یک مدل از نظر ماهیت و تنوع به نمونه‌های هدف نزدیک‌تر باشند، تعمیم‌پذیری مدل در ارزیابی برون‌مجموعه‌ای بهتر خواهد بود.

۴-۵- ایجاد مجموعه‌داده‌ی ترکیبی

پس از بررسی اهمیت تنوع تصاویر، مطابق با مرحله‌ی ۸ از شکل ۱، مجموعه‌داده‌ی MWD7[1] که یک مجموعه‌داده‌ی ترکیبی براساس مجموعه‌داده‌های قبلی می‌باشد، ایجاد شد تا بتوان مدل جدیدی ساخت که روی برخی دسته‌ها، با تصاویر بیشتر و متنوع‌تری آموزش ببیند. جدول ۷ دسته‌های مجموعه‌داده‌ی جدید و تعداد تصاویر هر دسته را نمایش می‌دهد.

[1] Mixed Waste Dataset, 7 classes

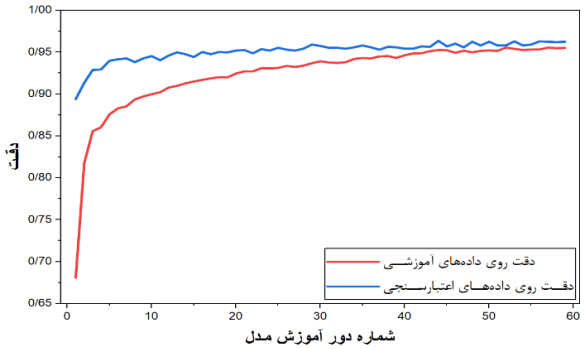
ردیف	نام دسته	معادل فارسی	تعداد تصویر	مجموعه‌داده‌ی بخش ۲-۲ تحقیق (RW)	مجموعه‌داده‌ی بخش ۲-۱ تحقیق (GC)
۱	Cardboard	کارتن	۱۳۵۲	Cardboard (کارتن)	cardboard (کارتن)
۲	Glass	شیشه	۲۴۳۱	Glass (شیشه)	brown glass, (شیشه قهوه‌ای) green glass, (شیشه سبز) white glass (شیشه سفید)
۳	Metal	فلز	۱۵۵۹	Metal (فلز)	metal (فلز)
۴	Organics	پسماند آلی	۱۸۳۲	Food Organics, (پسماند آلی غذایی) Vegetation (پسماند گیاهی)	biological (زیستی)
۵	Paper	کاغذ	۱۵۵۰	Paper (کاغذ)	paper (کاغذ)
۶	Plastic	پلاستیک	۱۷۸۶	Plastic (پلاستیک)	plastic (پلاستیک)
۷	Textile-Wearables	پسماند نساجی و پوشاک	۷۶۲۰	Textile Trash (پسماند نساجی)	clothes, (پوشاک) shoes (کفش)
تعداد کل تصاویر			۱۸۱۳۰		

جدول ۷- معرفی دسته‌های مجموعه‌داده‌ی MWD7

مطابق با جدول ۷، مجموعه‌داده‌ی MWD7 با ۱۸۱۳۰ تصویر در ۷ دسته، شکل گرفت. دو ستون سمت چپ جدول ۷ نشان می‌دهند هر دسته از MWD7، از ترکیب تصاویر کدام دسته‌های Realwaste و Garbage Classification شکل گرفته است. همانطور که از جدول ۷ نیز مشخص است، تصاویر دسته‌ی battery در Garbage Classification به دلیل نداشتن دسته‌ی هم‌خانواده و متناظر در Realwaste، با تصاویر هیچ دسته‌ای ترکیب نشدند و از مجموعه‌داده‌ی MWD7 کنار گذاشته شدند. البته مدل GC11-ENB0 که قبل‌تر معرفی شد، روی دسته‌ی battery نیز آموزش دیده است.

۴-۶- آموزش مدل روی مجموعه‌داده‌ی ترکیبی

پس از ایجاد مجموعه‌داده‌ی جدید، مطابق با مرحله‌ی ۹ از شکل ۱، مدل پردازش تصویر[1] MWD7-ENB0 با معماری EfficientNet-B0 و روش انتقال یادگیری روی MWD7 توسعه داده شد. شکل ۶، نمودار دقت روی داده‌های آموزشی و اعتبارسنجی را در حین آموزش مدل نشان می‌دهد.



[1] Mixed Waste Dataset, 7 classes, EfficientNetB0

شکل ۶- نمودار تغییرات دقت در حین آموزش مدل MWD7-ENB0

مدل MWD7-ENB0 پس از پایان دور آموزشی شماره ۵۹، به صورت خودکار متوقف شد و سه نسخه از مدل، ذخیره شدند. جدول ۸ نتایج حاصل از اعمال مدل بر روی ۱۰ درصد داده‌های آزمون نهایی را نشان می‌دهد.

جدول ۸- نتایج ارزیابی نهایی مدل MWD7-ENB0

معیار ارزیابی / مدل	بهترین/دقت اعتبارسنجی	بهترین/خطای اعتبارسنجی	مدل پایان آموزش
دقت	۰/۹۵۹۷۳۵	۰/۹۶۳۰۴۵	۰/۹۶۳۰۴۵
صحت میانگین کلان	۰/۹۳۹۲۲۲	۰/۹۴۳۴۹۷	۰/۹۴۳۴۹۷
یادآوری میانگین کلان	۰/۹۴۵۲۹۵	۰/۹۴۹۶۶۴	۰/۹۴۹۶۶۴
امتیاز ۱ میانگین کلان	۰/۹۴۲۰۷۱	۰/۹۴۶۳۶۱	۰/۹۴۶۳۶۱
صحت میانگین وزنی	۰/۹۶۰۲۰۲	۰/۹۶۳۶۴۹	۰/۹۶۳۶۴۹
یادآوری میانگین وزنی	۰/۹۵۹۷۳۵	۰/۹۶۳۰۴۵	۰/۹۶۳۰۴۵
امتیاز۱ میانگین وزنی	۰/۹۵۹۸۵۳	۰/۹۶۳۲۱۱	۰/۹۶۳۲۱۱

از جدول ۸، مشخص است که مدل MWD7-ENB0، موفق شد با دقت ۹۶/۳۰ درصدی تصاویر آزمون را در دسته‌های ۷گانه‌ی معرفی شده در جدول ۷، طبقه‌بندی کند. سایر معیارها نیز نزدیک به مقدار بیشینه قرار دارند بنابراین مدل از کیفیت خیلی خوبی در طبقه‌بندی ۱۸۱۳۰ تصویر مجموعه‌داده‌ی MWD7 برخوردار است.

۴-۷- بررسی تاثیر استفاده از معماری سایر اعضای خانواده‌ی EfficientNet در کیفیت مدل

در بخش مقدمه تحقیق عنوان شد که مدل‌های EfficientNet یک خانواده با ۸ عضو از B0 تا B7 هستند. در این بخش از پژوهش، مطابق با مرحله‌ی ۱۰ از شکل ۱، تاثیر استفاده از عضوی دیگری از این خانواده در کارآیی مدل در این پژوهش بررسی می‌گردد. برای این گام از پژوهش مدل EfficientNet-B2 انتخاب گردید؛ علت این انتخاب این بود که با عبور از مدل EfficientNet-B1 و انتخاب معماری EfficientNet-B2، تفاوت با معماری قبلی محسوس‌تر خواهد بود و از طرفی معماری انتخاب شده، همانند EfficientNet-B7 سنگین و با پارامترهای خیلی زیاد نخواهد بود. در خانواده‌ی EfficientNet، با افزایش شماره‌ی مدل، اندازه‌ی ورودی متعارف نیز افزایش می‌یابد و برای نسخه‌های B0، B1 و B2 به ترتیب ۲۲۴، ۲۴۰ و ۲۶۰ پیکسل گزارش شده است [۱۸]. این مرحله از تحقیق در دو حالت مختلف انجام شد؛ در حالت اول برای مقایسه‌ی مدل‌های EfficientNet-B0 و EfficientNet-B2، سایر تنظیمات آموزشی ثابت نگه داشته شد و هر دو مدل با اندازه‌ی ورودی ثابت ۲۲۴×۲۲۴ پیکسل آموزش داده شدند؛ بنابراین، تفاوت‌های مشاهده شده را باید در چارچوب همین شرایط تفسیر کرد. در حالت دوم، EfficientNet-B2 با اندازه‌ی ورودی ۲۶۰×۲۶۰ پیکسل نیز آموزش داده شد تا امکان مقایسه‌ی با نتایج حالت اول فراهم شود. در مرحله‌ی اول، معماری EfficientNet-B2 و روش انتقال یادگیری روی مجموعه‌داده‌ی MWD7 پیاده‌سازی شد. آموزش پس از پایان دور شماره ۳۹، متوقف و هر سه نسخه از مدل ذخیره شدند. جدول ۹ نتایج حاصل از آزمون نهایی مدل[1] MWD7-ENB2-224 را نمایش می‌دهد.

