مقاله پژوهشی

http://jmvip.sinaweb.net
www.ismvip.ir

این مقاله در قالب نهایی آن در مجله
ماشین‌بینایی و پردازش‌تصویر چاپ خواهد شد

تاریخ انتشار برخط: اسفندماه ۱۴۰۲

خوشه‌بندی عمیق تصاویر با رویکرد یادگیری گروهی و ویژگی‌های چندگانه شبکه‌های عصبی عمیق

ملیحه دانش[۱]، مصطفی نیکوسرشت[۲]

چکیده

یادگیری عمیق ابزار قدرتمندی در خوشه‌بندی تصاویر پیچیده و حجیم است. اغلب روش‌های خوشه‌بندی عمیق عمل دسته‌بندی تصاویر را بر اساس بردارهای بازنمایی حاصل از آموزش یک شبکه عصبی عمیق انجام می‌دهند به طوریکه ویژگی‌های مستخرج از لایه آخر شبکه مورد ارزیابی نهایی عمل خوشه‌بندی قرار می‌گیرد. بهره‌مندی از ویژگی‌های معنایی متفاوت حاصل از چندین شبکه عمیق می‌تواند نقش مؤثری در بهبود کارایی نتایج خوشه‌بندی حاصل ایفا کند که تاکنون تحقیقات کمی در این زمینه صورت گرفته است. در این مقاله رویکردی مبتنی بر خوشه‌بندی عمیق گروهی ارائه می‌کنیم که در آن با پیوند دادن روش‌های خوشه‌بندی عمیق و یادگیری گروهی، سعی در استفاده از مزایای شبکه‌های عصبی عمیق در کنار یکدیگر داریم. در این راستا ابتدا پنج مدل مختلف خودرمزگذار عمیق با لایه‌های کانولوشنی متعدد آموزش داده می‌شوند که در هر یک از آنها از رویکرد یادگیری انتقالی نیز جهت افزایش دقت و بهبود عملکرد استفاده می‌شود. پس از استخراج ویژگی‌های چندگانه تصاویر توسط مدل‌های مختلف عمیق، بازنمایی‌های حاصل خوشه‌بندی شده و طبق رویکرد یادگیری گروهی نتایج آنها با یکدیگر تجمیع می‌شود. نهایتاً دسته‌بندی نهایی تصاویر بر اساس اطلاعات مشترک خوشه‌بندی‌های پایه صورت می‌گیرد. نتایج حاصل از اعمال روش پیشنهادی بر روی چهار مجموعه داده استاندارد، نشان دهنده کارایی بهتر و مؤثرتر روش پیشنهادی نسبت به رویکردهای اخیر خوشه‌بندی عمیق تصاویر است.

کلید واژه ها

خوشه‌بندی تصاویر، شبکه عصبی عمیق، یادگیری گروهی، خودرمزگذار ، یادگیری انتقالی

۱- مقدمه

خوشه‌بندی تصاویر یک امر اساسی و در عین حال چالش برانگیز است که کاربردهای متنوعی در حوزه های مختلف علوم و پزشکی دارد. در گذشته دسته‌بندی اشکال و تصاویر طبق ویژگی‌های ساده و پیش پا افتاده آنها و به صورت دستی صورت می‌گرفت اما در زمان حال با افزایش روزافزودن حجم داده‌های تصویری و ویدیویی با پیچیدگی و ابعاد بالا، الگوریتم‌های سنتی خوشه‌بندی از دقت کافی برای دسته‌بندی این نوع داده‌ها برخوردار نبوده و نتیجه مطلوبی را در برندارد.

در سال‌های اخیر، یادگیری عمیق با قابلیت یادگیری بازنمایی ویژگی‌ها[۱]، که ابزار مؤثری در خوشه‌بندی داده‌های بسیار پیچیده است، توجه چشمگیری را به خود جلب کرده است. خودرمزگذار[۲] نوع خاصی از شبکه‌های عصبی مصنوعی است که نقش اساسی در

این مقاله در شهریورماه ۱۴۰۲ دریافت شد؛ در بهمن‌ماه بازنگری و پذیرفته گردید.

[۱و۲] گروه مهندسی کامپیوتر، دانشکده مهندسی برق و کامپیوتر، دانشگاه علم و فناوری مازندران

[۱] رایانامه: m.danesh@mazust.ac.ir

[۲] رایانامه: mostafa.nikosersht@gmail.com

نویسنده مسئول: ملیحه دانش

[۱] Feature representation learning

[۲] Autoencoder

یادگیری بدون ناظر[۱] شبکه‌های عمیق ایفا می‌کنند و هدف آن یادگیری بازنمایی داده‌هاست. استخراج ویژگی از مهم‌ترین بخش‌های حل یک مساله در یادگیری ماشین است که از آنها در راستای دسته‌بندی و تشخیص داده‌های مختلف استفاده می‌شود. مهمترین مزیت خودرمزگذارها انتخاب خودکار این ویژگی‌هاست، به طوریکه اطلاعات معنایی مهم و اساسی موجود در داده‌ها را می‌توان بدین وسیله استخراج کرد و به نوعی تاثیر داده‌های بلااستفاده و نویزی را کاهش داد.

بدین ترتیب خودرمزگذارها با انجام تبدیلاتی که بر روی داده‌ها انجام می‌دهند، بخش تاثیرگذار و مهم داده‌های ورودی را انتخاب می‌کنند که این امر به فشرده‌سازی و کاهش ابعاد ورودی‌ها می انجامد. لذا با توجه به محدودیت‌های زمانی و حافظه‌ای در عمل پردازش، این نوع شبکه‌های عصبی با استخراج ویژگی‌های مؤثر داده‌ها، هزینه‌های زمانی و حافظه‌ای پردازش را تا حد امکان کاهش می‌دهند، خصوصاً در مسائلی مانند طبقه‌بندی و خوشه‌بندی داده‌های تصویری و ویدیویی که عمدتاً در ابعاد و حجم بالا و دارای تعداد کلاس‌های زیادی هستند. به طور خلاصه مزایای استفاده از خودرمزگذارها در تصاویر شامل فشرده‌سازی و کاهش ابعاد داده‌ها، رفع نویز و وضوح بیشتر تصاویر مخدوش شده و همچنین بهینه کردن زمان و حافظه مصرفی جهت انجام عملیات پردازشی مختلف است. با توجه به مزایای این نوع شبکه عصبی، در مسئله خوشه‌بندی تصاویر از این معماری عمیق بهره برده و بردارهای بازنمایی داده‌ها را بر این اساس ایجاد می‌کنیم.

به دنبال قابلیت یادگیری ویژگی‌ها توسط شبکه‌های عصبی عمیق[۲](DNN)، روش‌های خوشه‌بندی زیادی بر اساس این شبکه‌ها توسعه یافته‌اند که به آنها خوشه‌بندی عمیق[۳] گفته می‌شود [۱، ۲]. این روش‌های خوشه‌بندی عمدتاً به دو دسته روش‌های "یک مرحله‌ای" و "دو مرحله‌ای" تقسیم‌بندی می‌شوند [۳]. روش‌های خوشه‌بندی عمیق یک مرحله‌ای، عملیات یادگیری بازنمایی ویژگی‌ها و تخصیص خوشه‌ها را به طور همزمان و در یک چارچوب مشترک انجام می‌دهند. در مقابل روش‌های خوشه‌بندی عمیق دو مرحله‌ای این عملیات را به طور مستقل و طی دو فرآیند جداگانه انجام می‌دهند، بدین صورت که ابتدا عمل یادگیری بازنمایی ویژگی‌ها انجام می‌شود؛ سپس تخصیص خوشه‌ها بر اساس بردارهای بازنمایی حاصل صورت می‌گیرد.

اگرچه روش‌های خوشه‌بندی عمیق به پیشرفت‌های قابل توجهی در یادگیری بدون ناظر بازنمایی‌ها و دسته‌بندی آنها دست یافتند، اما عمده آنها از بردارهای بازنمایی حاصل از آموزش یک شبکه عصبی (معمولا آخرین لایه شبکه) برای عمل خوشه‌بندی استفاده می‌کنند، در حالیکه چنانچه بتوان از بازنمایی‌های متعددی از داده‌ها به طور همزمان استفاده کرد، امکان بهره‌مندی از اطلاعات معنایی جامع‌تری فراهم شده و اشتراک ویژگی‌های مستخرج از آنها به دسته‌بندی مؤثرتری از داده‌ها منجر خواهد شد.

یادگیری گروهی[۴] روشی در یادگیری ماشین است که هدف آن استفاده از الگوریتم‌های یادگیری متعدد در جهت بهبود عملکرد هر یک از الگوریتم‌ها به تنهایی است [۴]. به عبارتی یادگیری گروهی سعی دارد با بهره‌مندی از خرد جمعی و ترکیب خروجی چندین مدل یادگیری، در راستای کاهش نویز و بهبود نتایج پیش‌بینی گام بردارد.

با توجه به عملکرد موفق روش یادگیری گروهی در مسائل مختلف یادگیری ماشین، استفاده از آن در زمینه‌های متعدد یادگیری با ناظر و بدون ناظر مورد توجه زیادی قرار گرفته است. در مسائل یادگیری بدون ناظر و به طور خاص در عمل خوشه‌بندی، معمولاً هیچ دانش قبلی در مورد ساختار زیربنایی یا ویژگی‌های خاص داده‌ها وجود ندارد. از این رو الگوریتم‌های خوشه‌بندی مختلف ممکن است با ارائه رویکردهای متفاوتی به داده‌ها، ساختار خاصی از آنها را شناسایی کرده و نتایج خوشه‌بندی متفاوتی را برای داده‌های مشابه ایجاد کنند. علاوه بر این، هیچ الگوریتم خوشه‌بندی وجود ندارد که بتواند به طور مداوم برای مسائل مختلف به خوبی عمل کند و دستورالعمل مشخصی برای انتخاب الگوریتم‌های خوشه‌بندی مناسب برای یک مسئله معین وجود ندارد.

