طراحی یک شبکهٔ عصبی کانولوشنال مبتنی‌بر بانک فیلتر برای طبقه‌بندی تصاویر اعداد دست‌نویس

دروگرمقدم، علی [۱]، کرمی‌ملایی، محمدرضا [۲*]، حسن‌زاده، محمدرضا [۳]

[۱] دانشجوی دکتری مهندسی پزشکی، گروه مهندسی پزشکی-بیوالکتریک، دانشکدهٔ مهندسی برق و کامپیوتر، دانشگاه صنعتی نوشیروانی بابل، بابل، ایران

[۲] دانشیار، گروه گروه مهندسی پزشکی-بیوالکتریک، دانشکدهٔ مهندسی برق و کامپیوتر، دانشگاه صنعتی نوشیروانی بابل، بابل، ایران

[۳] استادیار، گروه مهندسی برق-الکترونیک، دانشکدهٔ مهندسی برق و کامپیوتر، دانشگاه صنعتی نوشیروانی بابل، بابل، ایران

چکیده

در سال‌های اخیر شبکه‌های عصبی کانولوشنال به‌طور فزاینده‌ای در کاربردهای مختلف بینایی ماشین و به‌ویژه در شناسایی و طبقه‌بندی خودکار تصاویر استفاده شده‌اند. این نوع از شبکه‌های عصبی مصنوعی با شبیه‌سازی عملکرد قشر بینایی مغز قدرتمندترین ساختار را در تجزیه و تحلیل داده‌های بصری دارند. اما تنوع تصاویر دیجیتال و گوناگونی محتوا و ویژگی‌های آن‌ها ایجاب می‌کند برای دستیابی به کارایی بالاتر در هر مسئلهٔ طبقه‌بندی، شبکه‌های کانولوشنال به‌صورت اختصاصی طراحی و شاخص‌های آن‌ها به‌دقت تنظیم شوند. در این راستا، در پژوهش حاضر ضرایبی بهینه برای صافی‌های لایهٔ کانولوشن در شروع آموزش شبکه به‌کاررفته تا از این طریق دقت طبقه‌بندی در شبکه افزایش، و زمان آموزش کاهش یابد. این کار با طراحی و به‌کارگیری مجموعه‌ای از صافی‌های تخصصی برای لایهٔ کانولوشن در قالب یک بانک فیلتر و جای‌گذاری آن‌ها به‌جای صافی‌های تصادفی انجام شده و بر روی پایگاه دادهٔ تصاویر اعداد دست‌نویس MNIST ارزیابی شده‌است. آزمایش‌های ما بر روی شبکهٔ کانولوشنال تک‌لایه با سه نوع صافی‌گذاری (صافی‌های عدد ثابت، عدد تصادفی و بانک فیلتر) میانگین دقت طبقه‌بندی تصاویر اعداد دست‌نویس MNIST را در پنجاه بار آموزش شبکه به‌ترتیب ۷۴/۹۴، ۸۶/۴۷ و ۹۱/۸۹ درصد و برای شبکهٔ کانولوشنال سه‌لایه به ترتیب ۸۸/۸۲، ۹۶/۱۶ و ۹۹/۱۴ درصد نشان دادند. این نتایج نشان می‌دهند که صافی‌های به‌کاررفته در الگوی پیشنهادی در مقایسه با صافی‌های تصادفی ویژگی‌های مؤثرتری را از تصاویر استخراج کرده، و با شروع آموزش شبکه از نقطهٔ مناسب‌تر، بدون افزایش هزینهٔ محاسباتی دقت طبقه‌بندی را افزایش داده‌اند. بنابراین، می‌توان نتیجه گرفت که ضرایب اولیهٔ صافی‌های لایهٔ کانولوشن بر دقت طبقه‌بندی شبکه‌های کانولوشنال مؤثر است و با به‌کارگیری صافی‌های مؤثرتر در لایهٔ کانولوشن می‌توان این شبکه‌ها را خاص مسئله کرده و از این طریق کارآیی شبکه را افزایش داد.

واژگان کلیدی: شبکه‌های عصبی کانولوشنال، طبقه‌بندی تصاویر، اعداد دست‌نویس، پایگاه داده MNIST

Design of a Filter Bank-based Convolutional, Neural Network to Classify Handwritten Digit Images

Derogar Moghadam, Ali[1], Karami Mollaei, Mohammad Reza[2*], Hassanzadeh, Mohammadreza[3]

1 Ph.D. Student, Bioelectric Department, Electrical & Computer Faculty, Babol Noshirvani University of Technology, Babol, Iran

2 Associate Professor, Bioelectric Department, Electrical & Computer Faculty, Babol Noshirvani University of Technology, Babol, Iran

3 Assistant Professor, Electrical Engineering Department, Electrical & Computer Faculty, Babol Noshirvani University of Technology, Babol, Iran

* **Corresponding author**

* نویسندهٔ عهده‌دار مکاتبات

• تاریخ ارسال مقاله: ۱۴۰۱/۴/۷ • تاریخ پذیرش: ۱۴۰۲/۹/۵ • تاریخ انتشار: ۱۴۰۲/۱۰/۲۴ • نوع مطالعه: پژوهشی

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-25]

[DOI: 10.61186/jsdp.20.3.127]

Background: In recent years, convolutional, neural networks (CNNs) have been increasingly used in various applications of machine vision. CNNs simulate the function of the brain's visual cortex and they have a powerful structure to analyze visual images. However, the diversity of digital images and the diversity of their content and features require that, CNN networks are specially designed and their parameters are carefully adjusted to achieve higher efficiency in any classification problem. In this regard, in many previous studies, researchers have attempted to increase the efficiency of the CNNs by setting their adjustable parameters, more accurately.

New method: In this study, we presented a novel initializing method for the kernels of the first convolutional layer of the CNN networks. We designed a filter bank with specialized kernels and used them in the first convolution layer of the proposed models. These kernels compared to the random kernels (in traditional CNNs), without increasing the computational cost of the network extract more effective features from the input images and improve classification accuracy by covering all the important characteristics.

Results: The dataset used in this paper was the MNIST database of handwritten digits. We examined the performance of CNN networks when three different types of kernels were used in their first convolution layer. The first group of kernels had constant coefficients; the second group had random coefficients and finally the kernels of the third group were specially designed to extract a wide range of image features. Our experiments on a single-layer CNN network with three types of kernels (constant numbers, random numbers, and filter-bank), showed the average classification accuracy of MNIST images in 50 times of network training to be 74.94, 86.47, and 91.89 percent, respectively, and for a three-layer CNN network, 88.82, 96.16, and 99.14 percent, respectively.

Comparison with existing methods: Compared to the kernels with randomized coefficients, the use of specialized kernels in the first convolution layer of the CNN networks has several important advantages: 1) They can be designed to extract all important features of the input images, 2) They can be designed more effectively based on the problem in hand, 3) They cause the training start from a more appropriate point, and in this way, the speed of training and the classification accuracy of network increase.

Conclusion: This study provides a novel method to initialize kernels in convolution layers of CNN networks to enhance their performance in image classification works. Our results show that compared to random kernels, the kernels used in the proposed models extract more effective features from the images at different frequencies and increase the classification accuracy by starting the training algorithm from a more appropriate point, without increasing the computational cost. Therefore, it can be concluded that the initial coefficients of the convolution layer kernels are effective on the classification accuracy of CNN networks, and by using more effective kernels in the convolution layers, these networks can be made specific to the problem and in this way, increase the efficiency of the network.

Keywords: convolutional neural network, image classification, handwritten digit, MNIST database

۱- مقدمه

شناسایی و طبقه‌بندی خودکار تصاویر، یکی از مهم‌ترین و پرکاربردترین مسائل در حوزهٔ بینایی ماشین است و در یک دههٔ اخیر کاربردهای روزافزونی در پردازش تصاویر پزشکی، تصاویر ماهواره‌ای و سنجش از راه دور، صنایع غذایی، شناسایی چهره و حالات آن، شناسایی حروف و اعداد، و بسیاری از حوزه‌های مرتبط با تصاویر دیجیتال یافته‌است [۱]. اگرچه برای انسان شناسایی و طبقه‌بندی تصاویر با محتوای متفاوت بر اساس ویژگی‌های نهفته در آن‌ها کاری آسان است، اما رایانه‌ها نیازمند الگوریتم‌هایی بسیار پیچیده و زمان‌بر هستند تا بتوانند تصاویر مختلف را با دقت بالا از یکدیگر تشخیص دهند [۲]. در کلی‌ترین حالت می‌توان یک سامانهٔ خودکار برای طبقه‌بندی تصاویر را متشکل از دو بخش اصلی در نظر گرفت؛ بخش استخراج ویژگی و بخش طبقه‌بندی. بر این اساس، یک سامانهٔ قابل‌اعتماد و دقیق برای طبقه‌بندی تصاویر، نخست باید بتواند ویژگی‌هایی را از تصاویر ورودی استخراج کند که در تصاویر یک طبقه مشابه و در طبقات متفاوت به‌خوبی قابل تشخیص و تمایز باشند، و سپس این ویژگی‌ها را به‌وسیلهٔ یک طبقه‌بند مناسب در رده‌های موجود تقسیم‌بندی کند. ازاین‌رو، تاکنون روش‌های مختلفی برای استخراج ویژگی‌های مؤثّر از تصاویر و طبقه‌بندهای متفاوتی نیز برای انجام یک رده‌بندی باکیفیت ارائه و بر روی داده‌های تصویری با محتوای متفاوت آزموده شده‌اند [۳]، [۴].

بسیاری از پژوهش‌گران در مسئلهٔ طبقه‌بندی تصاویر بر شناسایی خودکار نویسه‌ها و اعداد دست‌نویس متمرکز بوده‌اند. چراکه شناسایی و طبقه‌بندی خودکار و دقیق ارقام دست‌نویس به روش‌های بینایی ماشین کاربردهای زیادی در دستگاه‌های خودخوان نشانی‌های پستی، چک‌ها و اسکناس‌ها، بارکدها، فرم‌های مالیاتی و بسیاری از مسائل آینده که ارتباط بین انسان و رایانه را تسهیل می‌کنند، دارد [۵]. اما، شناسایی و طبقه‌بندی خودکار ارقام دست‌نویس با چالش‌هایی مانند یکسان‌نبودن

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-25]

[DOI: 10.61186/jsdp.20.3.127]

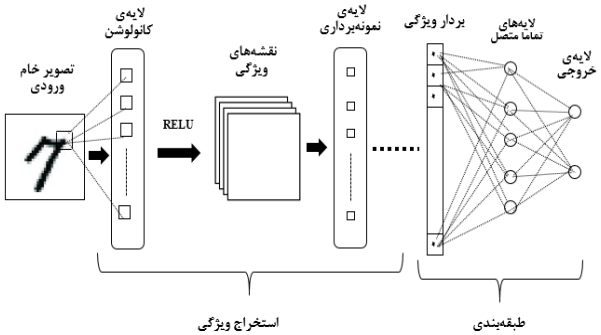
اندازه، ضخامت، جهت و موقعیت اعداد در تصویر و نیز تفاوت‌های فاحش در دست‌خط افراد روبه‌روست. بنابراین، الگوریتم‌های طبقه‌بندی تصویر در این حوزه باید از کارآیی بالایی برخوردار باشند تا به دقت و قابلیت اعتماد بالا دست یابند. الگوریتم‌های معرفی‌شده برای طبقه‌بندی ارقام دست‌نویس در یک تقسیم‌بندی کلی در چهار گروه الگوریتم‌های مبتنی‌بر تطبیق الگو[1]، روش‌های آماری[2]، روش‌های ساختاری[3] و شبکه‌های عصبی مصنوعی[4] قرار می‌گیرند [۶]. در روش‌های مبتنی‌بر تطبیق الگو، میزان شباهت بین دو بردار ویژگی (متشکل از گروه‌های پیکسلی، شکل‌ها، انحناها و غیره) در فضای ویژگی سنجش می‌شود. روش‌های آماری در اساس، بر مبنای برخی توابع تصمیم‌گیری آماری و مجموعه‌ای از معیارهای بهینه عمل و احتمال تعلق الگوی مشاهده‌شده به یک ردهٔ خاص را تعیین می‌کنند. روش‌هایی مانند K نزدیک‌ترین همسایه[5]، ماشین بردار پشتیبان[6]، الگوی مخفی مارکوف[7]، طبقه‌بند بیزین و استدلال فازی از این دسته هستند. در روش‌های ساختاری، ارقام به صورت واحدهای ساختاری درنظرگرفته‌می‌شوند، کمّی می‌شوند و رابطهٔ بین آن‌ها جستجو می‌شود. در نهایت، روش‌های مبتنی‌بر شبکه‌های عصبی مصنوعی که به‌طور وسیعی در مسائل طبقه‌بندی تصویر، و نیز در این پژوهش از آن‌ها استفاده شده‌است، ویژگی‌هایی مانند دقت و کیفیت بالا، آموزش‌پذیری خودکار بر اساس داده‌ها، عملکرد مناسب در طبقه‌بندی داده‌های نوفه‌ای، قابلیت پیاده‌سازی به‌صورت موازی و کارآمدی در یادگیری پایگاه داده‌های عظیم را دارند [۶].