جدول ۹- نتایج ارزیابی نهایی مدل MWD7-ENB2-224

معیار ارزیابی / مدل	بهترین/دقت اعتبارسنجی	بهترین/خطای اعتبارسنجی	مدل پایان آموزش
دقت	۰/۹۵۳۶۶۸	۰/۹۵۴۷۷۱	۰/۹۵۴۷۷۱
صحت میانگین کلان	۰/۹۲۹۵۵۸	۰/۹۳۰۶۶۳	۰/۹۳۰۶۶۳
یادآوری میانگین کلان	۰/۹۳۸۸۹۸	۰/۹۳۷۶۳۱	۰/۹۳۷۶۳۱

[1] Mixed Waste Dataset, 7 classes, EfficientNetB2, 224×224

امتیاز F1 میانگین کلان	۰/۹۳۳۹۰۴	۰/۹۳۳۸۶۲	۰/۹۳۳۸۶۲
صحت میانگین وزنی	۰/۹۵۴۵۰۶	۰/۹۵۵۵۷۲	۰/۹۵۵۵۷۲
یادآوری میانگین وزنی	۰/۹۵۳۶۶۸	۰/۹۵۴۷۷۱	۰/۹۵۴۷۷۱
امتیاز F1 میانگین وزنی	۰/۹۵۳۸۶۴	۰/۹۵۴۹۸۷	۰/۹۵۴۹۸۷

مطابق با مرحله‌ی ۱۱ از شکل ۱، با مقایسه‌ی جدول ۸ که معیارهای حاصل از آزمون نهایی مدل توسعه داده شده با استفاده از معماری EfficientNet-B0 بر روی مجموعه‌داده‌ی MWD7 است، با جدول ۹ که معیارهای حاصل از آزمون نهایی مدل توسعه داده شده با استفاده از معماری EfficientNet-B2 بر روی همان مجموعه‌داده است؛ می‌توان نتیجه گرفت مدل MWD7-ENB2-224 با این‌که از کیفیت خوبی برخوردار است، اما پارامترهای حاصل از آزمون مدل قبلی، مقداری برتری دارند. بنابراین در شرایط این پژوهش، با تعداد تصاویر تحقیق حاضر و با اندازه‌ی ورودی یکسان ۲۲۴×۲۲۴ برای هر دو معماری، معماری EfficientNet-B0 برای استفاده در روش انتقال یادگیری کافی است و در این شرایط نیازی به استفاده از مدل‌های سنگین‌تر خانواده‌ی EfficientNet احساس نمی‌شود. سپس با اندازه ورودی استاندارد، معماری EfficientNet-B2 و روش انتقال یادگیری روی مجموعه‌داده‌ی MWD7 پیاده‌سازی شد. در این حالت آموزش پس از دور شماره ۵۵، متوقف و هر سه نسخه از مدل ذخیره شدند. جدول ۱۰ نتایج حاصل از آزمون نهایی مدل MWD7-ENB2-260[1] را نمایش می‌دهد.

جدول ۱۰- نتایج ارزیابی نهایی مدل MWD7-ENB2-260

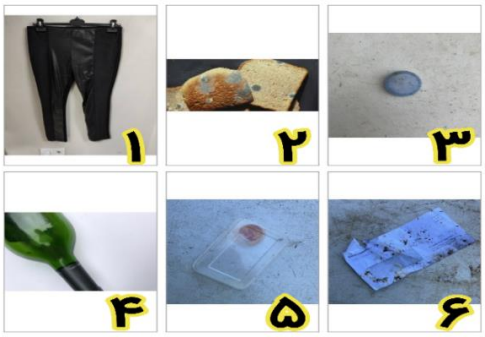
معیار ارزیابی / مدل	بهترین/دقت اعتبارسنجی	بهترین/خطای اعتبارسنجی	مدل پایان آموزش
دقت	۰/۹۶۹۱۱۲	۰/۹۷۰۲۱۵	۰/۹۷۰۲۱۵
صحت میانگین کلان	۰/۹۵۲۳۵۱	۰/۹۵۴۲۷۷	۰/۹۵۴۲۷۷
یادآوری میانگین کلان	۰/۹۵۹۸۷۴	۰/۹۶۱۲۴۰	۰/۹۶۱۲۴۰
امتیاز F1 میانگین کلان	۰/۹۵۶۰۱۸	۰/۹۵۷۵۶۰	۰/۹۵۷۵۶۰
صحت میانگین وزنی	۰/۹۶۹۵۶۵	۰/۹۷۰۸۱۰	۰/۹۷۰۸۱۰
یادآوری میانگین وزنی	۰/۹۶۹۱۱۲	۰/۹۷۰۲۱۵	۰/۹۷۰۲۱۵
امتیاز F1 میانگین وزنی	۰/۹۶۹۲۶۵	۰/۹۷۰۳۵۰	۰/۹۷۰۳۵۰

با مقایسه‌ی جدول ۱۰ با جدول‌های ۸ و ۹، می‌توان نتیجه گرفت که مدل MWD7-ENB2-260 بهترین کارآیی را داشته است، مدل با معماری B2، توانست به دقت ۹۷/۰۲ درصدی در طبقه‌بندی تصاویر مجموعه‌داده‌ی MWD7 برسد و از مدل با معماری B0 عبور کرد. این موضوع اهمیت دو چیز را نشان می‌دهد، ابتدا این‌که معماری با پارامترهای بیشتر توانست هرچند کوچک، مقداری معیارهای ارزیابی را بهبود دهد و مورد بعدی این‌که استفاده از اندازه‌ی ورودی‌های استاندارد و متناسب با هر مدل، می‌تواند در نتایج نهایی مدل تاثیرگذار باشد؛ همانطور که در این‌جا مدل با معماری EfficientNet-B2 و اندازه‌ی ورودی ۲۲۴×۲۲۴ نسبت به مدل سبک‌تر بهبودی ایجاد نکرد اما همین معماری با اندازه‌ی ورودی استاندارد خودش، معیارهای ارزیابی را بهبود بخشید.

۴-۸- بررسی معماری EfficientNet در تحلیل تصاویری با حضور هم‌زمان چند نوع پسماند

مطابق با مرحله‌ی ۱ از شکل ۱ در فاز دوم تحقیق، جهت ارزیابی کارآیی مدل MWD7-ENB0 که با معماری EfficientNet آموزش دیده بر روی تصویری که چندین نوع پسماند متفاوت را دربرمی‌گیرد، ابتدا به صورت تصادفی عددی بین ۲ تا ۷ انتخاب گردید و سپس در مجموعه‌داده‌ی MWD7، معادل با عدد تصادفی انتخاب شده، از بین ده درصد از تصاویری که در حین آموزش به مدل داده نشدند، تصاویر تصادفی انتخاب شدند تا ارزیابی بدون اعمال نظر شخصی باشد و همچنین تصاویری بررسی شوند که برای مدل، جدید باشند. شکل ۷، تصاویر انتخاب شده را مشخص می‌کند.

[1] Mixed Waste Dataset, 7 classes, EfficientNetB2, 260×260



شکل ۷- تصاویر انتخاب شده برای آزمایش

مدل MWD7-ENB0 در دو مرحله، ابتدا تک به تک روی تصاویر شکل ۷ اعمال شد و سپس روی تصویر چندشیئی شامل تمام تصاویر اعمال شد.