در راستای حل این چالش‌ها و با توجه به مزایای یادگیری گروهی، در این مقاله از رویکرد فوق جهت دستیابی به خوشه‌بندی کیفی‌تر از تصاویر استفاده می‌شود. اخیراً به شیوه‌های مختلفی از بحث یادگیری گروهی در خوشه‌بندی داده‌های تصویری استفاده شده است. در بسیاری از تحقیقات، دسته‌بندی اولیه تصاویر بدون استفاده از یادگیری عمیق و صرفاً با انواعی از روش‌های خوشه‌بندی سنتی صورت گرفته و سپس نتایج حاصل از آنها با یکدیگر ترکیب می‌شوند. در تحقیقات دیگری که مبتنی بر یادگیری عمیق هستند، به شیوه‌های مختلفی از شبکه‌های عمیق در روند خوشه‌بندی‌های پایه و اجماع نتایج حاصل استفاده شده است. در برخی از آنها ابتدا خوشه‌بندی بر روی بردارهای بازنمایی حاصل از شبکه‌های عصبی عمیق انجام می‌شود و سپس عمل اجماع نتایج صورت می‌گیرد و در بعضی دیگر، بردارهای بازنمایی حاصل از شبکه‌ها به روشی با یکدیگر ترکیب شده و سپس بردارهای حاصل از اجماع خوشه‌بندی می‌شوند؛ ضمن اینکه نوع شبکه‌های عصبی مورد استفاده و عملیات پیش پردازشی روی آنها نیز در خروجی نتایج تاثیرگذار هستند و از این حیث نیز تحقیقات متفاوتی در حال انجام است. روشی که در این مقاله پیشنهاد شده است به نوعی از ترکیبی از روش‌های فوق بهره می‌برد، بدین صورت که چندین مدل شبکه عصبی عمیق به طور همزمان آموزش داده

---

[۱] Unsupervised learning

[۲] Deep Neural Networks

[۳] Deep Clustering

[۴] Ensemble learning

می‌شوند تا بتوان اطلاعات مهم و مؤثر تصاویر را استخراج کرد. سپس با خوشه‌بندی بازنمایی‌های حاصل از هر شبکه و اجماع نتایج حاصل به صورت ماتریسی از بردارهای ویژگی، اطلاعات معنایی مستخرج از مدل‌ها با یکدیگر ترکیب شده و خوشه بندی می‌شوند تا بتوان به نتایج مطلوبی در امر خوشه‌بندی تصاویر دست یافت.

همچنین از آنجا که در اختیار داشتن اطلاعات همسایگی هر تصویر نقش اساسی در دسته‌بندی بهینه آنها دارد، از شبکه‌های عصبی کانولوشن[۱] (CNN) به عنوان یکی از مدل‌های موفق یادگیری بازنمایی تصاویر [۵] بدین منظور استفاده شده است. با بهره‌برداری از این شبکه‌ها ضمن آموزش اطلاعات سطح پیکسل، اطلاعات همسایه هر تصویر با استفاده از لایه‌های کانولوشن تا حد امکان استخراج می‌شود.

به علاوه، جهت افزایش نرخ یادگیری و بهبود کیفیت دسته‌بندی داده‌ها، تکنیک یادگیری انتقالی[۲] نیز مورد استفاده قرار گرفته است که این روش در داده‌های با مقیاس بالا منجر به سهولت یادگیری و افزایش دقت یادگیری می‌شود [۳]. در نهایت انتظار می‌رود که با استفاده از ویژگی‌های چندگانه شبکه‌های عصبی کانولوشنی در کنار یادگیری انتقالی، بتوان ضمن کاهش زمان آموزش مدل‌ها دقت خوشه‌بندی را تا حد امکان بهبود داد.

ادامه مقاله بدین صورت سازماندهی شده است: در بخش دوم پیشینه تحقیق مورد بررسی قرار گرفته و کارهای اخیر حوزه خوشه‌بندی عمیق و خوشه‌بندی گروهی مطرح می‌شود. در بخش سوم رویکرد پیشنهادی مقاله و مراحل مختلف آن توضیح داده می‌شود. نتایج حاصل از آزمایشات در بخش چهارم گزارش شده و نهایتاً خلاصه‌ای از تحقیق پیش رو در بخش پنجم بیان می‌شود.

۲- پیشینه تحقیق

در این مقاله رویکردی جهت خوشه‌بندی عمیق تصاویر ارائه می‌شود که در آن از ترکیبی از روش‌های خوشه‌بندی عمیق و خوشه‌بندی مبتنی بر یادگیری گروهی (خوشه‌بندی گروهی) استفاده می‌شود. کارهای مرتبط با هر یک از این روش‌های خوشه‌بندی به ترتیب در بخش‌های ۲-۱ و ۲-۲ مورد بررسی قرار می‌گیرند.

۲-۱- خوشه‌بندی عمیق

خوشه‌بندی عمیق به فرآیند یادگیری خودکار بازنمایی داده‌ها و تخصیص خوشه‌ها از طریق شبکه‌های عصبی اشاره دارد. الگوریتم‌های خوشه‌بندی سنتی همانند K-means، مدل آمیخته گاوسی[۳] (GMM) [۶]، فاکتورگیری ماتریس غیرمنفی (NMF)[۴] [۷] و خوشه‌بندی طیفی[۵] (SC) [۸] کاملاً به نحوه نمایش داده‌های ورودی وابسته هستند و از این رو برای داده‌های تصویری و متون کارایی خوبی ندارند. همچنین این الگوریتم‌ها اغلب برای داده‌هایی با ابعاد کم طراحی شده‌اند، لذا برای داده‌های حجیم و با ابعاد بالایی همچون تصاویر و ویدیوها عملکرد مناسبی ندارند.

روش خوشه‌بندی عمیق با قدرت یادگیری بازنمایی ویژگی‌ها توسط شبکه‌های عصبی عمیق این امکان را فراهم می‌سازد که داده‌های پیچیده و در مقیاس بزرگ را به بردارهای ویژگی با ابعاد کم تبدیل کرده و سپس از این ویژگی‌ها جهت خوشه‌بندی نهایی استفاده نماییم. در واقع، DNNها با اعمال تبدیل‌های غیرخطی پی در پی بر روی داده‌ها قادر هستند ویژگی‌های سطح پایین تا بالای داده‌ها را تا حد امکان استخراج نمایند. پیشرفت‌هایی که در زمینه یادگیری بازنمایی توسط DNNها صورت گرفته، به توسعه رویکردهای خوشه‌بندی عمیق متعددی انجامیده است که هدف آنها یادگیری مشترک بازنمایی نقاط داده‌ای و همچنین تخصیص خوشه به آنهاست.

در معماری‌های عمیق، یک شبکه عصبی ساده و در عین حال مؤثر، خودرمزگذار (AE[۶]) [۹] است که بسیاری از روش‌ها از AE جهت استخراج ویژگی‌ها برای عمل خوشه‌بندی استفاده می‌کنند. به عبارتی، خودرمزگذارها نمونه خاصی از DNNها هستند که برای تعبیه‌سازی داده‌ها در یک فضای برداری (معمولا متراکم و با ابعاد کم) آموزش داده می‌شوند و سعی می‌کنند ورودی‌ها را در این فضای برداری بازسازی کنند. جذابیت AEها در این واقعیت نهفته است که آنها می‌توانند بازنمایی‌ها را به روشی کاملاً بدون ناظر بیاموزند. در این راستا روش‌های خوشه‌بندی عمیق متعددی بر مبنای استفاده از AEها پیشنهاد شدند که در ادامه به ذکر مختصری از کارهای اخیر این حوزه می‌پردازیم.

یانگ و همکارانش [۱۰] از AE جهت تعبیه‌سازی داده‌ها بهره بردند و سپس با الگوریتم K-means خوشه‌بندی بردارهای تعبیه حاصل را انجام دادند. مرادی فرد و همکارانش [۱۱] یک رویکرد K-means عمیق پیشنهاد کردند که طی آن عملیات خوشه‌بندی و یادگیری بازنمایی‌ها را در یک فرآیند همزمان انجام می‌دهند. بدین منظور آنها از معماری خودرمزگذار عمیق جهت یادگیری بردارهای بازنمایی و ارائه آنها به الگوریتم خوشه‌بندی K-means (با تابع هدف اصلاح شده) استفاده کردند. اِل وی و همکارانش [۱۲] یک روش خوشه‌بندی زیرفضای عمیق[۷] مبتنی بر AE ارائه کردند. آنها تابع خطای بازسازی[۸] خودرمزگذار را با استفاده از یک معیار شباهت زوجی بین داده‌ها وزن‌دهی کردند تا بتوانند اطلاعات ساختاری محلی بین داده‌ها را تا حد امکان استخراج کنند. سپس بردارهای ویژگی حاصل را با استفاده از الگوریتم K نزدیک‌ترین

---

[۱] Convolutional Neural Networks
[۲] Transfer learning
[۳] Gaussian Mixture Model
[۴] Nonnegative Matrix Factorization
[۵] Spectral Clustering
[۶] AutoEncoder
[۷] Deep Subspace Clustering
[۸] Reconstruction loss

همسایه[۱] (KNN) خوشه‌بندی نمودند. دیالو و همکارانش [۱۳] از تکنیک‌های خودرمزگذار انقباضی[۲] و خودافزایشی[۳] برای توسعه یک مدل خوشه‌بندی عمیق استفاده کردند تا بتوانند ضمن کم کردن فاصله کردن نقاط همسایه به یکدیگر در خوشه‌ها توسط خودرمزگذار انقباضی، بازنمایی‌ها را هر چه بیشتر به موارد افزایشی آنها نزدیک‌تر کنند. در ادامه از خوشه‌بندی طیفی جهت گروه‌بندی بردارهای بازنمایی حاصل استفاده کردند. همچنین چانگ در [۱۴] مدل خودرمزگذار ترکیبی[۴] را برای خوشه‌بندی عمیق تصاویر ارائه کرد. در این مدل از ترکیبی از شبکه‌های خودرمزگذار متغیر[۵] (VAE) و مولد تخاصمی[۶] (GAN) استفاده کردند تا بتوانند بازنمایی‌های دقیق‌تر و در نتیجه دسته‌بندی بهتری از تصاویر انجام دهند. با توجه به عملکرد موفق خودرمزگذارها در شناسایی بردارهای بازنمایی داده‌های تصویری پیچیده و با ابعاد بالا، در این مقاله قصد داریم از این معماری در راستای دستیابی به ویژگی‌های سطح بالای تصاویر و در نهایت خوشه‌بندی مؤثرتری از آنها بهره ببریم.