شبکه‌های عصبی مصنوعی یک الگوی محاسباتی رایانه‌ای هستند که عملکرد آن‌ها از شبکه‌های عصبی زیستی در مغز انسان الهام گرفته شده‌است. در این شبکه‌ها پردازش اطلاعات از طریق تعداد زیادی واحدهای کوچک به نام نورون (گره) که در لایه‌های متوالی قرار گرفته و به هم پیوسته‌اند، انجام می‌شود. شبکه‌های عصبی مصنوعی عملکرد خیره‌کننده‌ای در بسیاری از مسائل شناسایی الگو و طبقه‌بندی دارند [۷]. اما طبقه‌بندی تصاویر به‌وسیلهٔ این شبکه‌ها همواره با چالش‌هایی روبه‌رو بوده‌است. یکی از مهم‌ترین این چالش‌ها عدم‌امکان تزریق مستقیم تصاویر به این شبکه‌ها (به‌ویژه تصاویر با ابعاد بالا) به‌دلیل تشدید افسارگسیختهٔ حجم و زمان محاسبات و نیز عملکرد پایین شبکه در طبقه‌بندی تصاویر به این روش است. به‌عبارت‌دیگر، طبقه‌بندی تصاویر به‌وسیلهٔ شبکه‌های عصبی مصنوعی در حالت کلی نیازمند مراحل پیش‌پردازشیِ استخراج و انتخاب ویژگی (کاهش ابعاد بردار ویژگی) به‌صورت دستی و قبل‌از اعمال تصویر به شبکه است. این امر به‌طور معمول، مطابق با مسئلهٔ پیش‌رو انجام شده و کاری بسیار پیچیده و وقت‌گیر است [۸]، [۹]؛ از این رو در سال‌های اخیر به شبکه‌های عصبی کانولوشنال[8] برای طبقه‌بندی تصاویر به‌شدت توجه شده‌است. این نوع از شبکه‌های عصبی مصنوعی، ساختار ویژه‌ای برای کار با تصاویر داشته و به‌جرأت می‌توان گفت که قدرتمندترین و مؤثّرترین پیکربندی را برای پردازش تصاویر دارند [۱۰]. مهم‌ترین برتری شبکه‌های کانولوشنال در هنگام کار با تصاویر، انجام عملیات استخراج ویژگی، انتخاب ویژگی و طبقه‌بندی در یک ساختار واحد و منسجم است. به‌عبارت‌دیگر، این شبکه‌ها برخلافِ سایر شبکه‌های عصبی مصنوعی نیازمند مراحل جداگانه‌ای برای استخراج ویژگی و کاهش ابعاد بردار ویژگی نیستند و به‌طور مستقیم، تصویر خام را در ورودی خود دریافت و آن را طبقه‌بندی می‌کنند [۱۱]. شبکه‌های کانولوشنال از پژوهش‌های دو پژوهشگر دانشگاه هاروارد (Hubel & Wiesel) که در دههٔ ۱۹۶۰ میلادی بر روی قشر بینایی مغز گربه آزمایش می‌کردند، منشأ گرفته‌اند [۱۲]. پژوهش‌های این پژوهشگران نشان می‌داد که در لایه‌های اولیهٔ قشر بینایی مغز، نورون‌ها به ویژگی‌های سطح پایین تصویر (مانند لبه‌ها و شدت روشنایی‌ها) و در لایه‌های بعد به ویژگی‌های پیچیده‌تر تصویر حساسیت نشان می‌دهند. بر این اساس در سال ۱۹۸۷، فوکوشیما یک الگوی نظری برای شبکه‌های کانولوشنال مطرح کرد و پس از آن، این شبکه‌ها به‌تدریج پیاده‌سازی و در عمل استفاده شدند [۱۳]. شبکه‌های کانولوشنال از این واقعیت که ورودی شبکه از جنس تصویر است، استفاده کرده و معماری شبکه را به‌طور معقولی محدود می‌کنند. این شبکه‌ها از سه نوع لایهٔ اصلی تشکیل می‌شوند: لایه‌های کانولوشن[9]، لایه‌های نمونه‌بردار کاهنده[10] و لایه‌های به‌طور کامل پیوسته[11]. به‌طوری‌که لایه‌های کانولوشن عملیات استخراج ویژگی، لایه‌های نمونه‌بردار کاهندهٔ عملیات انتخاب ویژگی و لایه‌های

[1] Template Matching
[2] Statistical Techniques
[3] Structural Techniques
[4] Artificial Neural Networks (ANN)
[5] k-Nearest-Neighbor (K-NN)
[6] Support Vector Machine (SVM)
[7] Hidden Markov Model (HMM)
[8] Convolutional Neural Network (CNN)
[9] Convolution Layer
[10] Pooling Layer
[11] Fully Connected Layers

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-25]
[DOI: 10.61186/jsdp.20.3.127]

به‌طور کامل پیوسته عملیات طبقه‌بندی را در معماری مرسوم یک شبکهٔ کانولوشنال انجام می‌دهند. شکل (۱) لایه‌های اصلی یک شبکهٔ کانولوشنال را که برای طبقه‌بندی اعداد دست‌نویس طراحی شده‌است، نشان می‌دهد.



(شکل-۱): لایه‌های اصلی یک شبکهٔ عصبی کانولوشنال در مسئلهٔ طبقه‌بندی اعداد دست‌نویس

(Figure-1): The main layers of a CNN network in handwritten digits classification

همان‌طورکه در شکل (۱) مشاهده می‌شود، ورودی شبکهٔ کانولوشنال یک تصویر خام (نه یک بردار از ویژگی‌های تصویر) است. همچنین، اگرچه یک شبکهٔ کانولوشنال می‌تواند بسته به کاربردی که دارد، تعداد و ترکیب متفاوتی از لایه‌های یادشده را داشته‌باشد، اما عملکرد این لایه‌ها در کل در دو بخش استخراج ویژگی و طبقه‌بندی تقسیم‌بندی می‌شود. تفاوت اصلی شبکه‌های کانولوشنال با سایر شبکه‌های عصبی نیز در بخش نخست این شبکه‌ها (بخش استخراج ویژگی) است. در این بخش، ویژگی‌های تصویر خام واردشده به شبکه، به‌وسیلهٔ صافی‌هایی (کرنل‌هایی) که با آن کانوالو می‌شوند، استخراج شده، سپس تعداد این ویژگی‌ها در گذر از لایه‌های بعدی شبکه کاهش می‌یابد. بنابراین، خروجی بخش استخراج ویژگی، یک بردار ویژگی است که در بخش دوم شبکه به‌وسیلهٔ یک ساختار به‌طور کامل پیوسته طبقه‌بندی می‌شود. بر این اساس، در یک شبکهٔ کانولوشنال ابعاد تصویر ورودی در عبور از لایه‌های شبکه کاهش یافته و همگام با رسیدن به انتهای شبکه، تصویر ورودی به یک بردار حاوی امتیاز دسته‌ها (کلاس‌ها) تبدیل و در نهایت طبقه‌بندی می‌شود. ویژگی‌های مهم شبکه‌های کانولوشنال عبارتند از: عملکرد ویژه بر روی ورودی‌های دو یا چندبعدی و از جنس تصویر، انجام عملیات استخراج ویژگی، انتخاب ویژگی و طبقه‌بندی در یک ساختار مجتمع و پردازش محلی ورودی با اتصال هر نورون به بخش کوچکی از حجم ورودی [۱۴].

امروزه شبکه‌های کانولوشنال افق‌های تازه‌ای را در کاربردهای مختلف بینایی ماشین روشن کرده، و مطالعات انجام‌شده بر روی آن‌ها عملکرد مناسب این شبکه‌ها را در بسیاری از زمینه‌ها به اثبات رسانده‌اند. در یک دههٔ اخیر شبکه‌های کانولوشنال در مسائل شناسایی تصاویر [۱۵]، [۱۶]، تشخیص اشیا در تصاویر [۱۷]، [۱۸]، [۱۹]، بخش‌بندی تصاویر [۲۰]، [۲۱] و طبقه‌بندی تصاویر [۲۲] عملکرد خیره‌کننده‌ای از خود نشان داده‌اند. این شبکه‌ها در شناسایی و طبقه‌بندی نویسه‌ها و اعداد دست‌نویس نیز استفاده شده‌اند. در این زمینه، نخستین بار در سال ۱۹۹۸، LeCun و همکارانش پایگاه دادهٔ MNIST[1] را طراحی و از یک شبکهٔ کانولوشنال برای طبقه‌بندی تصاویر مربوط به نویسه‌ها استفاده کردند [۲۳]. آن‌ها توانستند به کمک شبکهٔ پیشنهادی به بالاترین دقت تشخیص در آن زمان برسند. Simard و همکارانش یک شبکهٔ کانولوشنال با معماری ساده را برای طبقه‌بندی تصاویر اعداد دست‌نویس لاتین در پایگاه دادهٔ MNIST معرفی کردند [۲۴]. آن‌ها عملکرد شبکه‌های کانولوشنال را با شبکه‌های پرسپترون چندلایه[2] با ساختار به‌طور کامل پیوسته در دو لایه مقایسه کردند. در این پژوهش خطای طبقه‌بندی برای شبکهٔ کانولوشنال، چهاردهمِ درصد و برای شبکهٔ پرسپترون چندلایه، هفت‌دهمِ درصد گزارش شد. پس از اثبات کارایی شبکه‌های کانولوشنال در طبقه‌بندی تصاویر اعداد دست‌نویس در بسیاری از پژوهش‌های بعدی تلاش شد عملکرد این شبکه‌ها بهبود یابد. در این راستا، در برخی پژوهش‌ها این شبکه‌ها با الگوریتم‌های تکاملی، یا سایر روش‌های بهینه‌سازی ترکیب شدند تا انتخاب فراشاخصهای شبکه بهینه شده و دقت طبقه‌بندی افزایش یابد. در مواردی نیز تلاش بر این بود که الگوهای قوی‌تری از شبکه‌های کانولوشنال طراحی شود تا خطای طبقه‌بندی کاهش یابد.