جدول ۱۱- اطمینان مدل MWD7-ENB0 در طبقه‌بندی هر کدام از تصاویر شکل ۷ در تمامی دسته‌ها

شماره تصویر	اطمینان مدل جهت نسبت دادن تصویر به هر کدام از ۷ دسته در حالت تک شیء (اعمال مدل روی هر تصویر جداگانه) در بازه ۰ تا ۱						
۱	Textile-Wearables = ۰/۹۹۹۹۶۸۱۷	Plastic = ۰/۰۰۰۰۰۰۲۴	Paper = ۰/۰۰۰۰۰۸۷۴	Organics = ۰/۰۰۰۰۰۲۹۲	Metal = ۰/۰۰۰۰۰۲۳۶	Glass = ۰/۰۰۰۰۰۰۲۹	Cardboard = ۰/۰۰۰۰۱۷۲۰
۲	Textile-Wearables = ۰/۰۰۰۰۰۰۴۷	Plastic = ۰/۰۰۰۰۰۰۴۷	Paper = ۰/۰۰۰۰۰۰۰۰	Organics = ۰/۹۹۹۹۹۸۲۱	Metal = ۰/۰۰۰۰۰۰۰۱	Glass = ۰/۰۰۰۰۰۰۸۲	Cardboard = ۰/۰۰۰۰۰۰۰۱
۳	Textile-Wearables = ۰/۰۰۰۳۷۴۷۵	Plastic = ۰/۶۷۸۳۶۴۳۴	Paper = ۰/۰۰۱۶۲۷۳۰	Organics = ۰/۰۰۱۷۴۰۶۸	Metal = ۰/۱۳۶۸۵۷۱۴	Glass = ۰/۱۷۸۰۸۹۹۲	Cardboard = ۰/۰۰۲۹۴۵۹۱
۴	Textile-Wearables = ۰/۰۰۰۰۰۰۰۱	Plastic = ۰/۰۰۰۰۱۴۱۰	Paper = ۰/۰۰۰۰۰۰۱۳	Organics = ۰/۰۰۰۰۸۶۰۸	Metal = ۰/۰۰۰۰۰۰۰۱	Glass = ۰/۹۹۹۸۹۹۷۴	Cardboard = ۰/۰۰۰۰۰۰۰۱
۵	Textile-Wearables = ۰/۰۰۰۰۰۰۰۰	Plastic = ۰/۹۹۹۹۱۹۱۸	Paper = ۰/۰۰۰۰۰۰۰۱	Organics = ۰/۰۰۰۰۱۱۲۵	Metal = ۰/۰۰۰۰۰۰۳۲	Glass = ۰/۰۰۰۰۶۷۵۳	Cardboard = ۰/۰۰۰۰۰۱۷۱
۶	Textile-Wearables = ۰/۰۰۰۰۰۲۲۵	Plastic = ۰/۰۰۰۱۱۷۹۴	Paper = ۰/۹۷۰۷۰۲۴۷	Organics = ۰/۰۰۰۰۰۰۰۲	Metal = ۰/۰۰۰۰۰۰۳۷	Glass = ۰/۰۰۰۰۰۰۲۹	Cardboard = ۰/۰۲۹۱۷۶۶۱

راهنما: Textile-Wearables معادل پسماند نساجی و پوشاک، Plastic معادل پلاستیک، Organics معادل پسماند آلی، Metal معادل فلز، Glass معادل شیشه و Cardboard معادل کارتن می‌باشند.

جدول ۱۱ نتایج حاصل از اعمال مدل MWD7-ENB0 روی تمامی تصاویر شکل ۷، به صورت جداگانه را نمایش می‌دهد. مطابق با جدول ۱۱، مدل در تمامی موارد، دسته‌ی صحیح را پیش‌بینی کرده است. دسته‌ای که مدل اطمینان بالاتری به آن نسبت داده با رنگ زرد مشخص شده است. همچنین مدل در اکثر موارد با اطمینان خیلی بالایی دسته‌ی صحیح را پیش‌بینی کرده است؛ و به طور میانگین مدل تقریباً با اطمینان ۹۴ درصدی موفق شده است برچسب صحیح را تولید کند. جدول ۱۲ نشان از عملکرد مدل روی تصویری است که هم‌زمان شامل هر ۶ تصویر بوده است.

جدول ۱۲- اطمینان مدل MWD7-ENB0 در طبقه‌بندی تصویر ترکیبی در تمامی دسته‌ها

تصویر	اطمینان مدل جهت نسبت دادن تصویر ترکیبی شامل هر ۶ تصویر به هر کدام از ۷ دسته در حالت چند شیئی - در بازه ۰ تا ۱						
تصویر ترکیبی	Textile-Wearables = ۰/۲۳۶۸۹۵۹۶	Plastic = ۰/۱۵۷۲۴۳۵۸	Paper = ۰/۴۷۹۶۱۸۰۴	Organics = ۰/۰۴۴۴۸۵۶۵	Metal = ۰/۰۰۰۳۳۸۳۳	Glass = ۰/۰۶۳۸۶۴۹۲	Cardboard = ۰/۰۱۷۵۵۳۵۰

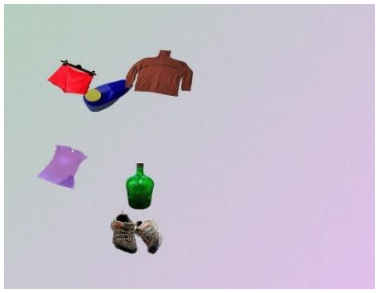
راهنما: Textile-Wearables معادل پسماند نساجی و پوشاک، Plastic معادل پلاستیک، Organics معادل پسماند آلی، Metal معادل فلز، Glassمعادل شیشه و Cardboard معادل کارتن می‌باشند.

از شکل ۷ و جدول ۱۱ مشخص است که تصاویر انتخاب شده، در ۵ دسته‌ی متنوع قرار دارند؛ در جدول ۱۱ اطمینان مدل برای این ۵ دسته، زمانی که تصویر هر کدام از این دسته‌ها، جداگانه بررسی شده، عدد بالایی است. درصورتی که این اطمینان‌ها در حالت تصویر چندشیئی، مطابق با جدول ۱۲ به طور قابل توجهی کاهش داشته اند. در حالت تحلیل تصویر چندشیئی، مدل برای آن دو دسته‌ای که تصویری از آن‌ها نبوده است کمترین اطمینان را دارد یعنی دسته‌های Cardboard و Metal، اما همچنان اطمینان دسته‌های Glass و Organics که تصویری از آن‌ها موجود است نیز بسیار پایین می‌باشد و سایر دسته‌ها نیز با این‌که اطمینان بالاتری دارند اما در قیاس با حالت تحلیل تک‌شیئی، اطمینان مدل در تشخیص آن‌ها کاهش شدیدی داشته است. بالاترین اطمینان مدل در تحلیل تصویر چندشیئی ۴۷/۹۶ درصد بوده درحالی که مدل در تحلیل تصاویر تک شیئی اکثرا با ۹۹ درصد اطمینان مواجه بوده است. این شواهد نشان از این دارد که معماری EfficientNet در تحلیل تصاویر با تک برچسب و یک نوع از پسماند، عملکرد عالی دارد اما زمانی که تصاویر پیچیده و چندشیئی می‌شوند، دیگر این معماری به اندازه کافی قابل اعتماد نیست. به منظور بررسی تکرار پذیر بودن شواهد، این آزمایش علاوه بر موردی که ذکر شد، ۹ بار دیگر به طور کامل تکرار شد و هربار همین نتیجه به دست آمد.

