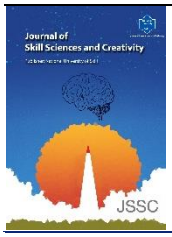


Journal of Skill Sciences and Creativity

National University of Skill

Summer 2024, Vol. 1, No. 2, p. 45-68

Journal Homepage: https://jssc.tvu.ac.ir/?lang=en

doi:10.48301/JSSC.2024.460608.1017

Handwritten Digit Recognition on MNIST Using Transfer Learning with VGG16

Kazem Taghandiki[1*]

[1]Faculty Member, Department of Computer Engineering, National University of Skills (NUS), Tehran, Iran.

ARTICLE INFO

Article Type:
Original Research

Received: 06.09.2024
Revised: 09.09.2024
Accepted: 11.11.2024

Keyword:
Handwritten Digit Recognition
MNIST Dataset
ImageNet Dataset
Deep Learning
VGG16 Model

***Corresponding Author:**
Kazem Taghandiki
Email: *ktaghandiki@tvu.ac.ir*

ABSTRACT

Handwritten digit recognition using the MNIST dataset is one of the fundamental problems in the field of deep learning and computer vision. In this study, the VGG16 transfer learning model was employed for recognizing handwritten digits. This model, which was previously trained on the ImageNet dataset, was retrained to adapt to the MNIST dataset. The performance of this model was evaluated using metrics such as accuracy, precision, recall, and F1 score, and the results were compared with other deep learning algorithms, including convolutional neural networks (CNNs), multilayer perceptrons (MLPs), and traditional machine learning algorithms. The results indicated that the VGG16 model, utilizing transfer learning, achieved an accuracy of 99% in recognizing handwritten digits, which is higher than that of models trained from scratch. Therefore, the use of pre-trained models can enhance the performance of deep learning models in handwritten digit recognition while reducing the required training time and computational resources.

E-ISSN: 3060-6691

EXTENDED ABSTRACT

Introduction

Handwritten digit recognition is a key application in computer vision, often evaluated using the MNIST dataset, a benchmark set containing thousands of labelled handwritten digits. Despite advancements in deep learning, achieving high accuracy on MNIST is challenging, especially for systems with limited resources. Convolutional Neural Networks (CNNs) excel at extracting features from complex images, making them highly effective for classification tasks. However, training CNNs from scratch can be resource-intensive.

Transfer learning provides an efficient alternative, allowing models pre-trained on large datasets to be fine-tuned for specific tasks. VGG16, a widely-used deep CNN pre-trained on ImageNet, is a strong candidate for transfer learning due to its robust feature extraction capabilities. This study evaluated VGG16's effectiveness in handwritten digit recognition on MNIST, comparing its performance to CNN, Multi-Layer Perceptron (MLP), and traditional machine learning methods like Naive Bayes, Decision Tree, and Support Vector Machine (SVM). The goal was to determine whether VGG16 can improve accuracy and efficiency for MNIST digit classification.

Methodology

The methodology used in the present study was implemented in Python 3.12.3, involving multiple distinct stages, each contributing to the overall model training, evaluation, and comparison process. The methodology follows these steps:

1- **Library Importation:** Key libraries, including NumPy, TensorFlow, and scikit-learn, are imported. NumPy is used for matrix operations, TensorFlow for building and training deep learning models, and scikit-learn for running traditional machine learning algorithms. This diverse toolkit allows for consistent pre-processing, efficient model training, and comprehensive evaluations.

2- **Data Loading and Preprocessing:** The MNIST dataset is loaded directly via TensorFlow. The dataset is split into training and testing sets, allowing for independent evaluations of unseen data. The images are normalized by scaling pixel values between 0 and 1 and reshaped to meet the input requirements of deep learning models. This normalization enhances model convergence during training while reshaping ensures compatibility across different model architectures.

3- **Label Encoding:** The multi-class nature of the MNIST dataset necessitates converting the digit labels to one-hot encoded vectors. This transformation is crucial for models designed to handle multi-class classification, as it enables them to predict the correct class probability.

4- **Model Training and Evaluation:**

- **VGG16:** The VGG16 model, pre-trained on ImageNet, is fine-tuned for the MNIST dataset through transfer learning. This is achieved by freezing the lower layers, which contain learned feature extraction capabilities, and training the higher layers on MNIST data.

- **CNN and MLP:** Additional custom CNN and MLP models are trained on MNIST for comparative analysis. The CNN comprises convolutional and pooling layers to capture spatial features, while the MLP leverages dense layers to identify patterns within the data.

5- **Traditional Classifiers:** The MNIST data is reshaped to fit the requirements of traditional classifiers, including Naive Bayes, Decision Tree, and SVM. While these models lack the hierarchical feature extraction of CNNs, they still provide valuable insights into the efficacy of traditional classification methods on MNIST data.