به‌عنوان مثال، Tabik و همکارانش تأثیر برخی روش‌های پیش‌پردازش تصاویر را بر دقت طبقه‌بندی چند الگوی پیش‌ساخته از شبکه‌های کانولوشنال (LeNet [۲۳]، Network3 [۲۵] و DropConnect [۲۶]) بررسی کردند. در این مقاله از تصاویر پایگاه دادهٔ اعداد دست‌نویس لاتین MNIST استفاده شد. آن‌ها، نخست دقت الگوهای مذکور را بر روی تصاویر پایگاه داده سنجیدند، سپس این تصاویر را تحت تبدیل‌های الاستیک، انتقال، چرخش و ترکیب‌های متفاوتی از آن‌ها قرار داده و مجدد عملکرد هر الگو را در طبقه‌بندی تصاویر به‌دست‌آمده بررسی کردند. نتایج

[1] National Institute of Standards and Technology database (MNIST)

[2] Multi-Layer Perceptron (MLP)

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-25]

[DOI: 10.61186/jsdp.20.3.127]

این پژوهش نشان می‌داد که عملیات پیش‌پردازشی مانند انتقال مرکز ثقل اعداد به مرکز تصویر، خنثی‌سازی چرخش و تغییرات الاستیک در تصاویر می‌تواند دقت طبقه‌بندی را در شبکه‌های کانولوشنال تا هفتادویک‌دهمِ درصد افزایش دهد [۲۷].

Ahlawat و همکارانش با هدف بهبود دقت طبقه‌بندی تصاویر اعداد دست‌نویس پایگاه دادهٔ MNIST تلاش کردند تا بهترین پیکربندی و تنظیمات را برای فراشاخصهای شبکهٔ کانولوشنال تعیین کنند. آزمایش‌های آن‌ها بر روی دو پیکربندی مختلف شبکهٔ کانولوشنال با سه و چهار لایهٔ کانولوشن و نیز تنظیمات متفاوت فراشاخصها نشان داد که بهترین دقت طبقه‌بندی شبکهٔ پیشنهادی آن‌ها ۹۹/۸۷ درصد است [۲۸]. وی و همکارانش در پژوهشی دیگر با ترکیب ویژگی‌های کلیدی شبکه‌های کانولوشنال و طبقه-بند دودویی ماشین‌بردار پشتیبان الگویی قدرتمند برای طبقه‌بندی خودکار تصاویر پایگاه دادهٔ اعداد دست‌نویس لاتین MNIST ارائه دادند. در الگوی پیشنهادی آن‌ها لایه‌های کانولوشن و نمونه‌بردار کاهنده در شبکهٔ کانولوشنال وظیفهٔ استخراج ویژگی‌های تصویر و طبقه‌بند دودویی ماشین‌بردار پشتیبان وظیفهٔ رده‌بندی بردار ویژگی را در خروجی شبکه بر عهده داشت. الگوی پیشنهادی آن-ها توانست تنها با یک لایهٔ کانولوشن به دقت ۹۹/۲۸ درصد برسد [۲۹].

Saqib و همکارانش با ترکیب الگوریتم ماشین با امکان یادگیری شدید[1] و شبکه‌های کانولوشنال تلاش کردند دقت طبقه‌بندی تصاویر اعداد دست‌نویس در پایگاه داده MNIST را افزایش دهند. آن‌ها لایه‌های به‌طور کامل پیوستهٔ شبکهٔ کانولوشنال را با ساختار ماشین با امکان یادگیری شدید جایگزین کردند. بنابراین، در ساختار پیشنهادی آن‌ها لایه‌های کانولوشن وظیفهٔ استخراج ویژگی‌های تصویر و ساختار ماشین با امکان یادگیری شدید وظیفهٔ طبقه‌بندی بردارهای ویژگی را بر عهده داشت. شبکهٔ پیشنهادی آن‌ها از سه لایهٔ کانولوشن و سه لایهٔ طبقه‌بندی تشکیل می‌شد. دقت گزارش‌شده برای این شبکه ۹۹/۸ درصد بود [۳۰].

Calderon و همکارانش، برای بهبود عملکرد شبکه‌های کانولوشنال از صافی‌های گابور در نخستین لایهٔ کانولوشن شبکه استفاده کردند تا ویژگی‌های تصویر را در فرکانس-های مختلف استخراج کنند. شبکهٔ پیشنهادی آن‌ها دارای دو لایهٔ کانولوشن بود و از الگوریتم پس‌انتشار[2] خطا برای تصحیح شاخص‌های شبکه استفاده می‌کرد. آن‌ها خطای طبقه‌بندی شبکهٔ کانولوشنال پیشنهادی را بر روی پایگاه دادهٔ تصاویر اعداد دست‌نویس لاتین MNIST، شصت‌وهشت‌صدمِ درصد گزارش کردند [۳۱].

Le و همکارانش، برای بهبود عملکرد شبکه‌های کانولوشنال در طبقه‌بندی تصاویر، الگوریتم آموزش شبکه را تغییر دادند. در این پژوهش از یک الگوریتم مبتنی‌بر حافظهٔ محدود به‌جای الگوریتم نزولی گرادیان تصادفی[3] در آموزش شبکه استفاده شد. نتایج این پژوهش در طبقه‌بندی تصاویر اعداد دست‌نویس پایگاه داده‌ی MNIST نشان می‌داد که تغییر الگوریتم آموزش در شبکه‌های کانولوشنال می‌تواند بر سرعت آموزش شبکه و دقت طبقه‌بندی مؤثّر باشد [۳۲].

باتوجه‌به مقدمهٔ ارائه‌شده می‌توان دریافت که شبکه‌های کانولوشنال زمینه‌ای جذاب و در حال رشد در کاربردهای مختلف بینایی ماشین هستند. این شبکه‌ها اگرچه به‌طور معمول، از ساختار استانداردی استفاده می‌کنند، شامل فراشاخص‌هایی هستند که با تنظیم دقیق آن‌ها می‌توان عملکرد شبکه را بهبود بخشید. در این مقاله، ما روشی جدید برای بهبود کارآیی شبکه‌های کانولوشنال ارائه داده-ایم که امکان دست‌یابی به نتایج بهتر را در مدت زمان کمتر فراهم می‌آورد. روش پیشنهادی همچنین قادر است تا تنظیمات شبکه را بر اساس مسئلهٔ پیش‌رو تغییر داده و از این طریق عملکرد بهتر شبکه را در طبقه‌بندی تصاویر در پی داشته‌باشد.

۲- مواد و روش‌ها

در این بخش، نخست، مشخصات مجموعه داده‌ای که برای ارزیابی الگوی پیشنهادی استفاده شد، ارائه می‌شود و پس از آن عملکرد لایه‌های مختلف شبکهٔ کانولوشنال تشریح و الگوی پیشنهادی ارائه می‌شود.

۲-۱- پایگاه داده

در این پژوهش برای ارزیابی کارآیی الگوی پیشنهادی و بررسی و مقایسهٔ دقت و قابلیت اطمینان آن نسبت به حالت مرسوم شبکه‌های کانولوشنال از تصاویر موجود در پایگاه دادهٔ MNIST استفاده شد [۳۳]. این پایگاه داده در مجموع شامل هفتادهزار تصویر سیاه و سفید از اعداد دست‌نویس لاتین (۰ تا ۹) است که از ۲۵۰ فرد مختلف گرفته شده‌است. این تصاویر همگی ابعاد ۲۸×۲۸ پیکسل

[1] Pooling Extreme Learning Machine (ELM)
[2] Back Propagation
[3] Stochastic Gradient Descent (SGD)

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-25]
[DOI: 10.61186/jsdp.20.3.127]

داشته و از ۲۵۶ سطح خاکستری تشکیل شده‌اند. تصاویر از نظر اندازه عادی‌سازی شده و در مرکز قرار گرفته‌اند. همچنین، برای مقایسهٔ بهتر نتایج در پژوهش‌های مختلف شصت‌هزار تصویر برای آموزش و ده‌هزار تصویر برای تست الگوریتم‌های طبقه‌بندی تصویر توسط طراحان این پایگاه داده به‌طور جداگانه دسته‌بندی شده‌است. شکل (۲) مثالی از تصاویر پایگاه دادهٔ MNIST را نشان می‌دهد.



(شکل-۲): نمونه‌هایی از تصاویر پایگاه داده‌ی MNIST
(Figure-2): Examples of MNIST database images

۲-۲ شبکه‌های کانولوشنال و الگوی پیشنهادی

۲-۲-۱ - لایهٔ ورودی

لایهٔ ورودی در شبکه‌های کانولوشنال بر اساس نوع ورودی مشخص می‌شود. این لایه می‌تواند به‌صورت یک‌بعدی (مانند سیگنال‌ها)، دوبعدی (مانند تصاویر خاکستری)، سه‌بعدی (مانند تصاویر رنگی) و یا حتی با ابعاد بالاتر باشد. اما ساختار و عملکرد شبکه‌های کانولوشنال در اساس، بر روی ورودی‌هایی از جنس تصویر بهینه است. در الگوی پیشنهادی لایهٔ ورودی آرایه‌هایی (تصاویری) با ابعاد ۲۸×۲۸ بودند که از پایگاه داده MNIST دریافت و قبل‌از تزریق به شبکه مقادیر شدت روشنایی آن‌ها به‌وسیلهٔ معادلهٔ (۱) در بازهٔ ۱-۰ طبیعی (نرمالیزه) شده‌بود.

$$X' = \frac{X - \min(X)}{\max(X) - \min(X)} \qquad (۱)$$

در معادلهٔ (۱)، X یک آرایهٔ ۲۸×۲۸ از شدت روشنایی پیکسل‌های تصویر و min(X) و max(X)، به‌ترتیب کمترین و بیشترین شدت روشنایی تصویر در این آرایه است.

۲-۲-۲ - لایهٔ کانولوشن

لایه‌های کانولوشن هستهٔ اصلی تشکیل‌دهندهٔ یک شبکهٔ کانولوشنال هستند و بخش مهمی از بار محاسباتی شبکه در این لایه‌ها انجام می‌شود. وظیفهٔ این لایه‌ها استخراج ویژگی‌های مؤثّر از تصاویر ورودی است. همان‌طورکه در شکل (۱) مشاهده می‌شود، هر لایهٔ کانولوشن از تعدادی صافی (یا کرنل) آموزش‌پذیر تشکیل شده‌است که عرض و ارتفاع مشخصی (بسیار کوچکتر از تصویر ورودی) دارند. این صافی‌ها یک‌به‌یک با تصویر ورودی کانوالو[1] می‌شوند و هریک از آن‌ها آرایه‌ای را از ویژگی‌های استخراج‌شده از تصویر می‌سازند که به آن نقشهٔ ویژگی[2] می‌گویند. به عبارت دیگر، هر صافی، در راستای عرض و ارتفاع تصویر ورودی با گام[3] مشخصی حرکت می‌کند و در هر بار توقف ضرب نقطه‌ای بین درایه‌های صافی و درایه‌های تصویر ورودی محاسبه می‌شود، که از کنار هم قرارگرفتن آن‌ها یک نقشهٔ ویژگی ساخته می‌شود. بنابراین، هر نقشهٔ ویژگی نشان‌دهندهٔ پاسخ منحصربه‌فرد یک صافی در موقعیت‌های مکانی مختلف است.