۴-۹- ایجاد مجموعه‌داده‌ی چندشیئی با حضور هم‌زمان چند نوع پسماند

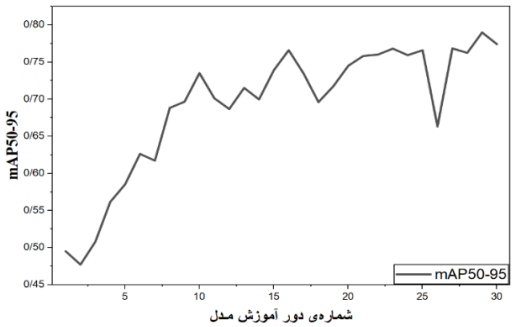
در این بخش از پژوهش و مطابق با مرحله‌ی ۲ از شکل ۱، مجموعه‌داده‌ی تصاویر چندشیئی تشکیل شد. از آنجا که تصاویر مورد بررسی تک برچسبی هستند و هر تصویر یک نوع پسماند را مشخص می‌کند، برای تحلیل تصاویر پیچیده باید ابتدا تصاویری شامل چندین شیء ایجاد کرد. ۱۸۱۳۰ تصویر مجموعه‌داده‌ی MWD7 به سه بخش آموزش، اعتبارسنجی و آزمون با نسبت‌های ۸۰، ۱۰ و ۱۰ درصد تقسیم شدند. سپس برای هر بخش به طور مستقل، تصاویر ترکیبی چندشیئی ساخته شد؛ به گونه‌ای که هر تصویر ترکیبی شامل ۲ تا ۷ شیء بود که تصادفی انتخاب شدند و هر تصویر خام فقط یک بار استفاده شد. این راهبرد باعث شد هم تنوع نمونه‌ها حفظ شود و هم از نشت داده میان بخش‌های آموزش، اعتبارسنجی و آزمون جلوگیری شود. برای هر تصویر انتخاب شده، ابتدا پس‌زمینه حذف شد، سپس مجموعه‌ای از تبدیلات شامل وارون‌سازی افقی، تغییر روشنایی، تغییر کنتراست، تغییر رنگ، چرخش تصادفی و افزودن نویز اعمال شد تا شباهت به شرایط واقعی افزایش پیدا کند. پس از آن، شیء روی یک پس‌زمینه‌ی جدید در کنار سایر اشیاء قرار گرفت. اندازه‌ی اشیاء نیز متناسب با تعداد اجزای هر تصویر چندشیئی تنظیم شد تا تراکم صحنه طبیعی‌تر شود. برای هر شیء، علاوه بر ذخیره‌ی موقعیت آن در تصویر ترکیبی، فایل برچسب استاندارد YOLO یعنی همان قالبی که YOLO برای آموزش نیاز دارد نیز تولید شد. از آنجا که مجموعه‌داده‌ی مبدا، یعنی MWD7، دارای ۱۸۱۳۰ تصویر است، تعداد تصاویر مجموعه‌داده‌ی مقصد یعنی MOWD[1] وابسته به الگوی تصادفی انتخاب تعداد شیء حاضر در هر تصویر چندشیئی است. اگر تمام تصاویر ترکیبی با حداقل تعداد شیء یعنی دو مورد ساخته می‌شدند، حداکثر تعداد تصاویر مجموعه‌داده‌ی چندشیئی یعنی ۹۰۶۵ تصویر حاصل می‌شد و در صورتی که هر تصویر چندشیئی با حداکثر تعداد شیء یعنی هفت مورد ساخته می‌شد، حداقل تعداد تصاویر مجموعه‌داده‌ی چندشیئی یعنی ۲۵۹۰ تصویر در مجموعه‌داده‌ی چندشیئی ایجاد می‌گردید. باتوجه به تصادفی بودن فرآیند، انتظار می‌رفت عددی که در اجرا حاصل می‌شود، چیزی بین این دو عدد باشد که در عمل نیز ۴۰۵۱ تصویر ایجاد شد. موقعیت قرارگیری اشیاء در پس‌زمینه‌ی جدید، تصادفی اما کنترل شده بود؛ به طوری که هر تصویر جدید اضافه شده، نباید بیش از حد با تصاویر قبلی هم‌پوشانی پیدا می‌کرد به گونه‌ای که تصویر قبلی قابل مشاهده نباشد. شکل ۸، یک نمونه از این تصاویر را نمایش می‌دهد.

[1] Multi-Object Waste Dataset



شکل ۸: نمونه‌ای از تصویر چندشیئی تولید شده

این تصاویر چندشیئی به عنوان مجموعه‌داده‌ی MOWD ذخیره شدند. از تعداد ۴۰۵۱ تصویر چندشیئی، ۳۲۵۱ تصویر آموزشی، ۳۹۹ تصویر اعتبارسنجی و ۴۰۱ تصویر برای آزمون نهایی درنظر گرفته شد. پس از ایجاد مجموعه‌داده‌ی چندشیئی، مطابق با مرحله‌ی ۳ از شکل ۱ در فاز دوم تحقیق، با معماری YOLO26s و روش انتقال یادگیری، مدل تشخیص شیء آموزش داده شد. شکل ۹، نمودار دورهای آموزش براساس معیار mAP50-95 برای داده‌های اعتبارسنجی در طول آموزش مدل را نمایش می‌دهد.



شکل ۹: نمودار تغییرات mAP50-95 در حین آموزش مدل چندشیئی

مطابق با شکل ۹، می‌توان مشاهده کرد که معیار mAP50-95 روند کلی صعودی داشته و با ادامه‌ی آموزش نیز همچنان این احتمال وجود دارد که مدل بتواند معیارهای ارزیابی را بهبود بخشد، به نوعی می‌توان گفت که شرط توقف زودهنگام تا پایان آموزش مدل، فعال نشد. بنابراین مدل ایجاد شده در این بخش یعنی مدل MOWD-YOLO[1] مدل بهینه نیست اما یک مدل خوب است. برای آموزش مدل MOWD-YOLO تعداد ۳۰ دور آموزش تعیین شد و آموزش بیش از این ادامه پیدا نکرد. از دلایل این موضوع می‌توان به چند مورد اشاره کرد؛ از شکل ۹ می‌توان متوجه شد که شیب کلی افزایش معیار mAP50-95 با افزایش تعداد دورهای آموزش کاهش پیدا کرده و این بدان معنا است که مدل تقریبا به بهینه شدن نزدیک شده است. دلیل دیگر توقف آموزش مدل، این است که در دورهای آخر آموزش مدل، مقدار mAP50-95 به بیش از ۰/۷۵ رسیده است که مقداری مناسب و باکیفیت در مسائل تشخیص اشیاء می‌باشد. اما یکی از دلایل کلیدی در اکتفا کردن به ۳۰ دور آموزشی درکنار رسیدن مدل به کیفیت مناسب، مدت زمان طولانی آموزش بود. مدل جدید در قیاس با مدل‌های قبلی پردازش‌های پیچیده‌تری انجام می‌دهد، بنابراین سرعت آموزش مدل، کندتر است. هر دور از آموزش و ارزیابی روی داده‌های اعتبارسنجی مدل MOWD-YOLO بین ۱ ساعت و ۴۸ دقیقه و ۴۰ ثانیه تا ۲ ساعت و ۳۲ دقیقه و ۵ ثانیه زمان برد و این بدان معناست که با توجه به ۳۰ دور آموزش انجام شده، کل زمان آموزش مدل تشخیص اشیاء در این تحقیق، بین ۲ تا ۳ شبانه‌روز کامل بوده است. البته ذکر این نکته ضروری است که مدت زمان آموزش وابسته به عوامل متعدد، از جمله قدرت سخت افزاری سیستم مورد استفاده می‌تواند متفاوت باشد؛ اعداد ذکر شده، تجربه‌ی تحقیق حاضر هستند. همچنین آخرین وزن‌های مدل در پایان دور ۳۰ ام، ذخیره شده اند و در صورت نیاز به بهبود معیارهای مدل تشخیص شیء، می‌توان از انتهای دور ۳۰ ام، آموزش مدل را ادامه داد و نیازی به آموزش از ابتدا نیست. همانند مدل‌های ساخته شده با معماری EfficientNet، مدل ساخته

[1] Multi-Object Waste Dataset, YOLO

شده با معماری YOLO نیز پس از اتمام آموزش، روی تصاویر آزمون که مدل درحین آموزش هرگز آن‌ها را ندیده است ارزیابی نهایی شد، جدول ۱۳ نتایج این ارزیابی را نشان می‌دهد.

جدول ۱۳- ارزیابی نهایی مدل MOWD-YOLO

معیار ارزیابی مدل	مقدار
mAP50-95	۰/۷۹۲
mAP50	۰/۸۲۳
صحت	۰/۷۵۵
یادآوری	۰/۷۵۸

باتوجه به جدول ۱۳، پس از ۳۰ دور آموزش مدل، تمامی معیارهای صحت، یادآوری، mAP50 و mAP50-95 بالای ۰/۷۵ هستند و معیار کلیدی mAP50-95 روی تصاویر آزمون، مقدار ۰/۷۹۲ داشته است که مقدار خوبی است و نشان از کیفیت خوب مدل دارد. تعدادی از مهم‌ترین نتایج تحقیق در جدول ۱۴ آورده شده اند.