شبکه‌های CNN کارایی بسیار خوبی در انواع عملیات پردازش تصویر از خود نشان داده‌اند، به این دلیل که ساختار لایه‌های کانولوشنی اطلاعات میان پیسکل‌های متصل به هم در تصاویر را به خوبی حفظ می‌کند و عملکرد خوبی در پردازش داده‌های تصویری دارد. از این رو این نوع شبکه‌ها در خوشه‌بندی عمیق تصاویر نیز مورد توجه زیادی قرار گرفتند. برای نمونه، یانگ و همکارانش [۱۵] ابتدا با استفاده از یک شبکه CNN عمل یادگیری بردارهای بازنمایی‌ها را انجام داده و سپس از یک شبکه RNN جهت پیاده‌سازی خوشه‌بندی تجمعی بر روی بازنمایی‌هایی حاصل از CNN بهره بردند. درمدل آنها، عمل آموزش هر دو شبکه طی یک فرآیند سراسری صورت گرفت. کارُن و همکارانش [۱۶] طی یک فرآیند همزمان پارامترهای شبکه CNN و تخصیص خوشه به ویژگی‌های حاصل را مشترکاً انجام می‌دهند. بدین صورت که به طور متناوب ویژگی‌های شبکه CNN را با الگوریتم خوشه‌بندی K-means گروه‌بندی کردند و سپس از خوشه‌های تخصیص یافته به روش نظارت شده، جهت به روزرسانی وزن‌های شبکه استفاده نمودند. همچنین سان و همکارانش [۱۷] یک روش نیمه نظارت شده به نام خوشه‌بندی عمیق فعال پیشنهاد کردند. در این روش آنها مدل‌های یادگیری فعال و خوشه‌بندی عمیق را با یکدیگر ترکیب کردند به طوریکه بتوانند با وارد کردن تفسیرهای انسانی در برچسب‌گذاری داده‌های غیرقطعی و نامشخص، از این اطلاعات جهت آموزش بهتر شبکه CNN و دستیابی به خوشه‌بندی کیفی‌تری از تصاویر استفاده نمایند. پنگ و لی [۱۸] نیز با هدف حفظ ویژگی‌های درون کلاسی، از یک مدل خوشه‌بندی عمیق تصاویر بر اساس تشابه برچسب داده‌ها و به حداکثر رساندن اطلاعات متقابل بین آنها پیشنهاد کردند. در این مدل آنها یک شبکه ترکیبی حاوی دو بلوک سطح کلاس و سطح نمونه ارائه کردند تا بتوانند بازنمایی داده‌ها را در هر دو سطح، با کمک یک شبکه عصبی عمیق مشترک به دست آورند. با خوشه‌بندی مستقل داده‌ها در هر یک از این سطوح، آنها توانستند میزان شباهت درون خوشه‌ای و تمایز بین خوشه‌ای را به ترتیب با استفاده از بلوک سطح نمونه و بلوک سطح کلاس به دست آورند.

با توجه به عملکرد مؤثر شبکه‌های CNN در یادگیری بازنمایی داده‌های تصویری، در این مقاله نیز از این نوع شبکه‌ها بهره برده و از لایه‌های کانولوشنی جهت افزایش دقت ویژگی‌های حاصل استفاده می‌شود.

۲-۲- خوشه‌بندی گروهی

هدف از یادگیری گروهی ترکیب نتایج چندین مدل پایه با یکدیگر در راستای دستیابی به عملکرد بهتر و نرخ یادگیری بیشتر نسبت به هر یک از الگوریتم‌های پایه است. این تکنیک به عنوان راهکاری در افزایش دقت خوشه‌بندی نیز مورد استفاده قرار گرفت و روش‌های متعددی سعی کردند با ادغام الگوریتم‌های خوشه‌بندی متفاوت یا یکدیگر (خوشه‌بندی گروهی)، دسته‌بندی پایدارتر و دقیق‌تری از داده‌ها به دست آورند.

روش‌های خوشه‌بندی گروهی عمدتاً به سه دسته تقسیم می‌شوند [۱۹]، شامل روش‌های مبتنی بر هم‌رویدادی زوج نقاط[۷]، روش‌های مبتنی بر افراز میانی[۸] و روش‌های مبتنی بر افرازبندی گراف[۹]. روش‌های مبتنی بر هم‌رویدادی [۲۰، ۲۱]، ابتدا یک ماتریس هم‌رویدادی بر اساس فرکانس قرارگیری جفت نقاط در گروه‌های یکسانی از خوشه‌بندی‌های پایه ایجاد می‌شود. سپس الگوریتم‌های خوشه‌بندی مبتنی بر ماتریس شباهت همچون سلسله‌مراتبی تجمعی بر روی ماتریس حاصل اعمال می‌شود تا دسته‌بندی داده‌ها طبق روش یادگیری گروهی ایجاد شود.

روش‌های مبتنی بر افراز میانی [۲۲، ۲۳] مسئله خوشه‌بندی گروهی را به یک مسئله بهینه‌سازی تبدیل می‌کنند که به دنبال یافتن یک خوشه‌بندی (یا افراز) میانی با به حداکثر رساندن شباهت بین این خوشه‌بندی و خوشه‌بندی پایه هستند. فرانک و جیانگ [۲۲] با تعبیه‌سازی خوشه‌ها در فضای برداری، مسئله خوشه‌بندی گروهی را به یک مسئله میانه اقلیدسی تبدیل کردند. هوانگ و همکارانش

[۱] K-Nearest Neighbors
[۲] Contractive Autoencoder
[۳] Self-augmentation
[۴] Fusion Autoencoder
[۵] Variational AutoEncoder
[۶] Generative Adversarial Network
[۷] Pairwise co-occurrence
[۸] Median partition
[۹] Graph partitioning

[۲۳] مسئله افراز میانی را به یک مسئله برنامه‌ریزی خطی دودویی فرموله کردند و آن را با مدل گراف عامل[۱] حل کردند.

روش‌های مبتنی بر افرازبندی گراف [۲۴، ۲۵] نیز معمولاً یک ساختار گراف را با جمع‌آوری اطلاعات خوشه‌بندی‌های پایه ایجاد می‌کنند و سپس خوشه‌بندی گروهی را از طریق برخی الگوریتم‌های افرازبندی گراف به دست می‌آورند. برای نمونه، در [۲۴] از مفهوم ابر لَبه[۲] استفاده شد و سه الگوریتم خوشه‌بندی مختلف مبتنی بر تقسیم‌بندی گراف را پیشنهاد کردند. نویسندگان در [۲۵] با استفاده از خوشه‌بندی‌های پایه و نقاط داده‌ای به عنوان گره‌های گراف، یک گراف دوبخشی ساختند و سپس گراف حاصل را با یکی از الگوریتم‌های خوشه‌بندی گراف افراز کردند.

اگرچه این روش‌های خوشه‌بندی گروهی عملکرد خوبی در بهبود دقت و پایداری نتایج حاصل از خود نشان دادند، اما عمدتاً وابسته به ویژگی‌هایی هستند که به طور دستی استخراج می‌شوند و عموماً برای داده‌هایی با ساختار پیچیده همانند تصاویر و ویدیو عملکرد مناسبی ندارند. در سال‌های اخیر یادگیری عمیق به عنوان ابزار قدرتمندی در یادگیری بازنمایی ویژگی‌های داده‌های بسیار پیچیده ارئه شده است، با این حال بهره‌مندی از ویژگی‌های چندگانه حاصل از شبکه‌های عصبی عمیق و تقویت اطلاعات معنایی تصاویر با کمک یادگیری گروهی، موضوعی است که جدیداً مورد توجه قرار گرفته و تاکنون تحقیقات کمی روی آن صورت گرفته است.

در این زمینه تحقیقی توسط هانگ و همکارانش در [۱۹] انجام شد که آنها یک روش خوشه‌بندی تصاویر مبتنی بر یادگیری گروهی را با هدف استخراج اطلاعات معنایی موجود در لایه‌های مختلف شبکه عصبی عمیق پیشنهاد کردند. بدین صورت که ابتدا یک شبکه CNN را به دو شکل مبتنی بر نمونه و مبتنی بر خوشه با پارامترهای وزن مشترک و به صورت بدون ناظر آموزش دادند. سپس بردارهای ویژگی حاصل از لایه‌های متعدد شبکه عصبی را با یکدیگر ترکیب کرده و خوشه‌بندی‌های پایه متعددی را از روی بردارهای بازنمایی حاصل ایجاد نمودند. در نهایت با تشکیل یک گراف دوبخشی وزن‌دار از تجمیع نتایج خوشه‌بندی‌های پایه و اعمال برش انتقال روی آن، عمل دسته‌بندی تصاویر را انجام دادند.