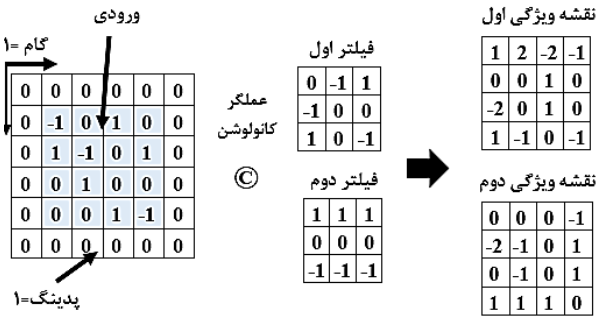
لایه‌های کانولوشن فراشاخصهایی مانند تعداد، ابعاد و ضرایب صافی‌ها و دو شاخص گام و پدینگ[4] دارند که تنظیم دقیق آن‌ها بر بهبود عملکرد شبکه مؤثّر است. به‌طوری‌که تعداد صافی‌ها در هر لایهٔ کانولوشن تعداد نقشه‌های ویژگی در خروجی لایه را مشخص می‌کند. تنظیم دقیق ابعاد صافی‌ها بر اساس ابعاد ویژگی‌های موردنظر در تصویر می‌تواند بر استخراج بهینه‌تر ویژگی‌ها مؤثّر باشد. ضرایب صافی‌ها که به‌طورمستقیم بر نوع ویژگی‌های استخراج‌شده از تصویر تأثیر می‌گذارند و در نهایت، گام میزان جابه‌جایی هر صافی بین دو کانوالو متوالی و پدینگ تعداد سطر و ستونی را که قبل‌از انجام کانولوشن به اطراف تصویر ورودی اضافه می‌شود، تعیین می‌کنند. بر این اساس، ابعاد هر نقشهٔ ویژگی در خروجی لایهٔ کانولوشن از معادلهٔ (۲) محاسبه می‌شود:

$$O = \left[\frac{X - W + 2P}{S}\right] + 1 \qquad (۲)$$

به‌طوری‌که O ابعاد نقشه ویژگی در خروجی لایهٔ کانولوشن، X ابعاد ورودی، P ابعاد پدینگ، S ابعاد گام و W ابعاد صافی‌های لایهٔ کانولوشن هستند. شکل (۳) عملکرد یک لایهٔ کانولوشن را در شبکهٔ کانولوشنال نشان می‌دهد.

[1] Convolve
[2] Feature Map
[3] Stride
[4] Padding

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-25]
[DOI: 10.61186/jsdp.20.3.127]



(شکل-۳): عملکرد لایهٔ کانولوشن در شبکهٔ کانولوشنال
(Figure-3): The function of convolution layer in CNN network

همان‌گونه‌که در شکل (۳) مشاهده می‌شود، شاخص‌های لایهٔ کانولوشن و به‌ویژه ضرایب صافی‌های این لایه نقش مهمی در استخراج مؤثّرتر ویژگی‌های تصویر و در نتیجه طبقه‌بندی دقیق‌تر تصاویر دارند. چراکه نقشه‌های ویژگی در خروجی لایهٔ کانولوشن در واقع، همان ویژگی‌های استخراج‌شده از تصویر ورودی هستند که پس‌از انجام عملیات کاهش بُعد به لایه‌های طبقه‌بندی شبکه ارسال می‌شوند. مطالعات ما نشان می‌دهند که در حالت مرسوم، ضرایب صافی‌های لایهٔ کانولوشن (همانند وزن‌ها و بایاس-های نورون‌ها در شبکه‌های عصبی) در شروع آموزش شبکه مقادیر تصادفی می‌گیرند و صافی‌هایی با ضرایب تصادفی را می‌سازند [۳۴]. اما، شروع آموزش شبکه با صافی‌های تصادفی می‌تواند کارآیی شبکهٔ کانولوشنال را کاهش دهد، چراکه این صافی‌ها نمی‌توانند تمامی ویژگی-های مهم در تصاویر ورودی را استخراج کنند. برای حل این مشکل در الگوی پیشنهادی از یک بانک فیلتر برای سازمان‌دهی صافی‌های لایهٔ کانولوشن استفاده شده‌است. در این بانک فیلتر، در شروع آموزش شبکه صافی‌هایی برای لایهٔ کانولوشن استفاده شده‌است که نخست، تمامی ویژگی‌های تصاویر ورودی که بر دقت طبقه‌بندی مؤثّرند، استخراج شوند، دوم، شباهت بین صافی‌ها به کمترین حد برسد و سوم، بسته به مسئلهٔ طبقه‌بندی ویژگی‌هایی اختصاصی‌تر و مؤثّرتر از تصاویر ورودی استخراج شود. برای روشن‌ترشدن موضوع، در ادامه نمونه‌هایی از صافی‌های تصادفی و صافی‌های به‌کاررفته در بانک فیلتر پیشنهادی به‌طور دقیق‌تر معرفی شده‌اند.

الف) صافی‌های تصادفی:

برای تولید صافی‌های تصادفی پس‌از تعیین تعداد و ابعاد صافی‌های هر لایهٔ کانولوشن در شبکهٔ کانولوشنال از دستورهای تولید اعداد تصادفی (مانند توابع rand یا randn در نرم‌افزار MATLAB) استفاده می‌شود. این دستورهای صافی‌هایی با ضرایب به‌طور کامل تصادفی تولید می‌کنند که در هر بار آموزش شبکه تغییر می‌کنند. شکل (۴) دو نمونه از صافی‌های تصادفی با ابعاد ۵×۵ را که به‌طور معمول در شروع آموزش شبکه‌های کانولوشنال استفاده می‌شوند، نشان می‌دهد.

-0.079	0.441	-0.982	0.541	-0.283
0.898	0.101	-0.093	0.389	-0.329
0.183	0.787	-0.435	0.751	0.902
0.291	-0.167	-0.168	0.778	-0.836
0.113	-0.854	-0.218	0.223	0.067

0.260	0.863	-0.234	0.126	-0.751
-0.613	-0.712	0.522	-0.431	0.052
0.326	0.089	0.712	-0.821	-0.744
-0.323	0.232	-0.188	-0.632	0.326
-0.442	0.456	-0.953	0.663	-0.016

(شکل-۴): نمونه‌ای از صافی‌های تصادفی با ابعاد ۵×۵
(Figure-4): Examples of random filters

همان‌طورکه در شکل (۴) مشاهده می‌شود، ضرایب صافی‌ها در این حالت به‌طور کامل تصادفی و بدون درنظرگرفتن هیچ پیش‌شرط خاصی تولید می‌شوند. این ضرایبِ تصادفی اگرچه در فرآیند آموزش شبکه به سمت ضرایب بهینه تغییر می‌کنند، در شروع آموزش هیچ ویژگی یا الگوی خاصی را از تصویر ورودی استخراج نمی-کنند. بنابراین، می‌توان گفت که انتخاب ضرایب تصادفی برای صافی‌های لایهٔ کانولوشن به دلایل زیر نامناسب است:

۱) ایجاد صافی‌هایی با ضرایب مشابه: درصورتی‌که دو یا چند صافی ضرایبی همگن از نظر مکانی داشته‌باشند، ویژگی‌هایی مشابه را از تصویر استخراج می‌کنند، که این امر عملکرد شبکه را کاهش می‌دهد.

۲) عدم‌پوشش تمامی ویژگی‌های مؤثّر: احتمال بسیار بالایی وجود دارد که صافی‌های تصادفی نتوانند تمامی ویژگی‌های مهم در تصویر را مانند لبه‌ها، شکل‌ها، خط‌ها، زاویه‌ها و ...، در تمامی جهت‌های ممکن استخراج کنند.

۳) ماهیت تصادفی: مقداردهی تصادفی به ضرایب صافی‌ها در لایه‌های کانولوشن سبب می‌شود تا الگوریتم در هر بار آموزش از نقطه‌ای متفاوت شروع به آموزش کرده و در نتیجه هم‌گرایی مناسبی در پاسخ‌ها وجود نداشته‌باشد. این امر در بسیاری از موارد سبب ایجاد پاسخ‌های پرت و دور از واقعیت می‌شود. ماهیت تصادفی ضرایب همچنین، امکان طراحی صافی‌ها بر اساس مسئلهٔ پیش‌رو را از بین می‌برد.

۴) افزایش هزینهٔ زمانی و محاسباتی شبکه: ضرایب ناکارآمد صافی‌ها و شروع آموزش از نقطه‌ای نامناسب به‌طورمستقیم بر هزینهٔ زمانی و محاسباتی شبکه تأثیر می‌گذارد و آن‌ها را به‌طور نامعقولی افزایش می‌دهد.

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-25]
[DOI: 10.61186/jsdp.20.3.127]

باتوجه‌به مطالب بیان‌شده، منطقی به نظر می‌رسد که آموزش شبکه از نقطه‌ای مناسب‌تر و توسط صافی‌هایی با ضرایب ازپیش‌تعیین‌شده شروع شود. به این منظور، در این مقاله از یک بانک فیلتر برای تخصیص صافی‌های کارآمد و خاص مسئله برای لایه‌های کانولوشن استفاده شده‌است.

ب) بانک فیلتر:

برای تولید یک بانک فیلتر می‌توان بر اساس ویژگی‌های موردنظر در تصویر از صافی‌های متنوعی استفاده کرد. به‌عنوان‌مثال، در مسئلهٔ تشخیص ارقام دست‌نویس لبه‌ها، گوشه‌ها و زوایای خطوط در جهت‌های مختلف ویژگی‌های مهمی هستند که در طبقه‌بندی و تمییز اعداد از یکدیگر کاربرد دارند. ازاین‌رو، می‌توان تأکید بیشتری بر صافی‌های لبه‌بردار و صافی‌های حساس به فرکانس‌های بالا داشت. شکل (۵) چند نمونه از صافی‌های طراحی‌شده با ابعاد ۵× ۵ را که در لایهٔ کانولوشن الگوی پیشنهادی استفاده شده‌است، نشان می‌دهد.

0.04	0.04	0.04	0.04	0.04
0.04	0.04	0.04	0.04	0.04
0.04	0.04	0.04	0.04	0.04
0.04	0.04	0.04	0.04	0.04
0.04	0.04	0.04	0.04	0.04

(1)

-0.04	-0.04	-0.04	-0.04	-0.04
-0.04	-0.04	-0.04	-0.04	-0.04
0.20	0.20	0.20	0.20	0.20
-0.04	-0.04	-0.04	-0.04	-0.04
-0.04	-0.04	-0.04	-0.04	-0.04

(2)

-0.04	-0.04	0.20	-0.04	-0.04
-0.04	-0.04	0.20	-0.04	-0.04
-0.04	-0.04	0.20	-0.04	-0.04
-0.04	-0.04	0.20	-0.04	-0.04
-0.04	-0.04	0.20	-0.04	-0.04

(3)

0.20	-0.04	-0.04	-0.04	-0.04
-0.04	0.20	-0.04	-0.04	-0.04
-0.04	-0.04	0.20	-0.04	-0.04
-0.04	-0.04	-0.04	0.20	-0.04
-0.04	-0.04	-0.04	-0.04	0.20

(4)

-0.04	-0.04	-0.04	-0.04	0.20
-0.04	-0.04	-0.04	0.20	-0.04
-0.04	-0.04	0.20	-0.04	-0.04
-0.04	0.20	-0.04	-0.04	-0.04
0.20	-0.04	-0.04	-0.04	-0.04

(5)

0.045	0.0467	0.056	0.047	0.045
0.047	0.317	0.715	0.317	0.047
0.056	0.715	-4.90	0.715	0.056
0.047	0.317	0.715	0.317	0.047
0.045	0.047	0.056	0.047	0.045

(6)

(شکل-۵): چند نمونه صافی طراحی‌شده با ابعاد ۵×۵ برای نخستین لایهٔ کانولوشن در الگوی پیشنهادی. (۱) صافی میانگین‌یاب، (۲) لبه‌یاب افقی، (۳) لبه‌یاب عمودی، (۴) لبه‌یاب ۴۵+، (۵) لبه‌یاب ۴۵-، (۶) صافی لگاریتمی.