جدول ۱۴- اقدامات و نتایج به دست آمده از تحقیق

نام مدل / اقدام	نتیجه
GC11-ENB0	رسیدن به دقت ۹۷/۰۳ درصدی در طبقه‌بندی تصاویر مجموعه‌داده‌ی Garbage Classification به‌جز دسته‌ی متفرقه
مقایسه‌ی GC11-ENB0 با GC11-SCNN	استفاده از معماری EfficientNet همراه با انتقال یادگیری و فریز کردن وزن‌های لایه‌های ابتدایی مدل، می‌تواند در مقایسه با آموزش مدل از صفر بدون استفاده از معماری EfficientNet موجب همگرایی سریع‌تر مدل و دست‌یابی به دقت بالاتر در مراحل ابتدایی آموزش و کاهش متوسط زمان هر دور آموزش مدل شود.
RW8-ENB0	رسیدن به دقت ۹۱/۳۱ درصدی در طبقه‌بندی تصاویر مجموعه‌داده‌ی Realwaste به‌جز دسته‌ی متفرقه
ارزیابی برون‌مجموعه‌ای	برجسته شدن اهمیت توجه دقیق به محتوای هر دسته و بررسی تنوع تصاویری که دربرمی‌گیرد، در استفاده از مدلی که با آن تصاویر آموزش داده شده است. مشخص شدن اهمیت تنوع تصاویر، در ارزیابی برون‌مجموعه‌ای با دامنه‌ی بیشتر. توسعه‌ی مجموعه‌داده‌ی MWD7 با تعداد و تنوع بیشتر در ۷ دسته‌ی تعریف شده.
MWD7-ENB0	رسیدن به دقت ۹۶/۳۰ درصدی در طبقه‌بندی تصاویر مجموعه‌داده‌MWD7
MWD7-ENB2-260	رسیدن به دقت ۹۷/۰۲ درصدی در طبقه‌بندی تصاویر مجموعه‌داده‌MWD7 و بهبود دقت حاصل از مدل B0
مقایسه‌ی MWD7-ENB0 با MWD7-ENB2-224 و MWD7-ENB2-260	استفاده از یکی از معماری‌های سنگین‌تر خانواده‌ی EfficientNet توانست کیفیت مدل را بهبود بخشد؛ اما این بهبود کیفیت، تنها در زمانی حاصل شد که از اندازه‌ی ورودی متفاوت با مدل سبک‌تر و متناسب با مدل سنگین‌تر استفاده شد.
پیاده‌سازی MWD7-ENB0 روی چند تصویر تصادفی از مجموعه‌ی آزمون در MWD7، در دو حالت: ۱. به طور جداگانه روی هر تصویر و ۲. روی تصویری با حضور هم‌زمان تمام تصاویر و تکرار آزمایش مجموعا به تعداد ۱۰ بار	معماری EfficientNet در تحلیل تصاویر با حضور هم‌زمان چندین نوع پسماند، محدودیت دارد و این تصاویر، منجر به کاهش قابل توجه اطمینان مدل می‌شوند. توسعه‌ی مجموعه‌داده‌ی MOWD با حضور هم‌زمان چندین نوع پسماند در یک تصویر.
MOWD-YOLO	رسیدن به mAP50-95 برابر با ۰/۷۹۲ در ۳۰ دور آموزش مدل تشخیص شیء با قابلیت تعیین نوع و موقعیت هر پسماند در تصویر چندشیئی بر روی مجموعه‌داده‌ی MOWD

۴-۱۰- مقایسه با برخی از مطالعات پیشین

در فاز اول تحقیق، این پژوهش مجموعاً شش مدل یادگیری عمیق را توسعه داد و با ترکیب کنترل کیفیت داده، مقایسه‌ی معماری‌ها و آزمون برون‌مجموعه‌ای، چارچوبی گسترده‌تر از بسیاری از مطالعات رایج داخلی و بین‌المللی ارائه کرد. در تحقیقات اخیر، پژوهش‌های

طبقه‌بندی پسماند با یادگیری عمیق رشد کرده اند، اما بخش قابل توجهی از آن‌ها هنوز بر یک مجموعه‌داده یا یک سناریوی آزمایشی محدود متمرکزند [۳۹]. برخی پژوهش‌های جدیدتر به سمت مقایسه‌ی چند مدل پیش آموزش دیده و چند مجموعه‌داده حرکت کرده اند؛ با این حال پروتکلی یکپارچه که هم‌زمان غربالگری صریح نشت اطلاعات، بررسی چند مجموعه‌داده و ارزیابی برون‌مجموعه‌ای، مقایسه‌ی آموزش از صفر در کنار انتقال یادگیری و مقایسه‌ی دو معماری در انتقال یادگیری را گزارش کند، کمتر دیده می‌شود [۴۲-۳۹]. همچنین معیارهای ارزیابی مدل‌ها در این تحقیق نشان می‌دهد مدل‌ها به خوبی آموزش دیده اند و از کیفیت مناسبی برخوردار هستند. بنابراین جای خالی پژوهش حاضر در بین مقالات این حوزه، به خصوص مقالات داخلی احساس می‌گردد و پژوهش حاضر تلاش می‌کند خلاهای قبلی را پر کند. هرچند پژوهش حاضر به همین‌جا ختم نمی‌شود و فاز دومی نیز برای آن طراحی شده است. در فاز دوم تحقیق، رویکرد متفاوتی نسبت به فاز اول مورد بررسی قرار گرفت. در ابتدا ضعف معماری EfficientNet در تحلیل تصاویر پیچیده با حضور هم‌زمان چندین نوع پسماند بررسی شد و سپس با خلق مجموعه‌داده‌ای متشکل از چندین نوع پسماند و استفاده از معماری تشخیص شیء YOLO، مدل جدیدی ایجاد شد که رویکردی متفاوت با مدل‌های قبل دارد و توانایی تشخیص نوع اشیاء و همینطور تشخیص موقعیت اشیاء در تصاویر را دارد، کارآیی مدل با انواع معیارها بررسی شد و نتایج نشان از کیفیت خوب مدل داشت. این مدل هفتمین مدل پردازش تصویر آموزش داده شده در پژوهش است که ساختار آن، می‌تواند فرآیند هوشمندسازی مدیریت پسماند را تسهیل بخشد. به طور مشخص‌تر در میان مطالعات قبلی، Zhang و همکاران در سال ۲۰۲۱ [۴۰] مدلی بر پایه‌ی معماری DenseNet169 و روش انتقال یادگیری، تنها روی یک مجموعه‌داده توسعه داده اند؛ Risfendra و همکاران در سال ۲۰۲۴ [۴۱] از معماری EfficientNet-B0 با انتقال یادگیری روی یک مجموعه‌داده‌ی شش‌کلاسه استفاده کرده اند و Kumala و همکاران در سال ۲۰۲۵ [۴۲] دو معماری MobileNetV2 و VGG16 را روی یک مجموعه‌داده با یکدیگر مقایسه کرده اند. در هیچ یک از این مطالعات، غربالگری صریح تکرار/شباهت میان داده‌ها، ارزیابی برون‌مجموعه‌ای مدل روی مجموعه‌داده‌ای مستقل، مقایسه‌ی تجربی آموزش از صفر در برابر انتقال یادگیری و یا گذار از طبقه‌بندی تک‌شیئی به تشخیص شیء در تصاویر چندشیئی گزارش نشده است.