6- **Performance Evaluation:** A custom evaluation function calculates metrics such as accuracy, precision, recall, and F1-score for each model. These metrics enable a thorough assessment of each model's predictive power, ability to generalize, and performance consistency across classes.

Results and discussion

The models exhibited varying levels of accuracy and reliability. Results were presented in terms of confusion matrices and comparative performance plots, with specific figures for reference:

- **Confusion Matrices (Figures 1, 2, and 3):** Confusion matrices illustrated each model's performance across individual digit classes. VGG16's matrix (Figure 2) demonstrated the highest consistency in accurate classification, followed by CNN (Figure 1) and MLP (Figure 3). These matrices revealed how each model performs on a class-by-class basis, highlighting VGG16's strong predictive accuracy.

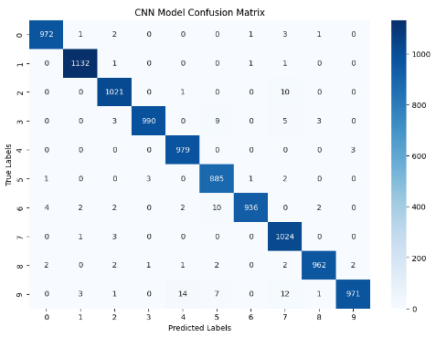


Figure 1. Confusion matrix generated by the CNN algorithm.

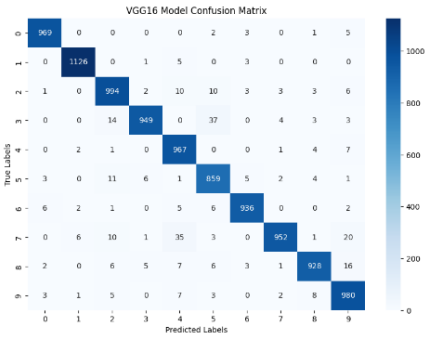


Figure 2. Confusion matrix created by the VGG16 algorithm.

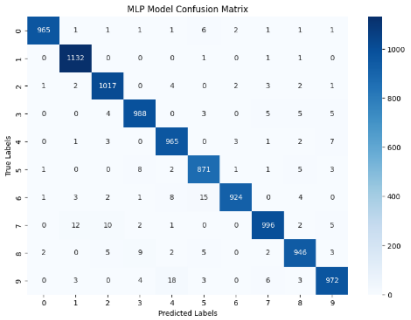


Figure 3. Confusion matrix created by the MLP algorithm.

– **Performance Comparison (Figure 4):** The performance of each model was quantified using accuracy, precision, recall, and F1-score. VGG16 achieved the highest accuracy of 99%, with CNN and MLP following closely at 98%. Traditional classifiers yielded mixed results; SVM reached 97% accuracy, Decision Tree achieved 89%, and Naive Bayes lagged at 55%. The precision, recall, and F1-score metrics further corroborated VGG16's dominance, as it consistently outperformed others in both precision and recall.

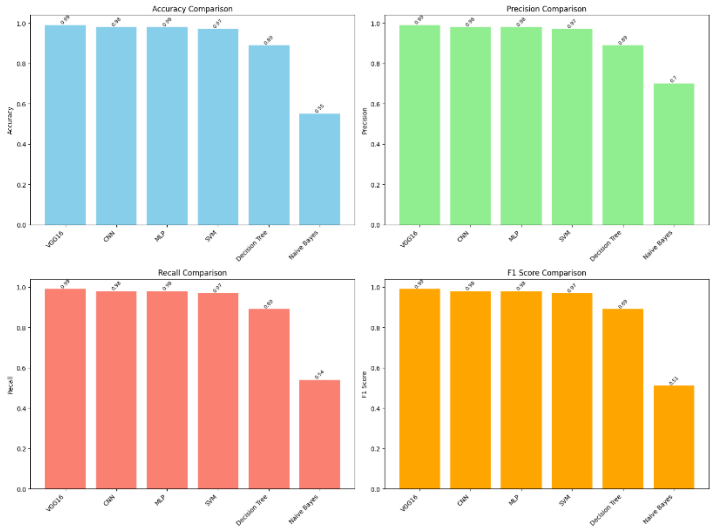


Figure 4. Performance metrics including Accuracy, Precision, Recall, and F1 Score for deep learning algorithms, traditional machine learning, and the VGG16 model.

– **ROC Curves (Figure 5):** The ROC curves highlighted the overall predictive capability of each model. VGG16 attained an area under the curve (AUC) of 0.98, showcasing its reliability in distinguishing between digit classes. CNN and MLP also performed well with AUCs of 0.96. Among traditional classifiers, SVM had an AUC of 0.92, followed by Decision Tree and Naive Bayes at 0.88 and 0.55, respectively.

The ROC curves underscore VGG16's robustness, which is attributed to the combination of deep feature extraction layers and fine-tuning.