۳- روش پیشنهادی

در این قسمت روش پیشنهادی مقاله تحت عنوان EDIC[۳] مطرح می‌شود که در آن از مزایای شبکه‌های عصبی عمیق در یادگیری بازنمایی داده‌های تصویری استفاده می‌شود. بدین منظور خودرمزگذارهای کانولوشنی به عنوان معماری شبکه عمیق و رویکرد یادگیری گروهی در یک چارچوب یکپارچه به کار برده می‌شوند. فلوچارت روش پیشنهادی در شکل (۱) نشان داده شده است.

همانطور که در شکل فوق پیداست، روند کلی الگوریتم EDIC بدین صورت است که ابتدا تصاویر ورودی در قالب ماتریسی از پیکسل‌ها وارد شبکه شده و پس از نرمال‌سازی و تنظیم ابعاد، وارد شبکه‌های خودرمزگذار عمیق مبتنی بر شبکه‌های عصبی کانولوشنی می‌شوند. در این مرحله، داده‌های ورودی با استفاده از ساختار رمزگذار هر شبکه عمیق آموزش دیده و به مدل‌های برداری با ابعاد کم تعبیه می‌شوند. سپس بردارهای تعبیه حاصل وارد ساختار رمزگشای شبکه شده و در آنجا مجددا بازسازی می‌شوند. نهایتاً بردارهای تعبیه حاصل از آموزش شبکه خودرمزگذار با حداقل خطای بازسازی برگردانده می‌شوند. به عبارتی خروجی هر مدل، بازنمایی‌هایی از تصاویر ورودی هستند که مستقلاً عمل خوشه‌بندی بر روی آنها انجام می‌شود. سپس نتایج خوشه‌بندی‌های حاصل از شبکه‌های مختلف طبق رویکرد یادگیری گروهی با یکدیگر تجمیع شده و ارزیابی نهایی خوشه‌بندی روی آنها انجام می‌شود.

درادامه مراحل مختلف روش پیشنهادی به تفصیل توضیح داده می‌شوند.

۳-۱- پیش پردازش تصاویر ورودی

جهت خوشه‌بندی هر یک از مجموعه تصاویر ورودی، ابتدا پیش پردازشی بر روی این داده‌ها اعمال می‌شود تا با سهولت و دقت بیشتری عمل آموزش شبکه عصبی عمیق و یادگیری بردارهای بازنمایی انجام شود. بدین منظور رویکرد یادگیری انتقالی به کار می‌رود تا بتوان از دانش شبکه‌های از پیش آموزش دیده در جهت کارایی بهتر مدل پیشنهادی بهره برد.

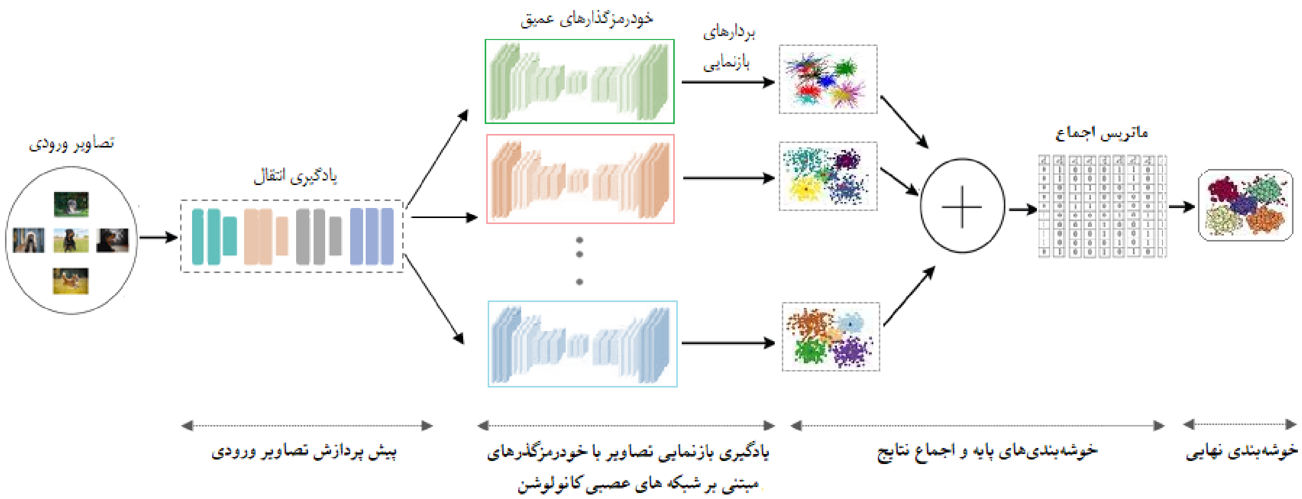
از آنجا که مجموعه تصاویر ورودی حجم بالایی دارند، این امر منجر به سرعت پایین یادگیری و کاهش دقت مدل‌ها می‌شود. لذا رویکرد یادگیری انتقالی سعی دارد با در اختیار قرار دادن مجموعه‌ای از ویژگی‌های پایه و دانش اولیه تصاویر، به ما در افزایش سرعت یادگیری و بهبود دقت نتایج یاری رساند. در این راستا از مدل‌های از پیش آموزش داده شده vgg16 و ResNet-50 استفاده می‌کنیم و ابعاد تصاویر ورودی را بر اساس ورودی مدل‌های فوق تنظیم کرده و آماده پردازش و ورود به شبکه می‌نماییم.

---

۱ Factor graph

۲ Hyper-edge

۳ Ensemble based Deep Image Clustering



شکل (۱) : روند کلی روش پیشنهادی EDIC

۳-۲- یادگیری بازنمایی تصاویر با خودرمزگذارهای مبتنی بر شبکه‌های عصبی کانولوشن

یک خودرمزگذار کانولوشنی[۱] (CAE) یک نوع شبکه عصبی بدون ناظر است که با استفاده از لایه‌های کانولوشنی در رمزگذار و رمزگشا، تصویر ورودی را در لایه خروجی بازسازی می‌کند. مزیت شبکه عصبی کانولوشن در این است که می‌تواند ویژگی‌های ساختاری تصویر و اطلاعات میان پیکسل‌های مجاور را تا حد امکان استخراج کرده و عملکرد خوبی در پردازش داده‌های تصویری داشته باشد. لذا ما از این نوع شبکه در معماری خودرمزگذار عمیق پیشنهادی استفاده کرده و لایه‌های کانولوشنی را در هر یک از مراحل آموزش و بازسازی تصویر به کار می‌بریم. روند کلی CAE مورد استفاده در شکل (۲) نشان داده شده است.

همانطور که در شکل فوق پیداست، ابتدا مجموعه تصاویر ورودی پس از پیش پردازش با مدل‌های از پیش آموزش دیده و استخراج دانش اولیه آنها، وارد شبکه رمزگذار می‌شوند. سپس طی لایه‌های مختلف کانولوشن در شبکه عصبی رمزگذار، بازنمایی‌هایی از آنها با ابعاد کم در فضای تعبیه ایجاد می‌شود؛ انتظار می‌رود با فشرده‌سازی تصاویر و استخراج ویژگی‌های مهم آنها، بخش‌های نویزی و پرت تصاویر در لایه تعبیه تا حد امکان حذف شده و بی‌اثر شوند. به دنبال آن رمزگشای کانولوشنی تصاویر فشرده شده را دریافت کرده و سعی می‌کند با استفاده از لایه‌های مختلف، نمونه‌هایی از تصاویر ورودی را بر اساس شکل فشرده شده آنها بازسازی کند.

پس از آموزش شبکه عصبی خودرمزگذار، بردارهای بازنمایی موجود در لایه تعبیه جهت عمل خوشه‌بندی اولیه تصاویر مورد استفاده قرار می‌گیرند.

تابع خطایی که جهت بهینه‌سازی کارایی شبکه CAE استفاده می‌شود، بررسی می‌کند که تا چه اندازه تصویر بازسازی شده طی فرآیند آموزش خودرمزگذار عمیق به تصویر اولیه شباهت دارد. در این راستا از تابع اختلاف آنتروپی دودویی[۲] (BCE) مطابق رابطه (۱) استفاده شده است که این تابع احتمال پیکسل های تصویر بازسازی شده و ورودی را با یکدیگر مقایسه می‌کند. بدین صورت که بر اساس میزان نزدیک یا دور بودن مقدار مورد انتظار از مقدار واقعی امتیازی را در بازه صفر و یک محاسبه کرده و برمی گرداند.

$$BCE = -(f_i . \log \hat{f}_i + (1-f_i) . \log(1-\hat{f}_i)) \qquad (۱)$$

در این رابطه $\hat{f}_i$ بیانگر مقدار عددی i اُمین خروجی مدل و f_i متناظر با مقدار هدف مربوطه است. هر چقدر تصویر بازسازی شده شباهت بیشتری به تصویر واقعی داشته باشد، مقدار این تابع خطا کمتر بوده (نزدیک‌تر به صفر) و بیانگر دقت بیشتر شبکه است. در مقابل به ازای تفاوت بیشتر مقدار بازسازی شده از واقعی، خطای BCE به یک نزدیک‌تر بوده و بیانگر عملکرد ضعیف شبکه است.