(Figure-5): Examples of the designed filters for the first convolution layer of the proposed model

همان‌طورکه در شکل (۵) مشاهده می‌شود، ضرایب صافی‌ها در این حالت به‌طور کامل، هدفمند تعیین می‌شوند. به‌طوری‌که هر صافی بتواند ویژگی خاصی را از تصویر ورودی استخراج کند. در این شکل صافی‌های اول تا ششم به ترتیب میانگین تصویر، لبه‌های افقی، لبه‌های عمودی، لبه‌های مورّب در زوایای ۴۵± و لبه‌های دوجانبه را استخراج می‌کنند. اگرچه در حالت کلی، بانک فیلتر می‌تواند بر اساس مسئلهٔ طبقه‌بندی، محتوای تصاویر ورودی و ویژگی‌های موردنظر در آن‌ها از صافی‌هایی با ابعاد و محتوای متفاوتی تشکیل شود و با یک فراخوانی ساده به‌عنوان صافی‌های لایهٔ کانولوشن استفاده شود. در الگوی پیشنهادی ما از یک بانک فیلتر که دارای ۱۶ صافی ۵×۵ بود، استفاده کردیم.

۳-۲-۲ - تابع فعال‌سازی یا واحد یک‌سوساز خطی[1]

تابع فعال‌سازی یا واحد یک‌سوساز خطی در واقع خروجی یک نورون عصبی را در لایه‌های کانولوشن شبکهٔ کانولوشنال الگوسازی‌سازی می‌کند. این تابع فعال‌سازی به‌طور معمول، بلافاصله بعداز هر لایهٔ کانولوشن به‌کاربرده‌می‌شود و بر روی نقشه‌های ویژگی تاثیر می‌گذارد. این تابع ویژگیِ غیرخطی‌بودن را به شبکه اضافه می‌کند. مطالعات نشان داده‌است که روش استاندارد برای الگوسازی خروجی نورون‌ها در شبکه‌های عصبی مصنوعی از طریق معادلهٔ (۳) انجام می‌پذیرد.

$$f(x) = \tanh(x) \quad or \quad (1+e^{-x})^{-1} \qquad (۳)$$

در این رابطه x ورودی نورون و f(x) خروجی نورون است. اما استفاده از این رابطه در شبکه‌های کانولوشنال با حجم بالای نورون‌ها سبب کاهش چشمگیر سرعت شبکه می‌شود. بنابراین، در عمل به‌جای استفاده از این رابطه از واحد یکسوساز خطی استفاده می‌شود. معادلهٔ (۴) تابع تبدیل یک واحد یکسوساز خطی را نشان می‌دهد:

$$f(x) = \max(0, x) = \begin{cases} x & if \quad x > 0 \\ 0 & if \quad x \le 0 \end{cases} \qquad (۴)$$

باتوجه‌به معادلهٔ (۴) تابع فعال‌سازی یکسوساز خطی عملیاتی عنصر به عنصر را بر روی نقشه‌های ویژگی انجام می‌دهد و به‌جای هر پیکسل با مقدار منفی، پیکسلی با مقدار صفر را جایگزین می‌کند. این امر با توجه به نقش کمتر پیکسل‌های دارای شدت روشنایی منفی در تصویر قابل توضیح است. در الگوی پیشنهادی نیز در خروجی هر لایهٔ کانولوشن از یک تابع فعال‌سازی یکسوساز خطی استفاده شده‌است.

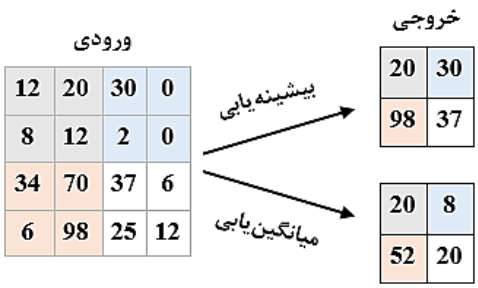
[1] Rectified Linear Unit (ReLU)

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-25]

[DOI: 10.61186/jsdp.20.3.127]

۲-۲-۴ - لایهٔ نمونه‌بردار کاهنده

لایهٔ نمونه‌بردار کاهنده وظیفهٔ انتخاب (کاهش) ویژگی را در شبکه‌های کانولوشنال بر عهده دارد و به‌طور معمول در معماری این شبکه‌ها پس از هر لایهٔ کانولوشن قرار می‌گیرد. ورودی این لایه نقشه‌های خام ویژگی و خروجی آن نقشه‌هایی با ابعاد کوچک‌تر از ویژگی‌های مؤثّرتر است. به عبارت دیگر، در این لایه عرض و ارتفاع فریم‌ها کاهش یافته و شاخص‌های شبکه و به‌دنبال آن حجم محاسبات شبکه، کمتر می‌شود تا از بیش‌برازش شبکه نیز جلوگیری شود. در این لایه یک پنجره با ابعاد مشخص از ورودی (نقشه ویژگی) جدا می‌شود و به کمک مقادیر داخل آن یک مقدار به‌عنوان خروجی انتخاب و یا محاسبه می‌شود. میانگین‌یابی و بیشینه‌یابی رایج‌ترین نوع عملیات نمونه‌برداری هستند. در میانگین‌یابی مقدار میانگین المان‌ها در پنجرهٔ جداشده محاسبه می‌شود و در بیشینه‌یابی مقدار بیشینه به‌عنوان خروجی درنظرگرفته‌می‌شود. عملیات نمونه‌برداری در لایهٔ نمونه‌بردار کاهنده همانند لایهٔ کانولوشن دارای شاخص‌های گام و پدینگ است و محاسبهٔ ابعاد خروجی آن نیز از معادلهٔ (۲) پیروی می‌کند. شکل (۵) عملکرد لایهٔ نمونه‌بردار کاهنده را در دو حالت میانگین‌یابی و بیشینه‌یابی مقایسه می‌کند.



(شکل-۶): مقایسهٔ عملکرد لایهٔ نمونه‌بردار کاهنده در دو حالت میانگین‌یابی و بیشینه‌یابی (گام =۲؛ پدینگ=۰؛ ابعاد پنجره= ۲×۲)

(Figure-6): Comparison of the Pooling layer performance in the two modes of averaging and maximizing

همان‌طورکه شکل (۶) نشان می‌دهد، ابعاد نقشه‌های ویژگی در عبور از لایهٔ نمونه‌بردار کاهنده کاهش می‌یابد. به‌عنوان‌مثال، با استفاده از صافی‌هایی با اندازهٔ ۲*۲ به‌همراه گام ۲، عرض و ارتفاع هر فریم ورودی نصف شده، و ۷۵ درصد مقادیر موجود در آن حذف می‌شود. در الگوی پیشنهادی پس از آزمایش‌های متعدد عملکرد لایهٔ نمونه‌بردار کاهنده در حالت بیشینه‌یابی تنظیم شده‌است.

۲-۱-۵ - لایه‌های طبقه‌بندی

لایه‌های طبقه‌بندی در معماری مرسوم شبکه‌های کانولوشنال از نوع لایه‌های به‌طور کامل پیوسته با نورون‌هایی دارای وزن‌ها و بایاس‌های آموزش‌پذیر هستند. به‌عبارت‌دیگر، نورون‌هایی که در یک لایهٔ به‌طور کامل پیوسته قرار دارند با تمام نورون‌های موجود در لایهٔ قبل‌از خود ارتباط دارند. لایه‌های به‌طور کامل متصل اغلب، بلافاصله پس‌از آخرین لایهٔ کانولوشن قرار می‌گیرند و هدف اصلی آن‌ها انجام طبقه‌بندی بر اساس ویژگی‌های استخراج‌شده از آخرین لایهٔ کانولوشن است. برای استفاده از این لایه‌ها در ساختار شبکهٔ کانولوشن نخست، خروجی آخرین لایهٔ کانولوشن در یک لایه به نام لایهٔ مسطح‌ساز[1] به یک بردار یک‌بعدی از ویژگی‌ها تبدیل شده و سپس این ویژگی‌ها برای طبقه‌بندی وارد لایه‌های به‌طورکامل پیوسته، می‌شوند. تعداد لایه‌های به‌طور کامل متصل و تعداد نورون‌های هرکدام جزء فراشاخص‌های شبکه هستند و بر اساس پیچیدگی مسئله و تعداد رده‌های خروجی تعیین می‌شوند. در الگوی پیشنهادی تعداد لایه‌های به‌طورکامل پیوسته متغیر درنظرگرفته‌شد و تعداد بهینهٔ این لایه‌ها و تعداد نورون‌های آن‌ها با انجام آزمایش مشخص شد.

۳- الگوریتم آموزش

آموزش شبکه‌های کانولوشنال به معنی تصحیح شاخص‌های آموزش‌پذیر شبکه بر اساس خطای محاسبه‌شده در خروجی است. در این شبکه‌ها آموزش به صورت نظارت‌شده[2] است و بر اساس زوج ورودی- خروجی صحیح انجام می‌شود. به این منظور، نخست، شاخص‌های شبکه (شامل ضرایب صافی‌های لایه‌های کانولوشن، وزن و بایاس‌های لایه‌های به‌طور کامل پیوسته و..) مقداردهی اولیه می‌شوند، سپس یک ورودی به شبکه اعمال و خروجی شبکه بر اساس وزن‌های اولیه محاسبه می‌شود و در نهایت، اختلاف خروجی به‌دست‌آمده با خروجی درست محاسبه، و این خطا به سمت عقب برگشته و شاخص‌ها را در جهت کاهش خطا اصلاح می‌کند. این فرایند آن‌قدر تکرار می‌شود تا خطای خروجی به کمترین مقدار خود برسد. پس از این فرایند شبکهٔ آموزش‌یافته به‌وسیلهٔ داده‌هایی که در فرایند آموزش حضور نداشته‌اند و به‌عنوان

[1] Flatten
[2] Supervised

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-25]
[DOI: 10.61186/jsdp.20.3.127]

داده‌های اعتبارسنجی کنار گذاشته شده‌بودند، اعتبارسنجی می‌شود. در الگوی پیشنهادی از الگوریتم پس‌انتشار خطا برای تصحیح شاخص‌های شبکه استفاده شده‌است.