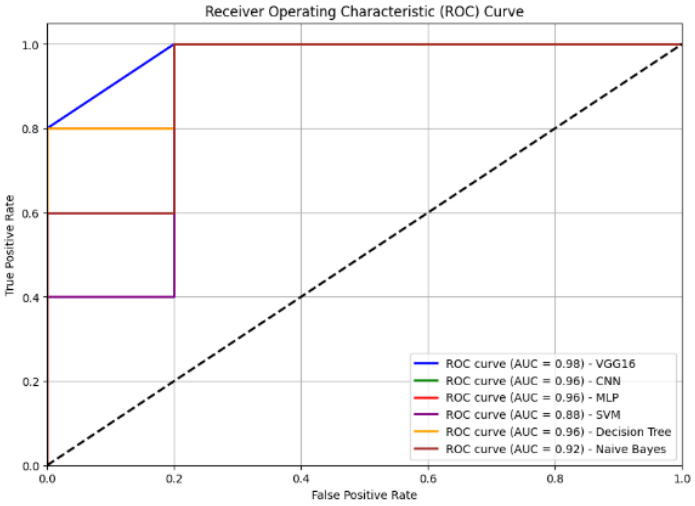


Figure 5. Performance metrics including Accuracy, Precision, Recall, and F1 Score for deep learning algorithms, traditional machine learning, and the VGG16 model.

Conclusion

The findings of this study reinforce the value of CNN architectures, specifically VGG16 with transfer learning, in handwritten digit recognition tasks. VGG16's pre-trained layers, designed for complex image classification on ImageNet, offer a significant advantage for simpler tasks like MNIST classification, as they reduce training time and computational requirements. This aligns with prior advancements in CNN research. Early architectures, such as LeNet-5, introduced the foundation for CNN-based digit recognition, while more complex designs like AlexNet, ZFNet, and VGGNet have progressively enhanced CNN effectiveness through deeper architectures and sophisticated optimization techniques.

Transfer learning with VGG16 not only simplifies the model development process but also improves computational efficiency by leveraging previously acquired knowledge. This approach aligns well with practical applications where data and computational resources may be limited. By adapting pre-trained models, researchers can achieve high accuracy on specific tasks without the need for large datasets or extensive training, thereby making deep learning more accessible across various domains.

نشریه علوم مهارتی و خلاقیت
دانشگاه ملی مهارت
تابستان ۱۴۰۳، دوره ۱، شماره ۲، ۶۸-۴۵
آدرس نشریه: https://jssc.tvu.ac.ir/
doi:10.48301/JSSC.2024.460608.1017

تشخیص اعداد دست‌نویس MNIST با استفاده از شبکه ترنسفری VGG16

کاظم تقندیکی[1*]

۱- عضو هیأت علمی گروه مهندسی کامپیوتر، دانشگاه ملی مهارت، تهران، ایران

اطلاعات مقاله

نوع مقاله: مقاله پژوهشی

دریافت مقاله: ۱۴۰۳/۰۳/۲۰
بازنگری مقاله: ۱۴۰۳/۰۶/۱۹
پذیرش مقاله: ۱۴۰۳/۰۸/۲۱

کلید واژگان:
تشخیص اعداد
مجموعه داده MNIST
مجموعه داده ImageNet
یادگیری عمیق
مدل VGG16

***نویسنده مسئول:** کاظم تقندیکی
پست الکترونیکی:
ktaghandiki@tvu.ac.ir

چکیده

تشخیص اعداد دست‌نویس با استفاده از مجموعه داده MNIST ازجمله مسائل اساسی در زمینه یادگیری عمیق و بینایی کامپیوتری است. در این تحقیق، از مدل یادگیری انتقالی VGG16 برای تشخیص اعداد دست‌نویس استفاده شده است. این مدل که قبلاً بر روی مجموعه داده ImageNet آموزش دیده بود، دوباره به‌منظور سازگاری با مجموعه داده MNIST آموزش داده شد. عملکرد این مدل با استفاده از معیارهای Accuracy، Precision، Recall و F1 score ارزیابی شد و نتایج آن با سایر الگوریتم‌های یادگیری عمیق مانند شبکه‌های عصبی پیچشی، چندلایه‌ای‌های چندگانه و الگوریتم‌های یادگیری ماشین سنتی مقایسه شد. نتایج نشان داد که مدل VGG16 با استفاده از یادگیری انتقالی، دارای دقت (Accuracy) ۹۹ درصد در تشخیص اعداد دست‌نویس می‌باشد که نسبت به مدل‌های آموزش‌دیده از ابتدا، دقت بالاتری دارد. از این‌رو استفاده از مدل‌های پیش‌آموزش‌شده می‌تواند عملکرد مدل‌های یادگیری عمیق را برای تشخیص اعداد دست‌نویس بهبود بخشد، در حالی که زمان آموزش و منابع محاسباتی موردنیاز را کاهش می‌دهد.