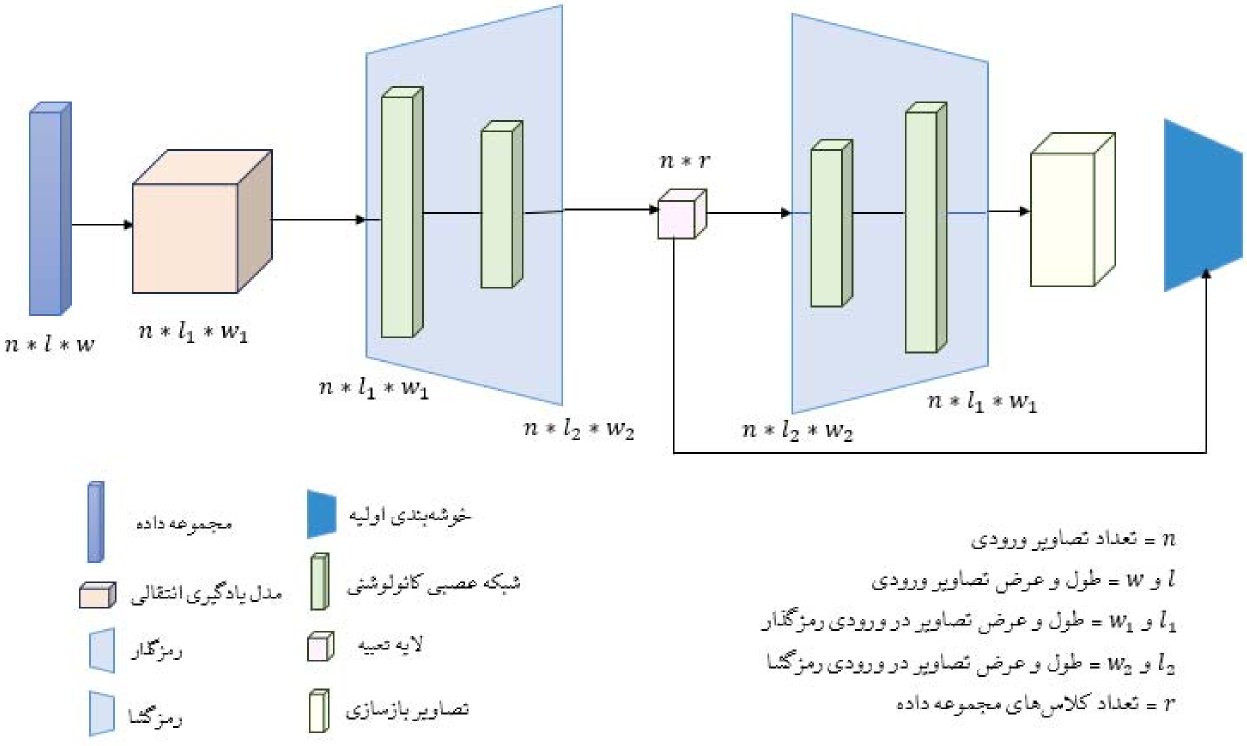
۳-۳- خوشه‌بندی‌های پایه و اجماع نتایج

روش خوشه‌بندی گروهی با الهام از رویکرد یادگیری گروهی و ترکیب چندین مدل مختلف خوشه‌بندی سعی دارد به نتیجه بهتری نسبت به هر یک از الگوریتم‌های خوشه‌بندی اولیه از نظر کارایی و ثبات دست یابد. این روش با تجمیع خوشه‌بندی‌های پایه، ساختار مشترک بین خوشه‌هایی را که به نمونه داده‌های مشابه تخصیص یافته‌اند، می‌یابد و بدین وسیله خوشه‌بندی پایدارتر و کیفی‌تری را نسبت به الگوریتم‌های پایه به دست می‌آورد.

---

[۱] Convolutional AutoEncoder

[۲] Binary Cross-Entropy loss



شکل (۲) : معماری خودرمزگذار کانولوشنی

با توجه به عملکرد موفق روش خوشه‌بندی گروهی در مسائل مختلف دنیای واقعی، ما نیز در این مقاله از رویکرد فوق استفاده کردیم و از اجماع خوشه‌بندی‌های مختلف در جهت بهبود نتایج حاصل و افزایش نرخ یادگیری بهره بردیم. در ادامه به توضیح روند استفاده از این رویکرد در مدل پیشنهادی خود می‌پردازیم.

جهت افزایش دقت و بهبود کیفیت نتایج با استفاده از یادگیری گروهی، روند کار بدین صورت بوده است که ابتدا پنج شبکه مختلف CAE با مجموعه‌ای از لایه‌های کانولوشنی مختلف آموزش داده می‌شوند که بتوان ویژگی‌های ساختاری موجود در داده‌ها را تا حد امکان استخراج کرد. جزئیات لایه‌های هر شبکه در جدول (۱) نشان داده شده است.

در ادامه هر شبکه CAE مستقلاً آموزش داده می‌شود و بردارهای بازنمایی حاصل از هر یک استخراج می‌شوند. جهت خوشه‌بندی بازنمایی‌های حاصل از هر شبکه، الگوریتم K-means استفاده می‌شود که یک الگوریتم خوشه‌بندی بر پایه تکرار است و سعی می‌کند مجموعه داده‌ها را به زیر گروه‌هایی بدون همپوشانی و کاملاً متمایز تقسیم کند. در الگوریتم K-means هر خوشه با یک مرکز ارتباط دارد و هدف اصلی آن به حداقل رساندن مجموع فواصل بین نقاط داده و مراکز خوشه‌های مربوطه است. بدین ترتیب که طی یک فرایند تکرارشونده بهترین مقدار را برای نقاط مراکز خوشه تعیین می‌کند. سپس هر نقطه داده را به نزدیک‌ترین مرکز c اختصاص می‌دهد به طوریکه آن نقاط داده که نزدیک به مرکز c هستند، تشکیل یک خوشه می‌دهند. هرچه تنوع کمتری در خوشه‌ها باشد، نقاط یک خوشه تشابه بیشتری با یکدیگر داشته و همگن‌تر هستند.

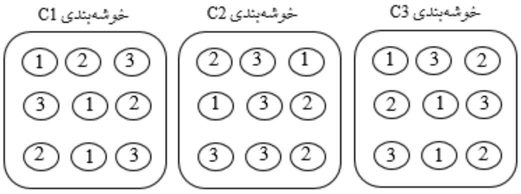
جدول (۱) : جزییات لایه‌های شبکه‌های عصبی مدل EDIC

شبکه عصبی	لایه‌های شبکه
CAE1	۲۰ لایه کانولوشن+ ۲ لایه BatchNormalization + ۲ لایه Dropout
CAE2	۱۸ لایه کانولوشن+ ۲ لایه BatchNormalization + ۲ لایه Dropout
CAE3	۱۸ لایه کانولوشن+ ۱ لایه BatchNormalization + ۳ لایه Dropout
CAE4	۲۰ لایه کانولوشن+ ۱ لایه BatchNormalization + ۲ لایه Dropout
CAE5	۱۶ لایه کانولوشن+ ۱ لایه BatchNormalization + ۲ لایه Dropout

پس از آن که خروجی حاصل از شبکه‌های عصبی عمیق بصورت مجزا خوشه‌بندی شد، به هر یک از نمونه‌ها (تصاویر) یک برچسب (شماره خوشه) نسبت داده می‌شود که ما این مقادیر را در لیست‌هایی ذخیره کرده و عمل اجماع نتایج را با استفاده از این برچسب‌ها انجام می‌دهیم.

جهت اجماع خوشه‌بندی‌های پایه با استفاده از برچسب‌های حاصل، روش کار بدین صورت است که یک آرایه دوبعدی A با ابعاد n سطر و m ستون ایجاد می‌شود که n بیانگر تعداد تصاویر ورودی و m تعداد شبکه‌های عصبی کانولوشنی است. هر عنصر A_{ij} آرایه متناظر با برچسبی است که به تصویر iاُم در خوشه‌بندی بردارهای بازنمایی حاصل از شبکه عصبی jاُم تخصیص داده شده

است. بـه عنـوان مثـال در شـکل (۳)، چگـونگی ترکیـب خوشه‌بندی‌های پایه و شکل‌گیری ماتریس اجماع به ازای شش داده ورودی (n=6)، سه خوشـه‌بندی پایـه (m=3) و سـه کـلاس در هـر خوشه‌بندی (r=3) نشان داده شده است.



ماتریس اجماع

	C1	C2	C3
X1	1	2	1
X2	2	3	3
X3	3	1	2
X4	3	1	2
X5	1	3	1
X6	2	2	3
X7	2	3	3
X8	1	3	1
X9	3	2	2

شکل (۳) : مثالی از چگونگی ایجاد ماتریس اجماع

۴-۳- خوشه‌بندی نهایی

پـس از تولیـد و ترکیـب خوشـه‌بندی‌های پایـه در قالـب آرایـه دوبعدی اجماع (خروجی مرحله قبل)، بـه ازای هـر تـصویر ورودی یک بـردار ویژگی خـواهیم داشـت کـه مقـادیر ایـن بـردار متنـاظر بـا برچسـب خوشـه‌هایی اسـت کـه بـه آن تـصویر در هـر خوشـه‌بندی حاصـل از شـبکه‌های عصـبی عمیـق نسـبت داده شـده اسـت. بـه عبارتی هـر سطر iاُم این آرایه، برداری به طول m (تعـداد شـبکه‌های عصبی) است که مقادیر آن حـاوی برچسـب‌های تـصویر iاُم در هـر یک از خوشه‌بندی‌های پایه است. به عنوان نمونه، در مثال شکل (۳) بردار ویژگی داده X1 به صورت X1=(1,2,1) می‌باشد. بـدین ترتیـب انتظـار مـی‌رود تـصاویری کـه بردارهـای ویژگـی آنهـا نقـاط مشترک (برچسب‌های یکسـان) بیشـتری بـا یکـدیگر داشـته باشـند، شـباهت بیشـتری بـا یکـدیگر داشـته و در دسـته‌های یکسـانی قـرار گرفته باشند.