۴- آزمایش‌ها

در پژوهش حاضر برای بررسی تأثیر ضرایب اولیهٔ صافی‌های لایهٔ کانولوشن بر عملکرد شبکه‌های کانولوشنال و ارزیابی عملکرد الگوی پیشنهادی مجموعه‌ای از آزمایش‌ها در دو گروه مجزا انجام شد. در گروه نخست آزمایش‌ها، یک شبکهٔ کانولوشنال ساده تنها با یک لایهٔ کانولوشن و در گروه دوم آزمایش‌ها یک شبکهٔ کانولوشنال با سه لایهٔ کانولوشن پیاده‌سازی و عملکرد آن‌ها در سه حالت مختلف صافی‌گذاری در لایهٔ کانولوشن (نخستین لایهٔ کانولوشن) مقایسه شد. سه نوع صافی‌گذاری عبارت بودند از: صافی‌های تصادفی، صافی‌های ثابت (با ضرایب یکسان) و صافی‌های ازپیش‌طراحی‌شده در یک بانک فیلتر. برای هر حالت از صافی‌گذاری شبکه پنجاه بار به‌وسیلهٔ ورودی‌هایی با ترتیب تصادفی آموزش داده‌شد و میانگین و بهترین دقت طبقه‌بندی شبکه ثبت شد. لایهٔ کانولوشن این شبکه‌ها ۱۶ صافی با ابعاد ۵×۵ داشت. صافی‌های تصادفی شامل اعداد تصادفی در بازهٔ [۱+،۱-] بودند و در هر بار اجرا به‌صورت تصادفی (با دستور randn) مقداردهی می‌شدند؛ صافی‌های ثابت شامل یک عدد ثابت در بازهٔ [۱+،۱-] بودند و در هر بار اجرا این مقدار تغییر می‌کرد؛ و در نهایت، صافی‌های به‌کاررفته در بانک فیلتر ترکیبی از صافی‌های لبه‌بردار مختلف در جهت‌های افقی، عمودی و ۴۵± درجه، صافی‌های میانگین‌گیر و غیره بود که امکان استخراج ویژگی‌های متنوع تصاویر ورودی را فراهم می‌کرد.

۴-۱- شبکهٔ کانولوشنال تک‌لایهٔ پیشنهادی

جدول (۱) جزئیات دقیق پیکربندی شبکهٔ کانولوشنال یک لایهٔ پیشنهادی را که پس‌از انجام آزمایش‌های متعدد تنظیم شده‌اند، نشان می‌دهد.

همان‌طورکه جدول (۱) نشان می‌دهد، الگوی پیشنهادی در این حالت دارای یک لایهٔ کانولوشن با ۱۶ صافی ۵×۵ و تابع فعال‌سازی یکسوساز خطی، یک لایهٔ نمونه‌بردار کاهنده با پنجره‌های ۳×۳ و گام ۲، دو لایهٔ به طور کامل پیوسته به‌ترتیب با صد و پنجاه نورون و در نهایت، لایهٔ خروجی با ده نورون و تابع فعال‌سازی بیشینه هموار[1] بود. جدول (۲) دقت طبقه‌بندی تصاویر پایگاه دادهٔ MNIST به‌وسیلهٔ این الگو با سه نوع صافی‌گذاری متفاوت در لایهٔ کانولوشن را نشان می‌دهد.

(جدول-۱): پیکربندی شبکهٔ کانولوشنال یک لایهٔ پیشنهادی
(Table-1): the configuration of the suggested single-layer CNN network

لایه‌ها	Size	N	S	P	A.F	O
input	-	-	-	-	-	۲۸×۲۸
Conv1	۵×۵	۱۶	۲×۲	۰	RELU	۱۳×۱۳×۱۶
Pooling	۳×۳	-	۲×۲	۰	Max	۶ ×۶
Flatten	-	-	-	-	-	۵۷۶ ×۱
FC1	-	۱۰۰	-	-	Relu	۱۰۰×۱
FC2	-	۵۰	-	-	Relu	۵۰ ×۱
Output	-	۱۰	-	-	Softmax	۱۰ ×۱

*** توضیحات:**
Size: ابعاد صافی‌ها؛ N: تعداد صافی‌ها و یا نورون‌ها؛ S: گام؛ P: پدینگ؛ A.F: تابع فعال‌سازی؛ O: ابعاد خروجی

(جدول-۲): مقایسهٔ عملکرد شبکهٔ کانولوشنال یک‌لایه با سه نوع صافی‌گذاری متفاوت در نخستین لایهٔ کانولوشن
(Table-2): Comparing the performance of a single-layer CNN network with three different types of filters in the first convolution layer

نوع صافی‌گذاری	میانگین دقت (٪)	بیشترین دقت (٪)
ثابت	۷۴/۹۴	۷۶/۷۲
تصادفی	۸۶/۴۷	۹۰/۱۶
بانک فیلتر	۹۱/۸۹	۹۳/۳۵

مقادیر به‌دست‌آمده در جدول (۲) نشان می‌دهند ضرایب اولیهٔ صافی‌های لایهٔ کانولوشن نقش مؤثّری بر دقت طبقه-بندی شبکهٔ کانولوشنال داشته‌اند. به‌طوری‌که میانگین دقت شبکه در حالتی که تمامی صافی‌ها ضرایب یکسان و ثابت داشته‌اند، به‌شدت پایین (۷۴/۹۴ درصد) بوده‌است. این نتایج همچنین، نشان می‌دهند که طراحی هدفمند صافی‌ها به‌وسیلهٔ یک بانک فیلتر در این لایه با استخراج ویژگی‌های مناسب‌تر و سمت‌وسودادن به الگوریتم آموزش، دقت طبقه‌بندی را افزایش داده‌است؛ به‌طوری‌که میانگین دقت طبقه‌بندی با صافی‌های تصادفی ۸۶/۴۷

[1] Softmax

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-25]
[DOI: 10.61186/jsdp.20.3.127]

درصد و با استفاده از صافی‌های به‌کارگرفته‌شده در بانک فیلتر ۹۱/۸۹ درصد بوده‌است. سپس برای تعیین اینکه آیا میانگین‌های بدست آمده از نظر آماری اختلاف معناداری داشته‌اند یا خیر، از آزمون t استفاده شد. نتایج این آزمون نشان داد که میانگین دقت مدل پیشنهادی در حالت بکارگیری بانک فیلتر در سطح معناداری ۰/۹۹ هم از فیلترهای ثابت و هم از فیلترهای تصادفی بیشتر بوده است. به عبارت دیگر، می‌توان گفت که بکارگیری بانک فیلتر در لایه‌ی کانولوشن شبکه‌ی تک لایه به طور معناداری سبب افزایش دقت طبقه‌بندی می‌شود.

۲-۴- شبکهٔ کانولوشنال سه‌لایهٔ پیشنهادی

در گروه دوم آزمایش‌ها، یک شبکهٔ کانولوشنال با سه لایهٔ کانولوشن پیاده‌سازی و همانند آزمایش‌ها قبل عملکرد آن در سه حالت مختلف صافی‌گذاری (در اولین لایهٔ کانولوشن شبکه) مقایسه شد. در مقالهٔ [۲۸] بهترین پیکربندی شبکهٔ کانولوشنال سه‌لایه در طبقه‌بندی ارقام دست‌نویس MNIST ارائه شده‌است. در آزمایش‌های این بخش از این پیکربندی استفاده و شبکهٔ کانولوشنال با همان مشخصات پیاده‌سازی شده‌است. جدول (۳) جزئیات دقیق پیکربندی شبکهٔ کانولوشنال سه‌لایهٔ پیشنهادی را نشان می‌دهد.

(جدول-۳): پیکربندی شبکهٔ کانولوشنال سه‌لایهٔ پیشنهادی
(Table-3): the configuration of suggested three-layer CNN network

لایه‌ها	Size	N	S	P	A.F	O
input	-	-	-	-	-	۲۸×۲۸
Conv1	۵×۵	۱۶	۱×۱	۲	RELU	۲۸×۲۸×۱۶
Pooling1	۳×۳	-	۲×۲	۱	Max	۱۴×۱۴
Conv2	۳×۳	۱۰	۱×۱	۱	RELU	۱۴×۱۴×۱۰
Pooling2	۳×۳	-	۲×۲	۱	Max	۷×۷
Conv3	۳×۳	۵	۱×۱	۱	RELU	۷×۷×۵
Pooling3	۳×۳	-	۲×۲	۱	Max	۴×۴
Flatten	-	-	-	-	-	۱۶ ×۱
FC1	-	۵۰	-	-	Relu	۵۰×۱
FC2	-	۲۵	-	-	Relu	۲۵ ×۱
Output	-	۱۰	-	-	Softmax	۱۰ ×۱

همان‌طورکه جدول (۳) نشان می‌دهد، الگوی پیشنهادی در این حالت دارای سه لایهٔ کانولوشن، سه لایهٔ نمونه‌بردار کاهنده، دو لایهٔ به‌طور کامل پیوسته و در نهایت، یک لایهٔ خروجی است. جدول (۴) دقت طبقه‌بندی تصاویر پایگاه دادهٔ MNIST به‌وسیلهٔ این الگو با سه نوع صافی‌گذاری متفاوت در لایهٔ کانولوشن را نشان می‌دهد.

(جدول-۴): مقایسهٔ عملکرد شبکهٔ کانولوشنال سه‌لایه با سه نوع صافی‌گذاری متفاوت در نخستین لایهٔ کانولوشن
(Table-4): Comparing the performance of a three - layer CNN network with three different types of filters in the first convolution layer

نوع صافی‌گذاری	میانگین دقت (٪)	بیشترین دقت (٪)
ثابت	۸۸/۸۲	۹۳/۶۹
تصادفی	۹۶/۱۶	۹۷/۸۸
بانک فیلتر	۹۹/۱۴	۹۹/۸۶

مقادیر به‌دست‌آمده در جدول (۴) نیز نشان می‌دهند که ضرایب اولیهٔ صافی‌ها در اولین لایهٔ کانولوشن شبکه‌های کانولوشنال چندلایه، نقش مؤثّری بر دقت طبقه‌بندی این شبکه‌ها داشته‌اند. باتوجه‌به این نتایج، میانگین دقت شبکه‌ای که به‌صورت هدفمند از بانک فیلتر استفاده می‌کند، ۹۹/۱۴ درصد و بالاترین دقت آن ۹۹/۸۶ درصد بوده‌است. سپس برای تعیین اینکه آیا میانگین‌های بدست آمده از نظر آماری اختلاف معناداری داشته‌اند یا خیر، بار دیگر از آزمون t استفاده شد. نتایج این آزمون نشان داد که میانگین دقت مدل پیشنهادی در حالت بکارگیری بانک فیلتر در سطح معناداری ۰/۹۵ هم از فیلترهای ثابت و هم از فیلترهای تصادفی بیشتر بوده است. به عبارت دیگر، می‌توان گفت که بکارگیری بانک فیلتر در اولین لایه‌ی کانولوشن شبکه‌های چند لایه نیز به طور معناداری سبب افزایش دقت طبقه‌بندی می‌شود.