شاپای الکترونیکی: ۳۰۶۰-۶۶۹۱

مقدمه

تشخیص اعداد دست‌نویس یکی از وظایف اساسی در حوزه بینایی ماشین[1] و شناسایی الگوها[2] است که به‌عنوان یک معیار استاندارد برای ارزیابی الگوریتم‌ها و مدل‌های جدید در زمینه یادگیری عمیق[3] مورد استفاده قرار می‌گیرد (لجئونه، ۲۰۲۰)[4]. مجموعه داده MNIST (لکان و همکاران، ۲۰۱۰)[5] که شامل ۶۰،۰۰۰ تصویر آموزشی و ۱۰،۰۰۰ تصویر آزمایشی از اعداد دست‌نویس است، به دلیل سادگی و قالب استاندارد خود به‌طور گسترده‌ای برای این منظور استفاده شده است (برنگارت، ۲۰۲۳)[6]. با وجود پیشرفت‌های قابل‌توجه در تکنیک‌های یادگیری عمیق در سال‌های اخیر، دستیابی به دقت بالا در تشخیص اعداد دست‌نویس همچنان یک معضل مهم است، به‌ویژه زمانی که نیاز به فرایندهای آموزشی کارآمد و مدیریت منابع وجود دارد (تای، ۲۰۲۳)[7].

یادگیری عمیق، به‌ویژه شبکه‌های عصبی پیچشی[8]، انقلابی در وظایف طبقه‌بندی تصویر، ازجمله تشخیص اعداد دست‌نویس، ایجاد کرده است (تای، ۲۰۲۳). شبکه‌های CNN برای یادگیری خودکار و تطبیقی سلسله‌مراتب ویژگی‌های فضایی از طریق پس‌انتشار با استفاده از بلوک‌های مختلفی مانند لایه‌های پیچشی[9]، لایه‌های تجمیع[10] و لایه‌های کاملاً متصل[11] طراحی شده‌اند (صالحی و همکاران، ۲۰۲۳)[12]. با این حال، آموزش شبکه‌های عمیق CNN از ابتدا نیازمند منابع محاسباتی قابل‌توجه و مجموعه داده‌های بزرگ است که همیشه امکان‌پذیر نیست.

برای مقابله با این معضلات، یادگیری انتقالی[13] به‌عنوان یک تکنیک قدرتمند در یادگیری عمیق ظهور کرده است (ایمان و همکاران، ۲۰۲۳). یادگیری انتقالی شامل استفاده از یک مدل ازپیش آموزش‌دیده می‌باشد که معمولاً بر روی یک مجموعه داده بزرگ مانند ImageNet[14] آموزش دیده است و تنظیم مجدد آن برای یک وظیفه خاص مانند طبقه‌بندی اعداد دست‌نویس MNIST (لکان و همکاران، ۲۰۱۰) می‌باشد (کوهن و همکاران، ۲۰۱۷؛ دنگ و همکاران، ۲۰۰۹؛ حسن و همکاران، ۲۰۲۴)[15]. این روش از قابلیت مدل ازپیش آموزش‌دیده برای استخراج ویژگی‌های عمومی از تصاویر بهره می‌برد بنابراین نیاز به داده‌ها و قدرت محاسباتی برای آموزش مدل جدید کاهش می‌یابد (چن و همکاران، ۲۰۲۴)[16].

یکی از شناخته‌شده‌ترین مدل‌هایی که در یادگیری انتقالی استفاده می‌شود، مدل VGG16[17] است که توسط گروه هندسه بصری در دانشگاه آکسفورد توسعه داده شده است و به دلیل سادگی و کارایی خود، معروف است و دارای ۱۶ لایه با فیلترهای پیچشی کوچک (۳۳) و عملکرد بالا در وظایف مختلف شناسایی تصویر می‌باشد (رودرگودا و همکاران،

---

[1] Machine Vision
[2] Pattern Recognition
[3] Deep Learning
[4] Lejeune
[5] Lecun
[6] Berngardt
[7] Taye
[8] Convolutional Neural Network (CNN)
[9] Convolutional Layers
[10] Pooling Layers
[11] Fully Connected Layers
[12] Salehi
[13] Transfer Learning
[14] Accessible at: https://www.image-net.org/download.php
[15] Cohen; Deng; Hassan
[16] Chen
[17] Visual Geometry Group 16

۲۰۲۳)[1]. اعمال VGG۱۶ به مجموعه داده MNIST (لکان و همکاران، ۲۰۱۰) شامل پیکربندی مجدد لایه‌های نهایی شبکه برای تطبیق با تعداد کلاس‌های موجود در مجموعه داده و تنظیم وزن‌ها بر اساس داده‌های جدید است (باکاسا و ویریری، ۲۰۲۳)[2].