در خصوص مطالعات داخلی، همانطور که در بخش مقدمه نیز اشاره شد، پژوهش‌های مبتنی بر یادگیری عمیق و تصاویر پسماند هنوز بسیار محدود هستند. در مقاله‌ای جدید که توسط بیات زاده و همکاران [۴۳]، منتشر شده است؛ مقاله بر یادگیری عمیق و تفکیک هوشمند زباله با استفاده از معماری پایه‌ی VGG16 تمرکز دارد. در این پژوهش که محققان عنوان می‌کنند پس از بررسی‌های انجام شده، در زمینه‌ی تفکیک زباله‌های خشک با استفاده از بینایی ماشین، کاری مشابه با پژوهش خود مشاهده نکرده اند؛ ۳۸۵۰ تصویر توسط محققان از Google جمع‌آوری و مدل با چهار طبقه‌ی بطری، شیشه، پاکت و قوطی ساخته شده است. علاوه بر تمامی نکاتی که درخصوص پژوهش حاضر بیان شد که نشان از تفاوت‌های این مقاله با مقاله‌ی بیات زاده و همکاران دارد، ذکر این نکته ضروری است که برخی مطالعات پیشین نشان داده اند که EfficientNet-B0 با تعداد پارامتر و هزینه‌ی محاسباتی کمتر نسبت به VGG16، می‌تواند در خیلی از مسائل عملکرد رقابتی یا حتی برتری داشته باشد [۱۸ و ۴۴]. در تحقیق داخلی دیگری، به بررسی اهمیت تفکیک هوشمند پسماند با استفاده از پردازش تصویر پرداخته شده است و فواید این کار، همچون بهینه شدن زمان و هزینه‌ها و شناسایی و تفکیک سریع پسماندهای قابل بازیافت مورد بررسی قرار گرفته اند [۴۵]. در یکی از مطالعات داخلی مرتبط با مدیریت هوشمند پسماند، ملک الکلامی و حسن‌زاده [۴۶]، رویکردی مبتنی بر اینترنت اشیاء را برای هوشمندسازی فرآیند جمع‌آوری پسماند ارائه کردند. در سامانه‌ی پیشنهادی این مطالعه، داده‌های حاصل از حسگرهای مرتبط با وضعیت مخازن از طریق اینترنت به سرور منتقل می‌شود و از این داده‌ها برای پایش وضعیت مخازن، انتخاب مخازن نیازمند جمع‌آوری، محاسبه‌ی مسیرهای جمع‌آوری و پیش‌بینی وضعیت آینده استفاده می‌شود. همچنین این مطالعه معماری لایه‌ای، زیرساخت‌های ارتباطی و الزامات سامانه‌های مدیریت هوشمند پسماند را مورد بحث قرار داده است. با این حال، ماهیت این مطالعه؛ تحقیقاتی و مفهومی است و تمرکز آن بر تبیین زیرساخت و ساز و کار یک سامانه‌ی مدیریت هوشمند مبتنی بر اینترنت اشیاء قرار دارد و در آن جمع‌آوری واقعی داده‌های حسگری، پیاده‌سازی عملی سامانه و توسعه و ارزیابی تجربی مدل‌ها گزارش نشده است. بنابراین تفاوت پژوهش حاضر با مطالعه‌ی مذکور تنها در استفاده از فناوری متفاوت نیست؛ بلکه در سطح اجرای پژوهش، نوع داده، روش تحلیل، فرآیند توسعه مدل و نوع خروجی نیز تفاوت اساسی وجود دارد. پژوهش حاضر یک مسیر تجربی از آماده‌سازی و کنترل کیفیت داده‌های تصویری تا توسعه، مقایسه و ارزیابی چندین مدل یادگیری عمیق و در نهایت تشخیص شیء در تصاویر چندشیئی را پیاده‌سازی کرده است. از این منظر، پژوهش حاضر را می‌توان یک گام اجرایی و تحلیلی در لایه‌ی بینایی ماشین مدیریت هوشمند

پسماند دانست که از نظر توسعه مدل، ایجاد مجموعه‌داده و اعتبارسنجی تجربی، دامنه‌ای فراتر از چارچوب مفهومی مورد بررسی در مطالعه‌ی پیشین دارد.

۵- بحث

در این پژوهش سعی شد رویکردی منسجم، در توسعه و ارزیابی مدل‌های پردازش تصویر پسماند در دو فاز مدل‌های طبقه‌بندی و مدل‌های تشخیص اشیاء، همراه با بارگیری دو مجموعه‌داده‌ی مستقل و توسعه‌ی یک مجموعه‌داده با تنوع بیشتر تصاویر و یک مجموعه‌داده‌ی چندشیئی، ایجاد شود. این تحقیق، روش انتقال یادگیری را با دو معماری برجسته‌ی EfficientNet و YOLO ادغام کرد و به نتایج قابل توجهی دست یافت. تحقیق حاضر می‌تواند یک گام موثر در تقویت شناخت فرآیند، همراه با اجرای عملی آن در راه هوشمندسازی مدیریت پسماند باشد که می‌تواند فواید متعدد و مهمی را در پی داشته باشد. با این حال، همانطور که قبل‌تر نیز در پژوهش عنوان شد، باید تعریف برچسب هر دسته، با دقت زیادی صورت گیرد. در پژوهش مشاهده شد که دسته‌ی Cardboard در هر دو مجموعه‌داده‌ی مستقل تحقیق دقیقاً با یک اسم یکسان وجود داشت، اما یک دسته دارای انواع متعددی از کارتن‌ها، پاکت‌ها در اندازه و شکل و رنگ‌های متفاوت بود و دسته‌ی دیگر تنوع بسیار کمتری داشت؛ نتایج نشان داد تعریف صحیح برچسب و استفاده از مدلی که انطباق بهتری با خواسته‌ی پژوهش دارد، در مشاهده‌ی کیفیت بهتر از مدل، تاثیر قابل‌توجهی دارد. بنابراین پیشنهاد می‌گردد در ساخت مدل‌های پردازش تصویر و یا ارزیابی مدل‌های موجود، حتماً قبل از انجام فرآیند، این مسئله به طور جدی بررسی شود. در این تحقیق مشاهده شد که برای تحلیل تصاویر چندشیئی، به صورت مصنوعی از تصاویر تک‌برچسبی، تصاویر چندشیئی ساخته شدند. تاحد امکان سعی شد تصاویر چندشیئی به واقعیت نزدیک باشند اما توصیه می‌گردد در پژوهش‌های آتی، تصاویر چندشیئی که ذاتاً ترکیبی از چند نوع پسماند به صورت هم‌زمان هستند نیز مورد استفاده قرار گیرند و مدل جدیدی با آن تصاویر آموزش داده شود و مقایسه با مدل معرفی شده در تحقیق حاضر صورت گیرد. در این پژوهش، انواع متعددی از مدل‌ها با اهداف مختلف توسعه داده شد، به دلیل محدودیت‌های سخت افزاری و سنگین و بسیار طولانی بودن فرآیند آموزش برخی مدل‌ها، شرط توقف زودهنگام در توسعه‌ی دو مدل GC11-SCNN و MOWD-YOLO، فعال نشد و مدل‌ها پس از تعداد مشخصی دور آموزش، متوقف شدند؛ بنابراین مدل‌های حاصل، ممکن است بهینه‌ترین مدل‌های قابل دست‌یابی با الگوی مورد استفاده‌ی پژوهش نباشند و پیشنهاد می‌گردد در پژوهش‌های آتی، این دو مدل تا زمان فعال شدن شرط توقف زودهنگام ادامه یابند و نتایج با پژوهش حاضر مقایسه شوند. در مقایسه‌ی مدل GC11-SCNN با GC11-ENB0، به دلیل محدودیت‌های سخت‌افزاری و زمان‌بر بودن آموزش مدل از صفر، آموزش GC11-SCNN تنها در ۱۰ دور ابتدایی انجام شد و با ۱۰ دور نخست آموزش GC11-ENB0 مقایسه گردید. بنابراین هدف این مقایسه ارزیابی عملکرد دو مدل در مراحل ابتدایی آموزش و بررسی سرعت همگرایی آن‌ها بود و نه مقایسه‌ی سقف نهایی دقت قابل دست‌یابی؛ نتایج نشان داد که مدل GC11-ENB0 در همان مراحل ابتدایی آموزش، همگرایی سریع‌تری داشته، به دقت اعتبارسنجی بالاتری دست یافته و میانگین زمان هر دور آموزش آن نیز کمتر بوده است. با این حال، از آنجا که مدل GC11-SCNN از ابتدا آموزش دیده و فرآیند آموزش آن پیش از رسیدن به معیار توقف زودهنگام متوقف شده است، نمی‌توان درباره‌ی حداکثر عملکرد قابل دست‌یابی این مدل قضاوت قطعی کرد. از این رو، پیشنهاد می‌شود در پژوهش‌های آینده، این مدل تا زمان فعال شدن شرط توقف زودهنگام آموزش داده شود تا امکان مقایسه‌ی کامل‌تر عملکرد نهایی دو رویکرد نیز فراهم شود.