۵- یافته‌ها و بحث

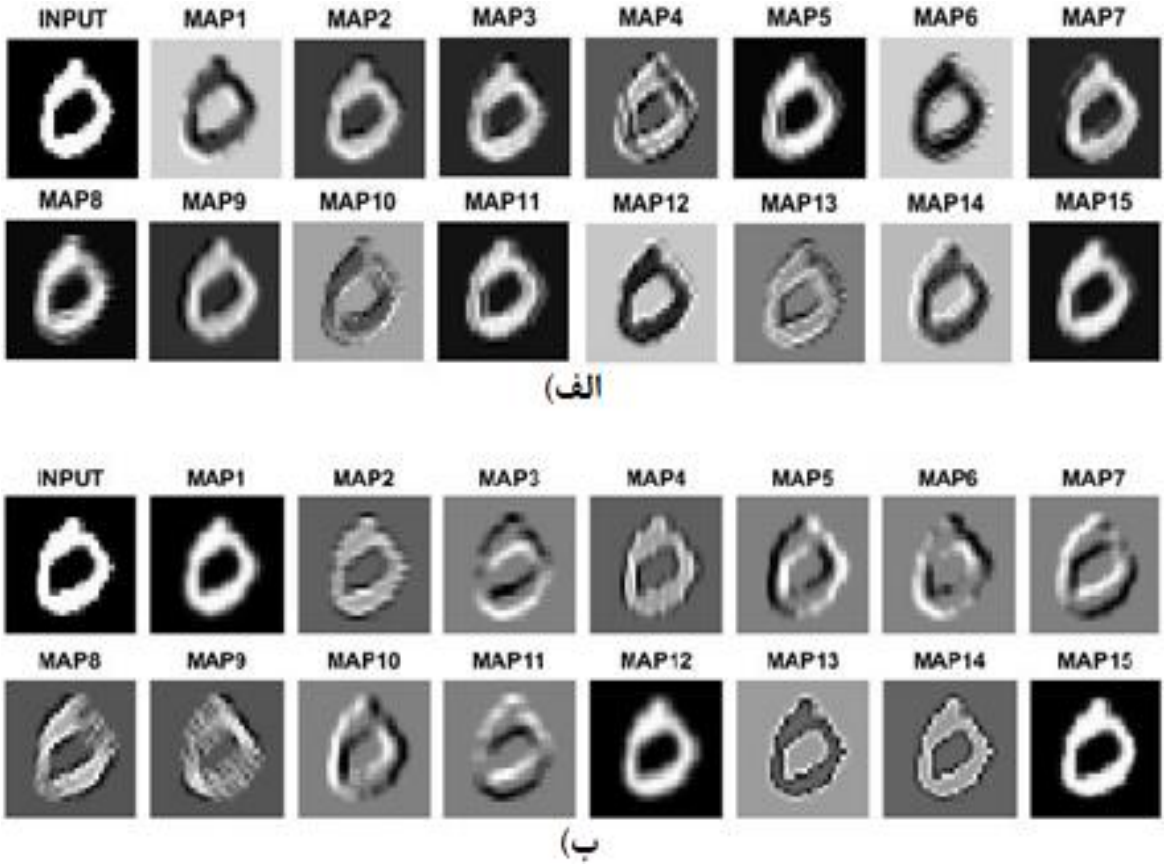
امروزه شبکه‌های کانولوشنال به‌طور روزافزون در کاربردهای مختلف بینایی ماشین و به‌ویژه در شناسایی و طبقه‌بندی تصاویر استفاده می‌شوند. این نوع از شبکه‌های عصبی مصنوعی با شبیه‌سازی عملکرد قشر بینایی مغز قدرتمندترین ساختار را در تجزیه و تحلیل داده‌های بصری دارند. اما گستردگی تصاویر دیجیتال و تنوع محتوا و ویژگی‌های موجود در آن‌ها مستلزم طرّاحی تخصصی و

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-25]
[DOI: 10.61186/jsdp.20.3.127]

دقیق این شبکه‌ها برای دستیابی به کارایی بالاتر است [۳۵]. در این راستا، در پژوهش حاضر تلاش شد تا دقت شبکه‌های کانولوشنال در طبقه‌بندی تصاویر با اختصاصی‌کردن صافی‌های لایهٔ کانولوشن به‌عنوان مهم‌ترین رکن استخراج ویژگی در این شبکه‌ها افزایش یابد. این کار از طریق طراحی مجموعه‌ای از صافی‌ها در قالب یک بانک فیلتر انجام شد.

آزمایش‌های ما بر روی شبکه‌های کانولوشنال با یک و سه لایهٔ کانولوشن نشان داد که به‌کارگیری صافی‌های تخصصی در نخستین لایهٔ کانولوشن شبکه سبب استخراج ویژگی‌های مؤثّرتر از تصاویر و افزایش قدرت و کارآیی شبکه در طبقه‌بندی تصاویر می‌شود. برای فهم بهتر چگونگی عملکرد بانک فیلتر و مقایسهٔ آن با صافی‌های تصادفی، خروجی اولین لایهٔ کانولوشن شبکه در شکل (۷) و خروجی آن پس‌از اعمال تابع فعال‌سازی یکسوساز خطی در شکل (۸) نشان داده شده‌است. این تصاویر نتایج حاصل از کانوالوشدن ۱۵ صافی غیرتصادفی در بانک فیلتر و ۱۵ صافی تصادفی با تصویر ورودی بوده و نقشه‌های ویژگی استخراج‌شده در اولین لایهٔ کانولوشن را نشان می‌دهند.

مقایسهٔ نقشه‌های ویژگی استخراج‌شده در شکل‌های (۷) و (۸) نشان می‌دهد که صافی‌های مورداستفاده در بانک فیلتر نسبت‌به صافی‌های تصادفی توانسته‌اند طیف وسیع‌تری از ویژگی‌های مهم را استخراج و در مجموع ویژگی‌های مؤثّرتر و معنادارتری از تصویر ورودی را استخراج کنند. به‌عنوان‌مثال، نقشه‌های ویژگی یک، شش و ده، که به‌وسیلهٔ صافی‌های تصادفی ایجاد شده‌اند، توانایی زیادی در استخراج ویژگی‌های تصویر ورودی نداشته‌اند. مقایسهٔ این تصاویر همچنین، نشان می‌دهد که تنوع ویژگی‌های استخراج‌شده به‌وسیلهٔ صافی‌های موجود در بانک فیلتر بسیار بیشتر از صافی‌های تصادفی بوده‌است. به‌طوری‌که برخی از این صافی‌ها ویژگی‌های فرکانس پایین و برخی دیگر ویژگی‌های فرکانس بالا را در تصویر به‌خوبی استخراج کرده‌اند.

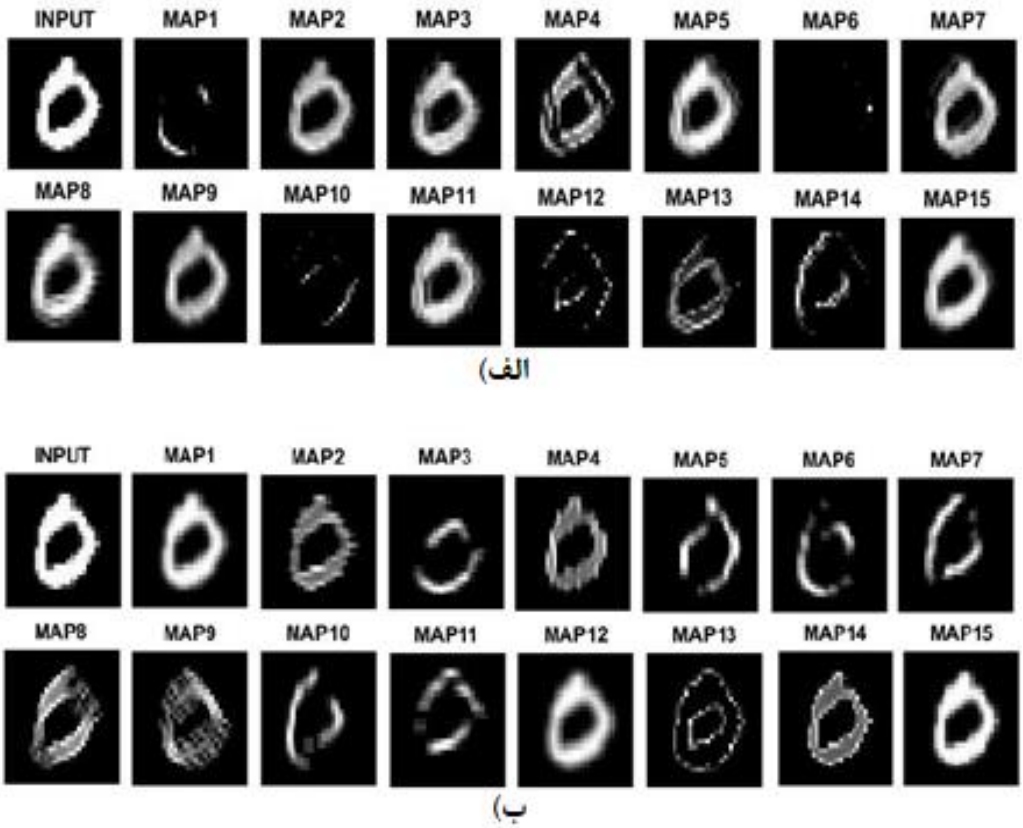


(شکل-۷): مقایسهٔ خروجی نخستین لایهٔ کانولوشن (نقشه‌های ویژگی). الف) صافی‌های تصادفی، ب) بانک فیلتر

(Figure-7): Comparison of the output of the first convolution layer (feature maps). A) Random filters, b) Filter bank

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-25]

[DOI: 10.61186/jsdp.20.3.127]



(شکل-۸): مقایسهٔ خروجی اولین لایهٔ کانولوشن پس از اعمال تابع فعال‌سازی یکسوساز خطی. الف) صافی‌های تصادفی، ب) بانک فیلتر

(Figure-8): Comparison of the output of the first RELU layer. A) Random filters, b) Filter bank

۱-۵- مقایسه با سایر روش‌ها

مطالعهٔ پیشینهٔ پژوهش نشان می‌دهد که بهبود عملکرد شبکه‌های کانولوشنال برای تشخیص و طبقه‌بندی تصاویر اعداد و یا نویسه‌های دست‌نویس موردتوجه بسیاری از پژوهش‌گران بوده‌است. اما بااین‌وجود، به‌کارگیری صافی‌های ازقبل‌طراحی‌شده برای لایه‌های کانولوشن شبکه در هیچ پژوهشی بررسی نشده‌است. تنها پژوهش نزدیک به مطالعهٔ حاضر توسط Calderon و همکارانش انجام شده‌است [۳۱]. آن‌ها پیشنهاد استفاده از صافی‌های گابور را در اولین لایهٔ کانولوشن شبکه‌های کانولوشنال مطرح کرده و نشان داده‌اند که این صافی‌ها بر دقت طبقه-بندی شبکه مؤثّرند. به‌منظور نشان‌دادن عملکرد روش پیشنهادی و اثبات کارآیی شبکه‌های کانولوشنال مبتنی‌بر بانک فیلتر، علاوه‌بر آزمایش‌هایی که در بخش قبل انجام شد، دقت روش پیشنهادی با برخی از پژوهش‌های گذشته، که از شبکه‌های کانولوشنال برای تشخیص اعداد دست‌نویس پایگاه دادهٔ MNIST استفاده کرده‌بودند، مقایسه شد. در جدول (۵) دقت شبکهٔ کانولوشنال سه‌لایهٔ پیشنهادی که در آن از بانک فیلتر استفاده شده‌است، با برخی از کارهای گذشته در این زمینه مقایسه شده‌است.

(جدول-۵): مقایسه‌ی عملکرد روش پیشنهادی با برخی از مطالعات گذشته

(Table-5): Comparing the performance of the proposed method with some past studies

پژوهش	روش	پایگاه داده	دقت (٪)
LeCun و همکارانش [۲۳]	CNN	Self-Created	۹۹
Simard و همکارانش [۲۴]	CNN MLP	MNIST	۹۹/۶ ۹۹/۳
Saqib و همکارانش [۳۰]	CNN+ELM	MNIST	۹۹/۸
Calderon و همکارانش [۳۱]	CNN + Gabor filters	MNIST	۹۹/۳۲
Le و همکارانش [۳۲]	CNN+SGD	MNIST	۹۹/۳۱
پژوهش حاضر	CNN + Filter Bank	MNIST	۹۹/۱۴

همانطور که در جدول (۵) مشاهده می‌شود، دقت کلاس-بندی تصاویر اعداد دست‌نویس در روش پیشنهادی (شبکه

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-25]
[DOI: 10.61186/jsdp.20.3.127]

کانولوشنال ۳ لایه به همراه بانک فیلتر) بسیار نزدیک به سایر روش‌های معرفی‌شده در مطالعات اخیر بوده است. با این تفاوت که در روش پیشنهادی تعداد لایه‌ها و حجم محاسباتی بسیار کمتر بوده و شبکه پیشنهادی تنها با سه لایهٔ کانولوشن و به کمک بانک فیلتر به دقت قابل قبولی رسیده است.