این مطالعه به ارزیابی کارایی استفاده از مدل VGG16 برای تشخیص اعداد دست‌نویس بر روی مجموعه داده MNIST (لکان و همکاران، ۲۰۱۰) از طریق یادگیری انتقالی می‌پردازد. به‌طور خاص، عملکرد مدل VGG16 با مدل‌های رایج دیگر یادگیری عمیق، مانند CNNهایی که از ابتدا آموزش داده شده‌اند، پرسپترون‌های چندلایه[3] و الگوریتم‌های یادگیری ماشین سنتی (ماشین بردار پشتیبان، درخت تصمیم و بیز ساده) با استفاده از معیارهایی مانند Accuracy، Precision، Recall و F1 score مقایسه خواهد شد.

ضرورت این پژوهش از نیاز فزاینده به مدل‌های کارآمد و دقیق ناشی می‌شود که بتوانند با منابع محاسباتی و داده‌های محدود آموزش ببینند. در کاربردهای عملی، مانند پردازش خودکار فرم‌ها و تشخیص اعداد در محیط‌های کم‌منبع، توانایی استقرار مدل‌های مؤثر بدون نیاز به آموزش گسترده، بسیار حیاتی است. علاوه بر این، با بررسی کاربرد مدل‌های ازپیش آموزش‌دیده مانند VGG16 برای تشخیص اعداد دست‌نویس MNIST (لکان و همکاران، ۲۰۱۰) می‌توان بینشی درباره قابلیت تعمیم این مدل‌ها و چگونگی تطبیق آنها با وظایف جدید اما مرتبط کسب کرد.

در نتیجه، هدف اصلی این پژوهش، نشان دادن کاربرد و مزایای یادگیری انتقالی، به‌ویژه استفاده از مدل VGG16، برای تشخیص اعداد دست‌نویس است. همچنین با ارائه مقایسه جامع با سایر روش‌های یادگیری عمیق و ماشین، به کارایی، دقت و عملی‌بودن استفاده از مدل‌های ازپیش آموزش‌دیده در حل مسائل اساسی بینایی ماشین تأکید خواهد شد. این بررسی نه‌تنها پتانسیل یادگیری انتقالی را نشان می‌دهد بلکه نیاز به روش‌های آموزشی کارآمدتر در زمینه یادگیری عمیق را نیز مد نظر قرار می‌دهد.

مبانی نظری

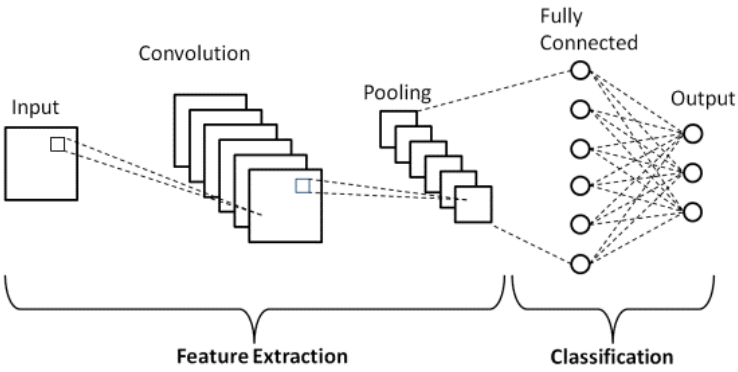
یادگیری عمیق و به‌ویژه شبکه‌های عصبی پیچشی، ازجمله روش‌های پیشرو در زمینه بینایی ماشین و شناسایی الگوها هستند (صالحی و همکاران، ۲۰۲۳). شبکه‌های عصبی پیچشی به دلیل توانایی‌های منحصربه‌فرد خود در استخراج ویژگی‌های پیچیده از داده‌های تصویری، به‌طور گسترده‌ای در مسائل مختلف ازجمله تشخیص اعداد دست‌نویس به‌کار گرفته شده‌اند (تئودوریس و همکاران، ۲۰۲۳)[4]. در این بخش، مبانی نظری مرتبط با یادگیری عمیق، شبکه‌های عصبی پیچشی، یادگیری انتقالی و کاربرد آنها در تشخیص اعداد دست‌نویس بررسی می‌شود.

یادگیری عمیق و شبکه‌های عصبی پیچشی

یادگیری عمیق، یک زیرمجموعه از یادگیری ماشینی است که از شبکه‌های عصبی مصنوعی با لایه‌های متعدد برای یادگیری ویژگی‌های پیچیده و الگوها در داده‌ها استفاده می‌کند (اصلانی و جاکوب، ۲۰۲۳)[5]. شبکه‌های عصبی پیچشی به‌طور خاص برای پردازش داده‌های تصویری، طراحی شده‌اند و از سه نوع لایه اصلی تشکیل شده‌اند: لایه‌های پیچشی،

[1] Rudregowda
[2] Bakasa & Viriri
[3] Multi-layer Perceptron (MLP)
[4] Theodoris
[5] Aslani & Jacob

لایه‌های تجمیع و لایه‌های کاملاً متصل. شکل ۱ معماری الگوریتم شبکه‌های عصبی پیچشی را نشان می‌دهد (شانگ و همکاران، ۲۰۲۴)[1].