۶- نتیجه‌گیری

در این پژوهش، با هدف کمک به ایجاد سامانه‌های تفکیک خودکار و پشتیبانی از مدیریت هوشمند پسماند؛ ابتدا دو مجموعه‌داده‌ی مستقل پسماند بارگیری و پس از انجام غربالگری تکرار و شباهت به منظور افزایش اعتبار نتایج، برای هر مجموعه‌داده به طور جداگانه، یک مدل طبقه‌بندی مبتنی بر معماری EfficientNet-B0 و روش انتقال یادگیری توسعه داده شد. نتایج نشان داد که این رویکرد در هر دو مجموعه‌داده، به عملکرد بسیار مطلوب مدل‌ها در طبقه‌بندی تصاویر پسماند منجر شده است. مدل GC11-ENB0 که روی ۱۴۸۱۸ تصویر مجموعه‌داده‌ی Garbage Classification، به جز دسته‌ی متفرقه آموزش دیده می‌تواند تصاویر پسماند را در ۱۱ دسته‌ی متفاوت پسماند و مدل RW8-ENB0 که روی ۴۲۵۷ تصویر مجموعه‌داده‌ی Realwaste، به جز دسته‌ی متفرقه آموزش دیده می‌تواند تصاویر پسماند را در ۸ دسته‌ی متفاوت پسماند طبقه‌بندی کند. همچنین مقایسه‌ی تجربی مدل مبتنی بر معماری

EfficientNet-B0 و انتقال یادگیری با CNN سفارشی آموزش دیده از صفر نشان داد که این مدل در مراحل ابتدایی آموزش، همگرایی سریع‌تر، دقت اعتبارسنجی بالاتر و میانگین زمان هر دور آموزش کمتری داشت؛ این نتایج نشان می‌دهد که در مسئله‌ی مورد بررسی، استفاده از معماری EfficientNet-B0 همراه با انتقال یادگیری، علاوه بر بهبود عملکرد مدل، از نظر کارآیی محاسباتی نیز رویکردی مناسب بوده است. علاوه بر این، ارزیابی برون‌مجموعه‌ای دو مدل GC11-ENB0 و RW8-ENB0 اهمیت ماهیت و تنوع تصاویر آموزشی را در مسائل پردازش تصویر نشان داد. بر همین اساس، با ترکیب دو مجموعه‌داده و ایجاد مجموعه‌داده‌ی ترکیبیMWD7، دامنه و تنوع نمونه‌های آموزشی افزایش یافت و مدل جدیدی بر پایه این مجموعه‌داده‌ی جدید ساخته شد. نتایج ارزیابی مدل MWD7-ENB0 نشان داد که این مدل در طبقه‌بندی ۱۸۱۳۰ تصویر مجموعه‌داده‌ی MWD7، در ۷ دسته‌ی متفاوت و پرکاربرد از تصاویر پسماند عملکرد بسیار خوبی داشته است. مدل آخر، توانایی طبقه‌بندی تصاویر پسماند در دسته‌های کارتن، شیشه، فلز، پسماند آلی، کاغذ، پلاستیک، پسماندهای نساجی و پوشاک را دارد. در گام بعدی تحقیق، یکی از معماری‌های سنگین‌تر خانواده‌ی EfficientNet برای بررسی تاثیر در کیفیت مدل مورد ارزیابی قرار گرفت. در این تحقیق با اندازه‌ی ورودی یکسان، استفاده از معماری سنگین‌تر خانواده‌ی EfficientNet بهبودی در معیارهای ارزیابی مدل ایجاد نکرد اما استفاده از اندازه‌ی ورودی استاندارد برای معماری با پارامترهای بیشتر، نشان داد که استفاده از معماری‌های سنگین‌تر با اندازه ورودی متناسب با هر مدل، می‌تواند در بهبود بهتر کارآیی مدل تاثیرگذار باشد. علاوه بر مدل MWD7-ENB0 که روی مجموعه‌داده‌ای با تصاویر بیشتر و متنوع‌تر توسعه داده شده، از مدل‌های GC11-ENB0 و RW8-ENB0 می‌توان برای طبقه‌بندی تصاویر در دسته‌های پسماند باتری، زیستی، شیشه‌ی قهوه‌ای، شیشه‌ی سبز، شیشه‌ی سفید، کفش، پوشاک، پسماندهای آلی غذایی، پسماند نساجی و پسماند گیاهی استفاده کرد. در فاز دوم تحقیق، پس از اجرای یک آزمایش و آشکارسازی محدودیت‌های معماری EfficientNet در تکرار کارآیی خود روی تصاویر با حضور یک نوع پسماند در تحلیل تصاویر با حضور هم‌زمان چند نوع پسماند، ابتدا مجموعه‌داده‌ی چندشیئیMOWD با ۴۰۵۱ تصویر با حضور هم‌زمان چند نوع پسماند ایجاد شد و سپس مدل مدرن YOLO26s روی این مجموعه‌داده آموزش یافت، نتایج نشان از کیفیت خوب مدل داشت، این مدل توانایی تشخیص نوع پسماندهای متعدد در تصویر و موقعیت قرارگیری هر پسماند را دارا می‌باشد. این رویکرد می‌تواند در سناریوهای واقعی، پرکاربرد باشد. مدل‌های خروجی تحقیق و همچنین پژوهش‌های تئوری مطرح شده، می‌توانند در سامانه‌های تفکیک خودکار پسماند، مدیریت منابع بازیافتی و در نتیجه‌ی آن، پروژه‌های عمرانی، حفاظت از محیط زیست و بهبود سلامت انسان و سایر جانداران تاثیر بسزایی داشته باشند و همچنین منجر به کاهش خطا و هزینه در فرایندهای مرتبط شوند.

۷-منابع

1. Gebrekidan, T. K., Weldemariam, N. G., Hidru, H. D., Gebremedhin, G. G., & Weldemariam, A. K. (2024). Impact of improper municipal solid waste management on fostering One Health approach in Ethiopia—challenges and opportunities: a systematic review. *Science in One Health*, *3*, 100081.
2. Haris, N. I. N., Al Edrus, S. S. O., Abdul Raof, N., Wondi, M. H., Khan, W. R., Sien, L. S., ... & Sawatdeenarunat, C. (2024). Toward low-carbon cities: A review of circular economy integration in urban waste management and its impact on carbon emissions. *Wiley Interdisciplinary Reviews: Energy and Environment*, *13*(5), e535.
3. Wu, Z., Pei, T., Bao, Z., Ng, S. T., Lu, G., & Chen, K., 2025. Utilizing intelligent technologies in construction and demolition waste management: From a systematic review to an implementation framework. Frontiers of Engineering Management, 12(1), 1-23. doi.org/10.1007/s42524-024-0144-4
4. Papamichael, I., Voukkali, I., Loizia, P., & Zorpas, A. A., 2023. Construction and demolition waste framework of circular economy: A mini review. *Waste Management & Research*, *41*(12), 1728-1740. doi.org/10.1177/0734242X231190804