در انتهـا جهـت دسـتیابی بـه خوشـه‌بندی نهـایی تصـاویر، بردارهای ویژگی حاصل مجدداً بـا اسـتفاده از الگـوریتم K-means افراز می‌شـوند. ایـن خروجـی بـه ازای هـر نمونـه تصـویر متنـاظر بـا خوشه‌ای است که به هر نمونه پس از عمل اجماع نسـبت داده شـده است. خوشه‌های حاصل به عنوان خروجی مـدل EDIC برگردانـده می‌شوند.

۴- آزمایشات و نتایج

در این بخش، کارایی روش خوشه‌بندی پیشـنهادی EDIC ارزیـابی شده و عملکرد آن بـا رویکردهـای اخیـر خوشـه‌بندی تصـاویر مـورد مقایسه قرار می‌گیرد. در ایـن راسـتا از مجموعـه داده‌هـای تصـویری متعـدد و معیارهـای ارزیـابی متنـوعی اسـتفاده شـده اسـت کـه توضیحات آنها در ادامه این بخش آمده است.

۴-۱- مجموعه داده‌ها و معیارهای ارزیابی

ارزیابی روش EDIC با استفاده از چهار مجموعه داده پرکاربرد در زمینه خوشه‌بندی تصاویر صورت گرفته که جزئیات تصاویر (داده‌های مورد آزمایش) و تعداد کلاس‌های آنها مطابق جدول (۲) می‌باشد. همچنین توصیفی از هر یک از این مجموعه داده‌ها در ادامه این بخش آمده است.

جدول (۲) : مجموعه داده‌های مورد آزمایش

مجموعه داده	تعداد تصاویر (n)	تعداد کلاس‌ها (r)
Fashion-Mnist	۷۰/۰۰۰	۱۰
Cifar-10	۶۰/۰۰۰	۱۰
Cifar-100	۶۰/۰۰۰	۱۰۰
ImageNet-10	۱۳/۰۰۰	۱۰

- Fashion-Mnist [۲۶]: پایگـاه داده‌ای از تصـاویر مُـد کـه حاوی ۱۰ کلاس، ۶۰ هزار داده آموزش و ۱۰ هزار داده تست است که هر کدام از داده‌ها در ابعاد ۲۸×۲۸ هستند.
- Cifar-10 و Cifar-100 [۲۷]: Cifar مجموعــــه داده‌ای از تصاویر رنگی در ابعـاد ۳۲×۳۲ اسـت کـه ۵۰ هـزار داده بـرای آمـوزش و ۱۰ هـزار داده بـرای تسـت دارد و ایـن نمونـه‌ها در نسخه های Cifar-10 و Cifar-100 به ترتیـب در ۱۰ و ۱۰۰ کلاس تقسیم بندی شده‌اند.
- ImageNet-10 [۲۸]: یکی از بزرگترین و جامع‌ترین مجموعه داده‌های موجود است که دارای ۱۰۰۰ کلاس مختلف حاوی ۱ میلیون و ۲۸۱ هزار و ۱۶۷ تصویر برای آموزش و ۱۰۰ هزار تصویر برای برای آموزش است. این مجموعه داده حجمی در حدود ۱۵۰ گیگابایت دارد. با توجه به حجم بسیار بالای مجموعه داده فوق، ما نسخه کوچکتری از آن را [۲۹] شامل ۱۳ هزار داده و ده کلاس را تفکیک کردیم و جهت ارزیابی مورد استفاده قرار دادیم.

برای محاسبه دقت نتایج روش خوشـه‌بندی پیشـنهادی از سـه معیار اندازه‌گیری مختلف به شرح زیر استفاده کردیم.

- اطلاعات متقابل نرمال شده (NMI[۱]) [۲۴]: معیاری جهت ارزیابی کیفیت خوشه‌بندی اسـت کـه میـزان اطلاعـات متقابـل بین خوشه‌های پیش‌بینی شـده را بـا کلاس‌هـای واقعـی داده‌هـا اندازه‌گیری می‌کند؛ ایـن معیـار مقـداری در بـازه صـفر تـا یـک دارد کـه بـه ازای همبسـتگی کامـل بـین نمونـه‌های پیش‌بینـی و واقعـی مقـدار یـک و در صـورتی کـه هـیچ اطلاعـات متقـابلی وجود نداشته نباشد، مقدار صفر خواهد داشت.

[۱] Normalized Mutual Information

- شاخص رُند اصلاح شده (ARI[۱]) [۲۹]: شاخصی جهت اندازه‌گیری شباهت میان خوشه‌های پیش‌بینی شده با خوشه‌های واقعی است که در صورتی که خوشه‌های فوق با یکدیگر هم پوشانی کامل داشته باشند، مقدار یک و در صورتی که برچسب‌گذاری به صورت کاملا تصادفی بوده و اشتراکی با هم نداشته باشند، مقدار صفر و در سایر موارد بسته به میزان همپوشانی نتایج، مقداری در بازه $[0,1]$ دارد.
- دقت (ACC[۲]) [۳۰]: معیار ACC دقت خوشه‌بندی حاصل را نسبت به برچسب‌های اولیه محاسبه کرده و تعداد نمونه‌هایی را که به درستی افراز شده‌اند را ارزیابی می‌کند. این معیار نیز مشابه معیارهای قبلی مقداری در بازه صفر تا یک دارد که هر چه دقت خوشه‌بندی بیشتر باشد، مقدار معیار ACC برای آن بیشتر خواهد بود.

۴-۲- تنظیمات پیاده‌سازی

جهت اجرای آزمایشات بر روی روش پیشنهادی، پارامترهای مربوط به بخش‌های مختلف بدین صورت تنظیم شده‌اند: در بخش پیش پردازش داده‌ها و جهت استفاده از یادگیری انتقال با شبکه‌های vgg16 و ResNet-50، ابعاد تصاویر ورودی $(l \times w)$ به ۴۸×۴۸ تغییر اندازه داده شد. جهت آموزش شبکه‌های CAE، تابع فعالیت Adam با نرخ یادگیری ۳-۱۰ و اندازه دسته[۳] با سایز ۱۲۸ مورد استفاده قرار گرفت. همچنین به منظور خوشه‌بندی بردارهای بازنمایی حاصل از هر شبکه و اجماع نتایج آنها از الگوریتم K-means با $T=100$ بار تکرار و تعداد خوشه‌هایی متناظر با تعداد کلاس‌های اولیه هر مجموعه داده $(r=k)$ استفاده شده است که در نهایت بهترین نتیجه حاصل از ۱۵ بار اجرای الگوریتم K-means جهت خروجی مدل برگردانده می‌شود.

۴-۳- ارزیابی نتایج

در ادامه به بررسی و مقایسه مدل پیشنهادی EDIC با روش‌های خوشه‌بندی پیشین می‌پردازیم. روش‌های مورد مقایسه شامل مجموعه ای از الگوریتم‌های کلاسیک خوشه‌بندی همانند K-means، خوشه‌بندی طیفی (SC) [۸]، خوشه‌بندی تجمعی (AC) [۳۱] و فاکتورگیری ماتریس غیرمنفی (NMF) [۷] و همچنین روش‌های اخیر خوشه‌بندی عمیق تصاویر همانند روش‌های مبتنی بر خودرمزگذار (AE) [۹]، خوشه‌بندی عمیق مبتنی بر تعبیه‌سازی (DEC[۴]) [۳۲] و یادگیری همزمان بدون ناظر (JULE) [۱۵] می‌باشند.

کارایی روش‌های خوشه‌بندی مختلف بر اساس هر یک از معیارهای NMI، ACC و ARI به ترتیب در جداول (۳)، (۴) و (۵) لیست شده‌اند. در هر جدول نتایج خوشه‌بندی الگوریتم‌های مختلف طبق معیار مورد بررسی در هر مجموعه داده مشخص شده است.

جدول (۳) : نتایج معیار NMI روش‌های خوشه‌بندی مختلف بر روی مجموعه داده‌های مختلف

	Fashion	CIFAR-10	CIFAR-100	ImageNet-10
K-means	۰/۵۱۲	۰/۰۸۷	۰/۰۸۴	۰/۱۱۹
SC	۰/۶۵۹	۰/۱۰۳	۰/۰۹	۰/۱۵۱
AC	۰/۵۶۴	۰/۱۰۵	۰/۰۸۹	۰/۱۳۸
NMF	۰/۴۲۵	۰/۰۸۱	۰/۰۷۹	۰/۱۳۲
AE	۰/۵۶۷	۰/۲۳۹	۰/۱۰	۰/۲۱
JULE	۰/۶۰۸	۰/۱۹۲	۰/۱۰۳	۰/۱۷۵
DEC	۰/۶۰۱	۰/۲۵۷	۰/۱۳۶	۰/۲۸۲
EDIC	**۰/۹۲**	**۰/۵۷۲**	**۰/۴۳**	**۰/۴۰۱**