شکل ۱. معماری الگوریتم شبکه‌های عصبی پیچشی.

- **لایه‌های پیچشی:** این لایه‌ها فیلترهایی را به‌کار می‌گیرند که تصاویر را برای استخراج ویژگی‌های محلی اسکن می‌کنند. این فیلترها به شناسایی ویژگی‌های پایه مانند لبه‌ها، گوشه‌ها و الگوهای تکراری کمک می‌کنند (تای، ۲۰۲۳).
- **لایه‌های تجمیع:** این لایه‌ها ابعاد ویژگی‌های استخراج‌شده را کاهش می‌دهند و ویژگی‌های مهم را حفظ می‌کنند. لایه‌های تجمیع معمولاً از عملیات ماکسیمم‌پولینگ[2] یا میانگین‌پولینگ[3] استفاده می‌کنند (صالحی و همکاران، ۲۰۲۳).
- **لایه‌های کاملاً متصل:** این لایه‌ها که در انتهای شبکه قرار دارند، ویژگی‌های استخراج‌شده را برای انجام وظیفه نهایی، مانند طبقه‌بندی به‌کار می‌گیرند (تئودوریس و همکاران، ۲۰۲۳).

شبکه‌های CNN با توانایی خود در یادگیری ویژگی‌های چندلایه، به ابزارهای بسیار مؤثری برای تشخیص و طبقه‌بندی تصاویر تبدیل شده‌اند. با این حال، آموزش این شبکه‌ها از ابتدا نیازمند منابع محاسباتی زیاد و مجموعه داده‌های بزرگ است (باکاسا و ویریری، ۲۰۲۳).

یادگیری انتقالی[4]

یادگیری انتقالی یک روش مؤثر برای کاهش نیاز به منابع محاسباتی و داده‌های بزرگ در آموزش مدل‌های یادگیری عمیق است (ایمان و همکاران، ۲۰۲۳). در این روش، یک مدل ازپیش آموزش‌دیده، که معمولاً بر روی یک مجموعه داده بزرگ مانند ImageNet آموزش دیده است، برای یک وظیفه جدید مانند تشخیص اعداد دست‌نویس، تنظیم مجدد[5] می‌شود (عزیزی و همکاران، ۲۰۲۳). یادگیری انتقالی بر این اصل استوار است که ویژگی‌های

[1] Shang
[2] Max Pooling
[3] Mean Pooling
[4] Transfer Learning
[5] Fine-Tuning

یادگرفته‌شده در یک دامنه می‌توانند به دامنه‌های دیگر انتقال یابند، به شرطی که دامنه‌ها ویژگی‌های مشترکی داشته باشند (چن و همکاران، ۲۰۲۴).

مدل‌های ازپیش آموزش‌دیده، مانند VGG16، قادر به استخراج ویژگی‌های عمومی از تصاویر هستند که می‌توانند در مسائل مختلف کاربرد داشته باشند (رودرگودا و همکاران، ۲۰۲۳). با تنظیم مجدد لایه‌های نهایی این مدل‌ها و آموزش مجدد آنها با داده‌های جدید، می‌توان به نتایج دقیقی دست یافت بدون اینکه نیاز به آموزش مدل از ابتدا باشد (ایمان و همکاران، ۲۰۲۳).

مجموعه داده MNIST

MNIST (لکان و همکاران، ۲۰۱۰) یک پایگاه داده استاندارد در حوزه یادگیری ماشین و بینایی کامپیوتری است که برای آموزش و ارزیابی الگوریتم‌های طبقه‌بندی تصویر به‌کار می‌رود (برنگارت، ۲۰۲۳). این مجموعه داده شامل ۶۰،۰۰۰ تصویر دست‌نویس از ارقام ۰ تا ۹ برای آموزش و ۱۰،۰۰۰ تصویر برای آزمون است (حسن و همکاران، ۲۰۲۴). هر تصویر در این مجموعه به‌صورت سیاه و سفید و با ابعاد ۲۸*۲۸ پیکسل ارائه شده است. MNIST (لکان و همکاران، ۲۰۱۰) به‌عنوان یکی از اولین و محبوب‌ترین مجموعه داده‌ها در یادگیری ماشین، به‌خصوص در زمینه تشخیص دست‌نویس، شناخته می‌شود (لجئونه، ۲۰۲۰). این پایگاه داده را یان لکان[1] و همکارانش در سال ۱۹۹۸ معرفی کردند (لکان و همکاران، ۱۹۹۸) که از آن زمان تاکنون به عنوان یک نقطه شروع استاندارد برای آزمایش و مقایسه الگوریتم‌های مختلف مورد استفاده قرار گرفته است. شکل ۲، نمونه‌ای از تصاویر دست‌نویس موجود در مجموعه داده MNIST (لکان و همکاران، ۲۰۱۰) را نشان می‌دهد.