5. Fang, B., Yu, J., Chen, Z., Osman, A. I., Farghali, M., Ihara, I., ... & Yap, P. S., 2023. Artificial intelligence for waste management in smart cities: a review. *Environmental Chemistry Letters*, *21*(4), 1959-1989. doi.org/10.1007/s10311-023-01604-3
6. Lu, W., & Chen, J., 2022. Computer vision for solid waste sorting: A critical review of academic research. *Waste Management*, *142*, 29-43. doi.org/10.1016/j.wasman.2022.02.009
7. Mansour, M. A., Ismail, M. H. B., Imran Latif, Q. B. A., Alshalif, A. F., Milad, A., & Bargi, W. A. A., 2023. A systematic review of the concrete durability incorporating recycled glass. *Sustainability*, *15*(4), 3568. doi.org/10.3390/su15043568
8. Kuoribo, E., Shokry, H., Asawa, T., & Mahmoud, H., 2024. A scientometric review of recycled glass waste as an alternative raw material in concrete production. *European Journal of Environmental and Civil Engineering*, *28*(11), 2507-2533. doi.org/10.1080/19648189.2024.2318402
9. Babafemi, A. J., Šavija, B., Paul, S. C., & Anggraini, V., 2018. Engineering properties of concrete with waste recycled plastic: A review. *Sustainability*, *10*(11), 3875. doi.org/10.3390/su10113875
10. Gu, L., & Ozbakkaloglu, T., 2016. Use of recycled plastics in concrete: A critical review. *Waste Management*, *51*, 19-42. doi.org/10.1016/j.wasman.2016.03.005
11. Tran, N. P., Gunasekara, C., Law, D. W., Houshyar, S., Setunge, S., & Cwirzen, A., 2022. Comprehensive review on sustainable fiber reinforced concrete incorporating recycled textile waste. *Journal of Sustainable Cement-Based Materials*, *11*(1), 28-42. doi.org/10.1080/21650373.2021.1875273
12. Sasi, S., Joseph, P., Haigh, R., Sandanayake, M., Vrcelj, Z., & Yaghoubi, E., 2024. A review on the effects of waste textile polymer fiber on concrete strength: Exploring the key parameters. *Buildings*, *14*(5), 1486. doi.org/10.3390/buildings14051486
13. Solahuddin, B. A., & Yahaya, F. M., 2021. Inclusion of waste paper on concrete properties: A review. *Civil Engineering*, *7*, 94-113. doi.org/10.28991/CEJ-SP2021-07-07
14. Solahuddin, B. A., & Yahaya, F. M., 2022. Properties of concrete containing shredded waste paper as an additive. Materials Today: Proceedings, 51, 1350-1354. doi.org/10.1016/j.matpr.2021.11.390
15. Mensah, R. A., Shanmugam, V., Narayanan, S., Razavi, N., Ulfberg, A., Blanksvärd, T., ... & Das, O., 2021. Biochar-added cementitious materials—A review on mechanical, thermal, and environmental properties. *Sustainability*, *13*(16), 9336. doi.org/10.3390/su13169336
16. Zhang, Y., He, M., Wang, L., Yan, J., Ma, B., Zhu, X., ... & Tsang, D. C., 2022. Biochar as construction materials for achieving carbon neutrality. *Biochar*, *4*(1), 59. doi.org/10.1007/s42773-022-00182-x
17. Kanyilmaz, A., Birhane, M., Fishwick, R., & Del Castillo, C., 2023. Reuse of steel in the construction industry: challenges and opportunities. *International Journal of Steel Structures*, *23*(5), 1399-1416. doi.org/10.1007/s13296-023-00778-4
18. Tan, M., & Le, Q., 2019. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning* (pp. 6105-6114). PMLR.

19. Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L., 2009, June. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition* (pp. 248-255). Ieee. doi.org/10.1109/CVPR.2009.5206848
20. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., ... & Fei-Fei, L., 2015. Imagenet large scale visual recognition challenge. *International journal of computer vision*, *115*(3), 211-252. doi.org/10.1007/s11263-015-0816-y
21. Plested, J., Phiri, M., & Gedeon, T., 2026. Deep transfer learning for image classification: a survey. *Artificial Intelligence Review*. doi.org/10.1007/s10462-026-11491-z
22. Kim, H. E., Cosa-Linan, A., Santhanam, N., Jannesari, M., Maros, M. E., & Ganslandt, T., 2022. Transfer learning for medical image classification: a literature review. *BMC medical imaging*, *22*(1), 69. doi.org/10.1186/s12880-022-00793-7
23. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 779-788).
24. Jocher, G., Qiu, J., Liu, M., Lyu, S., Akyon, F. C., & Kalfaoglu, M. E. (2026). Ultralytics yolo26: Unified real-time end-to-end vision models. *arXiv preprint arXiv:2606.03748*.
25. Alkılınç, A., Okay, F. Y., Kök, İ., & Özdemir, S., 2025. Deep Ensemble Learning Model for Waste Classification Systems. *Sustainability*, *18*(1), 24. doi.org/10.3390/su18010024.
26. Mohamed, M. (2021). Garbage Classification (12 Classes). Kaggle. Available at: https://www.kaggle.com/datasets/mostafaabla/garbage-classification
27. Single, S., Iranmanesh, S., & Raad, R., 2023. Realwaste: a novel real-life data set for landfill waste classification using deep learning. *Information*, *14*(12), 633. doi.org/10.3390/info14120633
28. National Institute of Standards and Technology., 2015. *Secure Hash Standard (SHS) (FIPS PUB 180-4)*. doi.org/10.6028/NIST.FIPS.180-4
29. McKeown, S., & Buchanan, W. J., 2023. Hamming distributions of popular perceptual hashing techniques. *Forensic Science International: Digital Investigation*, *44*, 301509. doi.org/10.1016/j.fsidi.2023.301509
30. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É., 2011. Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, *12*, 2825-2830.
31. Shorten, C., & Khoshgoftaar, T. M., 2019. A survey on image data augmentation for deep learning. *Journal of big data*, *6*(1), 1-48. doi.org/10.1186/s40537-019-0197-0
32. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, *15*(1), 1929-1958.
33. Li, C., Liu, K., & Liu, S., 2025. A survey of loss functions in deep learning. *Mathematics*, *13*(15), 2417. doi.org/10.3390/math13152417

34. Reyad, M., Sarhan, A. M., & Arafa, M., 2023. A modified Adam algorithm for deep neural network optimization. *Neural Computing and Applications*, *35*(23), 17095-17112. doi.org/10.1007/s00521-023-08568-z

35. Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

36. Hussein, B. M., & Shareef, S. M., 2024. An empirical study on the correlation between early stopping patience and epochs in deep learning. In *ITM Web of Conferences* (Vol. 64, p. 01003). EDP Sciences. doi.org/10.1051/itmconf/20246401003

37. Padilla, R., Passos, W. L., Dias, T. L., Netto, S. L., & Da Silva, E. A. (2021). A comparative analysis of object detection metrics with a companion open-source toolkit. *Electronics*, *10*(3), 279.

38. Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., ... & Zitnick, C. L. (2014, September). Microsoft coco: Common objects in context. In *European conference on computer vision* (pp. 740-755). Cham: Springer International Publishing.

39. Fotovvatikhah, F., Ahmedy, I., Noor, R. M., & Munir, M. U., 2025. A systematic review of AI-based techniques for automated waste classification. *Sensors*, *25*(10), 3181. doi.org/10.3390/s25103181

40. Zhang, Q., Yang, Q., Zhang, X., Bao, Q., Su, J., & Liu, X., 2021. Waste image classification based on transfer learning and convolutional neural network. *Waste Management*, *135*, 150-157. doi.org/10.1016/j.wasman.2021.08.038

41. Risfendra, R., Ananda, G. F., & Setyawan, H., 2024. Deep learning-based waste classification with transfer learning using EfficientNet-B0 model. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, *8*(4), 535-541. doi.org/10.29207/resti.v8i4.5875

42. Kumala, R. A., Sari, C. A., & Rachmawanto, E. H., 2025. A Comparison of MobileNetV2 and VGG16 Architectures with Transfer Learning for Multi-Class Image-Based Waste Classification. *Journal of Applied Informatics and Computing*, *9*(4), 1610-1624. doi.org/10.30871/jaic.v9i4.9958

۴۳. بیات زاده، مسعود و حسینی رحمان، سید ابوالقاسم و کربلایی حسنی، حمید و شیخی، محمد،۱۴۰۴، تشخیص زباله های خشک مبتنی بر سیستم هوشمند شبکه عصبی کانولوشن درسازمان مدیریت پسماند شهرداری اراک، https://civilica.com/doc/2333531

44. Ali, H., Shifa, N., Benlamri, R., Farooque, A. A., & Yaqub, R. (2025). A fine tuned EfficientNet-B0 convolutional neural network for accurate and efficient classification of apple leaf diseases. *Scientific Reports*, *15*(1), 25732.

۴۵. رنجبری، سمانه و ریگی، مصطفی،۱۴۰۲،تفکیک پسماند مبتنی بر پردازش تصویر در شهر هوشمند، شانزدهمین کنفرانس بین المللی نوآوری و تحقیق در علوم مهندسی، https://civilica.com/doc/1885403

۴۶. ملک الکلامی، میلا و حسن زاده، محمد،۱۴۰۱،مدیریت هوشمند پسماند بر پایه ی اینترنت اشیاء: گامی در مسیر شهر دانش، https://civilica.com/doc/1949898