۶- نتیجه‌گیری

در این مقاله روشی جدید برای بهبود عملکرد شبکه‌های کانولوشنال در طبقه‌بندی تصاویر ارائه شد. در الگوی پیشنهادی، باتوجه‌به نقش اساسی لایه‌های کانولوشن (و به‌ویژه ضرایب صافی‌های آن) در استخراج ویژگی‌های مؤثّر از تصاویر و تأثیرات مستقیم آن‌ها بر عملکرد و دقت شبکه، از یک بانک فیلتر استفاده شد. بانک فیلتر امکان طراحی صافی‌های لایهٔ کانولوشن را به‌صورت خاص مسئله و مبتنی‌بر ویژگی‌های تصاویر ورودی فراهم می‌آورد. در این روش به‌جای وزن‌دهی تصادفی به ضرایب صافی‌های لایهٔ کانولوشن در آغاز آموزش شبکه، ضرایب صافی‌ها به‌طور دقیق‌تر، متنوع‌تر و خاص مسئلهٔ مقداردهی شدند. استفاده از بانک فیلتر پیچیدگی‌های اضافی به شبکه تحمیل نمی‌کند، با شروع آموزش شبکه از نقطهٔ بهینه مدت‌زمان آموزش را کاهش می‌دهد، کارآیی لایهٔ کانولوشن را با به کمترین حد رساندن ویژگی‌های مشابه و افزایش ویژگی‌های مؤثّر افزایش می‌دهد و در نهایت امکان طراحی لایه‌های کانولوشن شبکه را برای استخراج ویژگی‌های مبتنی‌بر محتوای تصویر فراهم می‌کند، و در نتیجه، دقت طبقه‌بندی را افزایش می‌دهد. در این پژوهش بررسی‌های انجام‌شده بر روی تصاویر پایگاه دادهٔ اعداد دست‌نویس MNIST نشان داد که استفاده از بانک فیلتر به‌طور قابل‌ملاحظه‌ای دقت شبکهٔ کانولوشنال را در طبقه‌بندی تصاویر افزایش می‌دهد. به‌طوری‌که هم برای شبکه‌های کانولوشنال تک‌لایه و هم برای شبکه‌هایی با تعداد بیشتری لایه‌های کانولوشن، ضرایب اولیهٔ صافی‌ها در اولین لایهٔ کانولوشن بر دقت طبقه‌بندی تاثیرگذار بود. آزمایش‌ها بر روی شبکهٔ کانولوشنال تک‌لایه با سه نوع صافی‌گذاری، صافی‌های مقدار ثابت، تصادفی و بانک فیلتر میانگین دقت طبقه‌بندی را در ۵۰ بار آموزش شبکه به ترتیب ۷۴/۹۴، ۸۶/۴۷ و ۹۱/۸۹ درصد و برای شبکهٔ کانولوشنال سه‌لایه به ترتیب ۸۸/۸۲، ۹۶/۱۶ و ۹۹/۱۴ درصد نشان دادند. این نتایج نشان می‌دهد که

۱) ضرایب اولیهٔ صافی‌های لایهٔ کانولوشن بر دقت طبقه‌بندی شبکه مؤثّر است؛

۲) با به‌کارگیری صافی‌های مؤثّرتر در لایهٔ کانولوشن می‌توان شبکهٔ کانولوشنال را خاص مسئله ساخته و از این طریق کارآیی شبکه را افزایش داد. بنابراین، پیشنهاد می‌شود تا به‌جای استفاده از ضرایب تصادفی برای صافی‌های لایه‌های کانولوشن در شروع آموزش شبکه‌های کانولوشنال، این ضرایب برای استخراج مؤثّرتر ویژگی‌های موردنظر در تصویر طراحی و مقداردهی شوند.

۷- مراجع 7-Refrence

[1] Javidi, B., "Image recognition and classification: algorithms, systems, and applications". 2002: CRC press.

[2] Lu D, Weng Q. "A survey of image classification methods and techniques for improving classification performance". International journal of Remote sensing. 2007;28(5):823-70.

[3] Nath, S.S ,.et al. "A survey of image classification methods and techniques". International conference on control, instrumentation, communication and computational technologies (ICCICCT). 2014, IEEE.

[4] Dat, P.K.K.O., "Image Classification Difficulties".

[5] Jain, G. and J. Ko, "Handwritten digits recognition". Multimedia Systems, Project Report, University of Toronto, 2008: p. 1-3.

[6] Jain, A.K., R.P.W. Duin, and J. Mao, "Statistical pattern recognition: A review". IEEE Transactions on pattern analysis and machine intelligence, 2000. 22(1): p. 4-37.

[7] Egmont-Petersen, M., D. de Ridder, and H. Handels, "Image processing with neural networks: a review". Pattern recognition, 2002. 35(10): p. 2279-2301.

[8] Ghosh, M.M.A. and A.Y. Maghari. "A comparative study on handwriting digit recognition using neural networks". in 2017 international conference on promising electronic technologies (ICPET). 2017. IEEE.

[9] Bala, R. and D. Kumar, "Classification using ANN: A review". Int .J. Comput. Intell. Res, 2017. 13(7): p. 1811-1820.

[10] Bhatnagar, S., D. Ghosal, and M.H. Kolekar. "Classification of fashion article images using convolutional neural networks". in 2017 Fourth International Conference on Image Information Processing (ICIIP). 2017 IEEE.

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-25]

[DOI: 10.61186/jsdp.20.3.127]

[11] Chen, F., et al., "Assessing four neural networks on handwritten digit recognition dataset (MNIST)". arXiv preprint arXiv:1811.08278, 2018.
[12] Hubel, D.H. and T.N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex". The Journal of physiology, 1962. 160(1): p. 106.
[13] Fukushima, K., "Neural network model for selective attention in visual pattern recognition and associative recall". Applied Optics, 1987. 26(23): p. 4985-4992.
[14] Momeny M, Sarram M A, Latif A, Sheikhpour R. A Convolutional Neural Network based on Adaptive Pooling for Classification of Noisy Images. JSDP 2021; 17 (4) :139-154.
[15] Traore, B.B., B. Kamsu-Foguem, and F. Tangara, "Deep convolution neural network for image recognition". Ecological Informatics, 2018. 48: p. 257-268.
[16] Duan, M., G. Wang, and C. Niu, "Method of small sample size image recognition based on convolution neural network". Computer Engineering and Design, 2018. 39(1): p. 224-229.
[17] Zhou, X., et al. "Application of deep learning in object detection". in 2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS). 2017. IEEE.
[18] Ren, S., et al., "Faster r-cnn: Towards real-time object detection with region proposal networks". Advances in neural information processing systems, 2015. 28.
[19] Liu, Z., et al., "Salient object detection for RGB-D image by single stream recurrent convolution neural network". Neurocomputing, 2019. 363: p. 46-57.
[20] Lang, R., L. Zhao, and K. Jia. "Brain tumor image segmentation based on convolution neural network". in 2016 9th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI). 2016. IEEE.
[21] Sultana, F., A. Sufian, and P. Dutta, "Evolution of image segmentation using deep convolutional neural network: a survey". Knowledge-Based Systems, 2020. 201: p. 106062.
[22] Al-Saffar, A.A.M., H. Tao, and M.A. Talab. "Review of deep convolution neural network in image classification". in 2017 International conference on radar, antenna, microwave, electronics, and telecommunications (ICRAMET). 2017. IEEE.
[23] LeCun, Y., et al., "Gradient-based learning applied to document recognition". Proceedings of the IEEE, 1998. 86(11): p. 2278-2324.
[24] Simard, P.Y., D. Steinkraus, and J.C. Platt. "Best practices for convolutional neural networks applied to visual document analysis". in Icdar. 2003.
[25] Nielsen, M.A., "Neural networks and deep learning". Vol. 25. 2015: Determination press San Francisco, CA, USA.
[26] Wan, L., et al. "Regularization of neural networks using dropconnect". in International conference on machine learning. 2013. PMLR.
[27] Tabik, S., et al., "A snapshot of image pre-processing for convolutional neural networks: case study of MNIST". 2017.
[28] Ahlawat, S., et al., "Improved handwritten digit recognition using convolutional neural networks (CNN)". Sensors, 2020. 20(12): p. 3344.
[29] Ahlawat, S. and A. Choudhary, "Hybrid CNN-SVM classifier for handwritten digit recognition". Procedia Computer Science, 2020. 167: p.2554-2560.
[30] Ali, S., et al., "An effective and improved CNN-ELM classifier for handwritten digits recognition and classification". Symmetry, 2020. 12(10): p. 1742.
[31] Calderon, A., S. Roa, and J. Victorino, "Handwritten digit recognition using convolutional neural networks and gabor filters". Proc. Int. Congr. Comput. Intell, 2003: p. 1-9.
[32] Le, Q.V., et al. "On optimization methods for deep learning". in ICML. 2011.
[33] LeCun, Y., C. Cortes, and C.J. Burges, "The MNIST database of handwritten digits". Website. 1998.
[34] Yamashita, R., et al., "Convolutional neural networks: an overview and application in radiology". Insights into imaging, 2018. 9(4): p. 611-629.
[35] Cui, H. and J. Bai, "A new hyperparameters optimization method for convolutional neural networks". Pattern Recognition Letters, 2019. 125: p. 828-834.



محمدرضا کرمی ملایی، کارشناسی خود را در رشتهٔ مهندسی برق الکترونیک از دانشگاه فردوسی مشهد، کارشناسی ارشد خود را در رشتهٔ پردازش سیگنال از دانشگاه گرونوبل فرانسه و دکتری خود را در رشتهٔ مهندسی پزشکی-بیوالکتریک از دانشگاه نانسی فرانسه دریافت کرده و هم‌اکنون دانشیار گروه مهندسی پزشکی، دانشکدهٔ مهندسی برق و کامپیوتر، دانشگاه صنعتی نوشیروانی بابل است.

نشانی رایانامهٔ ایشان عبارت است از:

mkarami@nit.ac.ir



محمدرضا حسن‌زاده، مدارک کارشناسی تا دکتری خود را در رشتهٔ مهندسی برق الکترونیک به‌ترتیب از دانشگاه‌های صنعتی شریف، صنعتی

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-25]

[DOI: 10.61186/jsdp.20.3.127]

امیرکبیر و دانشگاه مازندران دریافت کرده و هم اکنون استادیار گروه مهندسی برق الکترونیک، دانشکده مهندسی برق و کامپیوتر، دانشگاه صنعتی نوشیروانی بابل است.
نشانی رایانامهٔ ایشان عبارت است از:
m.hasanzadeh@nit.ac.ir



علی دروگرمقدم، کارشناسی ارشد خود را در رشتهٔ مهندسی پزشکی- بیوالکتریک از دانشگاه حکیم سبزواری در سال ۱۳۹۲ دریافت کرده و هم‌اکنون در مرحلهٔ دفاع از رسالهٔ دکترای خود با موضوع پیش‌بینی حمله‌های صرعی از روی سیگنال الکتروانسفالوگرام به کمک شبکه‌های عصبی کانولوشنال در همین رشته در دانشگاه صنعتی نوشیروانی بابل است. وی به‌مدت سه سال عضو هیأت علمی (طرح سربازی) پردیس فناوری‌های نوین دانشگاه حکیم سبزواری بوده‌است.
نشانی رایانامهٔ ایشان عبارت است از:
ali_derogarmoghadam@stu.nit.ac.ir

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-25]

[DOI: 10.61186/jsdp.20.3.127]