شکل ۲. نمونه‌ای از تصاویر مجموعه داده MNIST (عزیزی و همکاران، ۲۰۲۳).

یکی از دلایل محبوبیت MNIST (لکان و همکاران، ۲۰۱۰)، ساده و در عین حال چالش‌برانگیزبودن آن است. به عبارت دیگر، این مجموعه داده به اندازه کافی پیچیده است تا بتواند توانایی الگوریتم‌های مختلف را به‌خوبی محک بزند امّا در عین حال به اندازه کافی ساده است که حتی مبتدیان نیز بتوانند با آن کار کنند و نتایج قابل‌قبولی به‌دست آورند (لکان و همکاران، ۱۹۹۸). استفاده گسترده از MNIST (لکان و همکاران، ۲۰۱۰) در جامعه تحقیقاتی، امکان مقایسه مستقیم عملکرد مدل‌های مختلف را فراهم می‌آورد و به پیشرفت تکنیک‌های جدید در یادگیری ماشین کمک می‌کند.

[1] Yann LeCun

مدل VGG16

مدل VGG16 که گروه هندسه بصری دانشگاه آکسفورد توسعه داده‌اند، یکی از شناخته‌شده‌ترین مدل‌های یادگیری عمیق است که بر روی مجموعه داده ImageNet آموزش دیده است (باکاسا و ویریری، ۲۰۲۳). این مدل دارای ۱۶ لایه با فیلترهای پیچشی کوچک (۳*۳) است که به دلیل سادگی و کارایی خود در استخراج ویژگی‌های تصویر شناخته شده است. VGG16 به دلیل ساختار لایه‌ای عمیق خود، توانایی یادگیری ویژگی‌های بسیار پیچیده و دقیق از تصاویر را دارد (رودرگودا و همکاران، ۲۰۲۳).

کاربرد یادگیری انتقالی در تشخیص اعداد دست‌نویس

استفاده از یادگیری انتقالی با مدل VGG16 برای تشخیص اعداد دست‌نویس در مجموعه داده MNIST (لکان و همکاران، ۲۰۱۰) یک راهکار مؤثر برای بهره‌گیری از ویژگی‌های ازپیش یادگرفته شده است (چاندوره و اینامدار، ۲۰۲۳)[1]. با تنظیم مجدد لایه‌های نهایی مدل VGG16 برای تطبیق با تعداد کلاس‌های MNIST (لکان و همکاران، ۲۰۱۰) و آموزش مجدد آن با داده‌های این مجموعه، می‌توان به دقت بالا در تشخیص اعداد دست یافت. این روش نه‌تنها باعث افزایش دقت مدل می‌شود بلکه نیاز به منابع محاسباتی و داده‌های آموزشی زیاد را نیز کاهش می‌دهد (فاتح و همکاران، ۲۰۲۱).

پارامترهای ارزیابی

در ارزیابی مدل‌های انتقال یادگیری، یادگیری عمیق و یادگیری ماشین سنتی، معمولاً از چندین پارامتر مختلف برای سنجش عملکرد مدل‌ها و کیفیت نتایج استفاده می‌شود. در زیر پارامترهای اصلی ارزیابی این نوع مطالعات به همراه توضیح و رابطه ریاضی ارائه شده است:

۱- **نرخ صحت (Accuracy):** نسبت تعداد نمونه‌های درست تشخیص‌داده‌شده توسط مدل به کل تعداد نمونه‌ها. نرخ صحت به‌صورت رابطه ۱ محاسبه می‌شود (تقندیکی، ۲۰۲۳).

$$\frac{TP + TN}{TP + TN + FP + FN} \quad (۱)$$

که TP تعداد نمونه‌های مثبت درست، TN تعداد نمونه‌های منفی درست، FP تعداد نمونه‌های مثبت اشتباه و FN تعداد نمونه‌های منفی اشتباه است.

۲- **نرخ کامل‌بودن (Recall):** نسبت تعداد نمونه‌های مثبت درست تشخیص داده‌شده توسط مدل به کل تعداد نمونه‌های مثبت. نرخ کامل‌بودن به‌صورت رابطه ۲ محاسبه می‌شود (غفاریان و بامحبت، ۲۰۲۳؛ تقندیکی و همکاران، ۲۰۲۳).

$$\frac{TP}{TP + FN} \quad (۲)$$

۳- **نرخ دقت واقعاً مثبت (Precision):** نشان می‌دهد که چه درصدی از نمونه‌هایی که مدل به‌عنوان مثبت پیش‌بینی کرده است، واقعاً مثبت هستند. نرخ دقت واقعاً مثبت به‌صورت رابطه ۳ محاسبه می‌شود (تقندیکی، ۲۰۲۳).