جدول (۴) : نتایج معیار ACC روش‌های خوشه‌بندی مختلف بر روی مجموعه داده‌های مختلف

	Fashion	CIFAR-10	CIFAR-100	ImageNet-10
K-means	۰/۴۷۴	۰/۲۲۹	۰/۱۳۱	۰/۲۴۱
SC	۰/۵۸۳	۰/۲۴۷	۰/۱۳۶	۰/۲۴۷
AC	۰/۵۰	۰/۲۲۸	۰/۱۳۸	۰/۲۴۲
NMF	۰/۴۳۴	۰/۱۹	۰/۱۱۸	۰/۲۳
AE	۰/۵۴	**۰/۳۱۴**	۰/۱۶۵	۰/۳۱۷
JULE	۰/۵۶۳	۰/۲۷۲	۰/۱۳۷	۰/۳۰
DEC	۰/۵۹	۰/۳۰	**۰/۱۸۵**	**۰/۳۸۱**
EDIC	۰/۶۰	۰/۳۰۱	۰/۱۸۱	۰/۲۵

جدول (۵) : نتایج معیار ARI روش‌های خوشه‌بندی مختلف بر روی مجموعه داده‌های مختلف

	Fashion	CIFAR-10	CIFAR-100	ImageNet-10
K-means	۰/۳۴۸	۰/۰۴۹	۰/۰۲۸۹	۰/۰۵۷
SC	۰/۴۶۸	۰/۰۸۵	۰/۰۲۲	۰/۰۷۶
AC	۰/۳۷۱	۰/۰۶۵	۰/۰۳۴	۰/۰۶۷
NMF	–	۰/۰۳۴	۰/۰۲۶	۰/۰۶۵
AE	۰/۳۹۱	۰/۱۶۹	۰/۰۴۸	۰/۱۵۲
JULE	–	۰/۱۳۸	۰/۰۳۳	۰/۱۳۸
DEC	۰/۴۴۶	۰/۱۶۱	۰/۰۵	۰/۲۰۳
EDIC	**۰/۹۲**	**۰/۲۹**	**۰/۱۲**	**۰/۳۵**

همانطور که از مقایسه نتایج حاصل از روش‌های مختلف پیداست، تقریباً در هر سه معیار ارزیابی مدل پیشنهادی EDIC عملکرد بهتری نسبت به سایر روش‌ها در هر چهار مجموعه داده

[۱] Adjusted Rand Score

[۲] Accuracy

[۳] Batch size

[۴] Deep Embedding Clustering

داشته است. به طور خاص، روش‌های خوشه‌بندی عمیق کارایی قابل توجهی نسبت به روش‌های خوشه‌بندی کلاسیک در اکثر مجموعه داده‌ها داشته که این امر به دلیل توانایی یادگیری بازنمایی‌های تصاویر توسط شبکه‌های عصبی عمیق و استخراج ویژگی‌های مهم تصاویر است. در میان روش‌های خوشه‌بندی عمیق، روش EDIC توانسته به نتایج بهتری نسبت به دیگر روش‌های عمیق دست یابد. به عنوان نمونه روش ما به لحاظ معیار NMI در هر یک از مجموعه داده‌های Fashion، CIFAR-10، CIFAR-100 و ImageNet-10 به ترتیب به خروجی ۰/۹۲، ۰/۵۷۲، ۰/۴۳ و ۰/۴۰۱ دست یافته است در حالیکه بهترین مقدار این معیار در سایر روش‌های خوشه‌بندی عمیق شامل ۰/۶۵۹، ۰/۲۵۷، ۰/۱۳۶ و ۰/۲۸۲ می باشد. ضمن اینکه این مقادیر به یک روش مشخص مربوط نمی‌شود و بدین معنی است که روش‌های خوشه‌بندی فوق الزاماً در همه این معیارها دقت مناسبی ندارند و عملکرد آنها از جنبه‌های مختلف متفاوت بوده و پایدار نیست. در مورد نتایج حاصل از معیار ACC در مجموعه داده ImageNet-10 که مدل ما دقت کمتری نسبت به برخی روش‌ها به دست آورده، لازم به توضیح است که این امر به دلیل استفاده از زیرمجموعه‌ای از این مجموعه داده (به دلیل حجم بالا) جهت ارزیابی کارایی و نامتوازن شدن اندازه خوشه‌های واقعی موجود بین این داده‌ها اتفاق داده که مشکل فوق با ارزیابی از طریق معیارهای NMI و ARI برطرف شد و روش ما طبق هر دوی این معیارها به بهترین خروجی نسبت به سایر روش‌ها دست یافته است.

لذا می‌توان نتیجه گرفت که استفاده از ویژگی‌های چندگانه شبکه‌های عصبی عمیق طبق رویکرد یادگیری گروهی توانسته اطلاعات معنایی بیشتری را در مورد میزان شباهت/عدم شباهت میان داده‌های تصویری استخراج کند و لذا به نتایج بهتر و خوشه‌بندی مؤثرتری از داده‌های تصویری دست یابد.

۴-۴- تحلیل پیچیدگی محاسباتی

در این بخش به بررسی پیچیدگی محاسباتی مراحل مختلف روش پیشنهادی پرداخته و منابع مورد نیاز را به لحاظ زمانی و حافظه مورد ارزیابی قرار می‌دهیم. در تمامی موارد n بیانگر تعداد تصاویر ورودی، d بیانگر ابعاد هر تصویر ورودی ($d = w*l$) و m متناظر با تعداد خودرمزگذارهای کانولوشنی است.

- **پیش پردازش:** در این گام، از رویکرد یادگیری انتقال جهت افزایش سرعت و دقت مدل پیشنهادی استفاده شده است که طی آن یکسری ویژگی‌های اولیه تصاویر بر اساس مدل‌های از قبل آموزش دیده استخراج شده و پس از تنظیم ابعاد تصاویر مورد استفاده قرار می‌گیرند. کلیه این عملیات با مرتبه $O(nd)$ انجام می‌شود که طی آن n تصویر d بعدی در زمان ثابتی پیش پردازش شده و آماده ورود به شبکه می‌شوند.

- **یادگیری بازنمایی تصاویر با خودرمزگذارهای مبتنی بر شبکه‌های عصبی کانولوشن:** در این مرحله تعداد ثابت m خودرمزگذار عمیق جهت یادگیری بردارهای بازنمایی تصاویر مورد استفاده قرار می‌گیرند که طی آن هر خودرمزگذار با استفاده از چند لایه کانولوشنی در رمزگذار و رمزگشا، تصاویر ورودی را در لایه خروجی بازسازی می‌کند. در هر شبکه عمیق، تصاویر پیش پردازش شده h بعدی $(h = w_1 * l_1)$ طی یک روال تکرار شونده با T تکرار آموزش دیده و تابع خطای BCE طبق رابطه (۱) در هر مرحله محاسبه می‌شود. این تابع شامل یکسری عملیات جمع، تفریق و ضرب برداری است که مجموعه این عملیات رابطه خطی با ابعاد تصویر دارند. بدین ترتیب پیچیدگی زمانی آموزش شبکه‌های خودرمزگذارهای کانولوشنی به صورت $O(mTnh)$ خواهد بود. از آنجا که طبق نتایج حاصل، به ازای $T \leq 200$ تعبیه‌سازی تصاویر با دقت خوبی انجام می‌شود و با توجه به مقدار کوچک m $(m = 5)$ می‌توان پیچیدگی زمانی این مرحله را از مرتبه $O(nh)$ در نظر گرفت.

- **خوشه‌بندی‌های پایه و اجماع نتایج:** پس از تولید بردارهای بازنمایی هر تصویر در گام قبل، بازنمایی‌های حاصل از هر شبکه مستقلاً خوشه‌بندی شده و سپس نتایج خوشه‌بندی‌های حاصل با یکدیگر تجمیع می‌شوند. جهت انجام خوشه‌بندی‌ها از الگوریتم K-means استفاده کردیم که به ازای حداکثر k خوشه، دارای پیچیدگی $O(knh)$ می‌باشد. لذا خوشه‌بندی بردارهای حاصل از تمامی خودرمزگذارها از مرتبه $O(mknh)$ خواهد بود. در مرحله اجماع، ماتریسی با ابعاد $n \times m$ خواهیم داشت که هر سطر آن متناظر با برچسب هر تصویر در خوشه‌بندی‌های حاصل از m شبکه عصبی عمیق است. لذا پیچیدگی زمانی کل این مرحله به صورت $O(mknh)$ می‌باشد که با توجه به مقادیر ثابت و کوچک m و k، مرتبه آن را به صورت $O(nh)$ می‌توان در نظر گرفت.

- **خوشه‌بندی نهایی:** در آخرین مرحله از مدل پیشنهادی، از بردارهای ویژگی حاصل از مرحله قبل (هر سطر ماتریس اجماع) جهت دسته‌بندی نهایی تصاویر استفاده شد. بدین منظور نیز الگوریتم K-means به کار برده شد که بردارهای ویژگی را در مرتبه زمانی $O(knh)$ خوشه‌بندی می‌کند.

با جمع‌بندی پیچیدگی زمانی مراحل مختلف می‌توان نتیجه گرفت که مرتبه زمانی کل روش EDIC به صورت $O(nd + nh)$ است که به صورت $O(nd)$ خلاصه می‌شود. از آنجا که ابعاد تصاویر ورودی (d) نسبت به تعداد تصاویر ورودی (n) بسیار کوچکتر است، می‌توان نتیجه گرفت روش EDIC، عمل خوشه‌بندی تصاویر را در زمانی خطی از مرتبه تعداد تصاویر ورودی انجام می‌دهد.

در مورد پیچیدگی حافظه EDIC، حافظه مورد نیاز جهت آموزش شبکه‌های عمیق و تولید بردارهای بازنمایی تصاویر از مرتبه $O(mnd)$ است. جهت ذخیره سازی ماتریس اجماع نیز به

حافظه‌ای با پیچیدگی $O(mn)$ نیاز است. لذا پیچیدگی حافظه کل الگوریتم EDIC به صورت $O(mnd)$ می‌باشد.

با توجه به اهمیت زمان مصرفی الگوریتم‌های خوشه‌بندی تصاویر، در جدول (۶) پیچیدگی زمانی روش‌های مختلف به ازای تعداد n داده تصویری و ابعاد d بعد به ازای هر تصویر مورد مقایسه قرار گرفته است.

جدول (۶) : پیچیدگی زمانی روش‌های خوشه‌بندی مختلف

روش خوشه‌بندی	پیچیدگی زمانی
K-means	$O(n^2)$
SC	$O(n^3)$
AC	$O(n^3)$
NMF	$O(nd)$
AE	$O(nd)$
JULE	$O(nd)$
DEC	$O(nd)$
EDIC	$O(nd)$

همانطورکه از مقادیر جدول فوق پیداست، الگوریتم‌های مختلف دارای پیچیدگی زمانی حداقل از مرتبه $O(nd)$ هستند. لذا مدل EDIC نیز با دارا بودن حداقل مرتبه زمانی، کارایی مناسبی در مقایسه با روش‌های پیشین دارد.

۵- خلاصه

در این مقاله روشی با عنوان EDIC جهت خوشه‌بندی عمیق تصاویر پیشنهاد شد، بدین صورت که ابتدا ویژگی‌های چندگانه داده‌های تصویری با خودرمزگذارهای کانولوشنی مختلف استخراج شد. سپس نتایج حاصل از آنها طبق رویکرد یادگیری گروهی با یکدیگر ترکیب شده و عمل خوشه‌بندی نهایی با در اختیار داشتن اطلاعات مؤثرتری از میزان شباهت و تمایز میان تصاویر صورت گرفت. نتایج حاصل از اجرای روش EDIC بر روی چهار مجموعه داده تصویری پراستناد و طبق سه معیار ارزیابی مختلف نشان‌دهنده برتری روش پیشنهادی نسبت به الگوریتم‌های خوشه‌بندی سنتی و رویکردهای اخیر خوشه‌بندی عمیق می‌باشد.

در راستای توسعه روش پیشنهادی می‌توان انواعی از روش‌های خوشه‌بندی سنتی و اخیر را در دسته‌بندی اولیه و نهایی تصاویر مورد بررسی قرار داد و همچنین از روش‌های دیگری نیز در جهت اجماع نتایج پایه استفاده کرد. همچنین یکپارچه کردن عملیات تعبیه‌سازی و خوشه‌بندی در یک فرآیند همزمان و احتساب خطای خوشه‌بندی در کنار خطای بازسازی تصاویر در شبکه‌های عمیق در جهت افزایش دقت بردارهای بازنمایی حاصل، یکی دیگر از جنبه‌های تحقیقاتی پیش‌رو است.

مراجع

[1] M. Danesh, M. Dorrigiv and F. Yaghmaee, “DGCU: A new deep directed method based on Gaussian embedding for clustering uncertain graphs,” Computers and Electrical Engineering, vol. 101, 108066, 2022.

[2] M. Danesh, M. Dorrigiv and F. Yaghmaee, “A survey of clustering large probabilistic graphs: Techniques, evaluations, and applications, Expert Systems, vol. 40, no. 6, e13248, 2023.

[3] N. Adaloglou, F. Michels, H. Kalisch and M. Kollmann, “Exploring the Limits of Deep Image Clustering using Pretrained Models,” arXiv preprint arXiv: 2303.17896, 2023.

[4] M. Danesh, M. Dorrigiv and F. Yaghmaee, “Ensemble-based clustering of large probabilistic graphs using neighborhood and distance metric learning,” The Journal of Supercomputing, vol. 77, pp. 4107–4134, 2021.

[5] N. Khare, P.S. Thakur, P. Khanna, and A. Ojha, “Analysis of Loss Functions for Image Reconstruction Using Convolutional Autoencoder,” in Proc. of Computer Vision and Image Processing (CVIP 2021), vol. 1568, 2022.

[6] C. M. Bishop and N. M. Nasrabadi, Pattern recognition and machine learning. Springer, vol. 4, no. 4, 2006.

[7] D. Cai, X. He, X. Wang, H. Bao, and J. Han, “Locality preserving nonnegative matrix factorization,” in Proc. of International Joint Conference on Artificial Intelligence (IJCAI), 2009.

[8] A. Y. Ng, M. I. Jordan, and Y. Weiss, “On spectral clustering: Analysis and an algorithm,” in Proc. of Advances in Neural Information Processing Systems (NeurIPS), pp. 849–856, 2002.

[9] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, “Greedy layer-wise training of deep networks,” Advances in Neural Information Processing Systems (NeurIPS), vol. 19, 2006.

[10] B. Yang, X. Fu, N. D. Sidiropoulos, and M. Hong, “Towards k-means friendly spaces: Simultaneous deep learning and clustering,” in Proc. of International Conference on Machine Learning (ICML), pp. 3861–3870, 2017.

[11] M.M. Fard, T. Thonet, and E. Gaussier, “Deep k-means: Jointly clustering with k-means and learning representations,” Pattern Recognition Letter, vol. 138, pp. 185–192, 2020.

[12] J. Lv, Z. Kang, X. Lu, and Z. Xu, “Pseudo-supervised deep subspace clustering,” IEEE Transactions on Image Processing, vol. 30, pp. 5252–5263, 2021.

[13] B. Diallo, J. Hu, T. Li, G. Khan, and et al., “Deep embedding clustering based on contractive autoencoder,” Neurocomputing, , vol. 433, pp. 96–107, 2021.

[14] S. Chang, “Deep clustering with fusion autoencoder,” arXiv preprint arXiv: 2201.04727, 2022.

[15] J. Yang, D. Parikh, and D. Batra, “Joint unsupervised learning of deep representations and image clusters,” in Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5147–5156, 2016.

[16] M. Caron, P. Bojanowski, A. Joulin, and M. Douze, “Deep clustering for unsupervised learning of visual features,” in Proc. of European Conference on Computer Vision (ECCV), pp. 132–149, 2018.

[17] B. Sen, P. Zhou, L. Du, and X. Li, “Active deep image clustering,” Knowledge Based Systems, vol. 252, 2022.

[18] D. Huang, D-H. Chen, X. Chen and et al., “DeepCluE: Enhanced Image Clustering via Multi-layer Ensembles in Deep Neural Networks,” arXiv preprint arXiv: 2206.00359, 2022.

[19] N. Iam-On, T. Boongoen, S. Garrett, and C. Price, “A link-based approach to the cluster ensemble problem,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 33, no. 12, pp. 2396–2409, 2011.

[20] J. Yi, T. Yang, R. Jin, A. K. Jain, and M. Mahdavi, “Robust ensemble clustering by matrix completion,” in Proc. of IEEE International Conference on Data Mining, pp. 1176–1181, 2012.

[21] L. Franek and X. Jiang, “Ensemble clustering by means of clustering embedding in vector spaces,” Pattern Recognition, vol. 47, no. 2, pp. 833–842, 2014.

[22] D. Huang, J. Lai, and C.-D. Wang, “Ensemble clustering using factor graph,” Pattern Recognition, vol. 50, pp.131–142, 2016.

[23] A. Strehl and J. Ghosh, “Cluster ensembles: A knowledge reuse framework for combining multiple partitions,” Journal of Machine Learning Research, vol. 3, pp. 583–617, 2003.

[24] X. Z. Fern and C. E. Brodley, “Solving cluster ensemble problems by bipartite graph partitioning,” in Proc. of International Conference on Machine Learning (ICML), 2004.

[25] H. Xiao, K. Rasul, and R. Vollgraf, “Fashion-MNIST: A novel image dataset for benchmarking machine learning algorithms,” arXiv preprint arXiv:1708.07747, 2017.

[26] A. Krizhevsky, G. Hinton et al., “Learning multiple layers of features from tiny images,” 2009.

[27] Y. Li, P. Hu, Z. Liu, D. Peng, J. T. Zhou, and X. Peng, “Contrastive clustering,” in Proc. of AAAI Conference on Artificial Intelligence (AAAI), 2021.

[28] D. Huang, C.-D. Wang, J.-H. Lai, and C.-K. Kwoh, “Toward multidiversified ensemble clustering of high-dimensional data: From subspaces to metrics and beyond,” IEEE Transactions on Cybernetics, in press, 2021.

[29] N. Nguyen and R. Caruana, “Consensus clusterings,” in Proc. of IEEE International Conference on Data Mining (ICDM), pp. 607–612, 2007.

[30] A. K. Jain, “Data clustering: 50 years beyond k-means,” Pattern Recognition Letters, vol. 31, no. 8, pp. 651-666, 2010.

[31] J. Xie, R. Girshick, and A. Farhadi, “Unsupervised deep embedding for clustering analysis,” in Proc. of International Conference on Machine Learning (ICML), pp. 478–487, 2016.

ملیحه دانش مدرک کارشناسی و کارشناسی ارشد خود را در مهندسی کامپیوتر گرایش نرم‌افزار از دانشگاه فردوسی مشهد دریافت کرد. او مدرک دکتری خود در مهندسی کامپیوتر را از دانشگاه سمنان در گرایش هوش مصنوعی و رباتیکز اخذ نمود. ایشان هم اکنون استادیار دانشکده مهندسی برق و کامپیوتر دانشگاه علم وفناوری مازندران هستند. زمینه‌های تحقیقاتی ایشان شامل یادگیری ماشین، گراف کاوی، بینایی ماشین و تحلیل شبکه‌های اجتماعی است.

مصطفی نیکوسرشت دانش آموخته کارشناسی مهندسی کامپیوتر در دانشگاه علم و فناوری مازندران است. زمینه‌های پژوهشی ایشان یادگیری ماشین، پردازش تصویر و پردازش متن است.