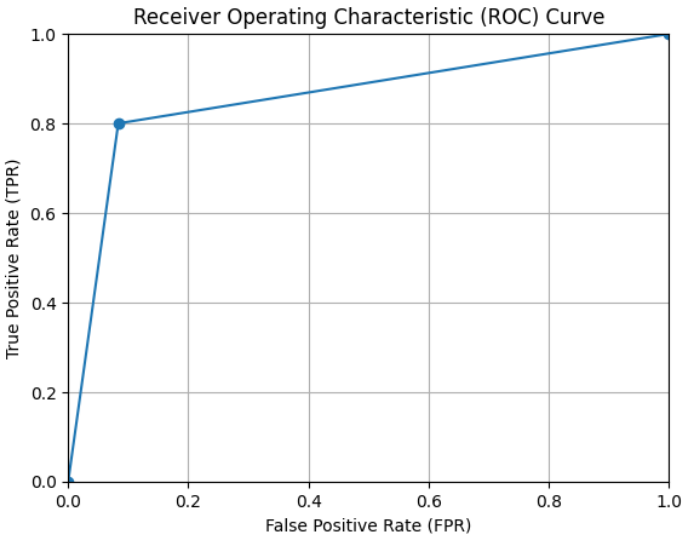
[1] Chandure & Inamdar

$$\frac{TP}{TP+FP} \quad (۳)$$

۴- **نرخ F1 Score**: میانگین هندسی F1، برای مواقعی که دسته‌بندی نمونه‌ها نیاز به توازن بین نرخ کامل‌بودن و نرخ واقعاً مثبت دارد، مورد استفاده قرار می‌گیرد. معیار F1 به صورت رابطه ۴ محاسبه می‌شود (نامجوی‌راد و دادگرپور، ۲۰۲۱؛ تقندیکی و همکاران، ۲۰۲۳).

$$\frac{TP}{TP+\frac{1}{2}(FP+FN)} \quad (۴)$$

۵- **منحنی ROC**[1]: یک رسم نمودار از نرخ واقعی مثبت (TPR[2]) در مقابل نرخ اشتباه مثبت (FPR[3]) برای مدل‌های طبقه‌بندی منحنی ROC به‌عنوان یک معیار برای اندازه‌گیری عملکرد مدل‌های طبقه‌بندی استفاده می‌شود (کریژوسکی و همکاران، ۲۰۱۲)[4]. شکل ۳، نمونه‌ای از یک منحنی ROC را نشان می‌دهد.



شکل ۳. نمودار ROC.

[1] Receiver Operating Characteristic
[2] True Positive Rate
[3] False Positive Rate
[4] Krizhevsky

مرور پیشینه

۱- LeNet-5

یان لکان و همکارانش در سال ۱۹۹۸ یکی از اولین کاربردهای موفق CNN در تشخیص اعداد دست‌نویس را معرفی کردند. مدل LeNet-5 شامل دو لایه پیچشی[1] و دو لایه تجمیع[2] به همراه سه لایه کاملاً متصل بود. این مدل با استفاده از ویژگی‌های محلی تصاویر و کاهش پیچیدگی محاسباتی توانست با دقت بالایی اعداد دست‌نویس را شناسایی کند، نرخ صحت اولیه این مدل ۹۵ درصد بود (لکان و همکاران، ۱۹۹۸).

۲- AlexNet

کریژوسکی در سال ۲۰۱۲ مدل AlexNet را توسعه دادند که به‌طور قابل‌توجهی عمیق‌تر از LeNet-5 بود. اگرچه AlexNetبرای مجموعه داده ImageNet طراحی شده بود، اصول و معماری آن برای مجموعه داده‌های کوچک‌تر مانند MNIST (لکان و همکاران، ۲۰۱۰) نیز به‌خوبی قابل تعمیم بود. این مدل با استفاده از لایه‌های پیچشی بیشتر و تکنیک‌هایی مانند ReLU و Dropout، عملکرد بسیار خوبی در شناسایی تصاویر داشت، نرخ صحت این مدل ۸۰ درصد بود (کریژوسکی و همکاران، ۲۰۱۲).

۳- ZFNet

زیلر و فِرگاس[3] در سال ۲۰۱۴ مدل ZFNet را معرفی کردند که به بهبود معماری AlexNet پرداخت. این مدل با استفاده از تکنیک‌های بصری‌سازی توانست عملکرد بهتری در تشخیص ویژگی‌ها و بهبود دقت تشخیص اعداد دست‌نویس داشته باشد. این مدل توانست با تحلیل و بهینه‌سازی لایه‌های پیچشی، ویژگی‌های مهم‌تری از تصاویر را استخراج کند، Accuracy این مدل ۷۰ درصد بود (زیلر و فرگاس، ۲۰۱۴).

۴- شبکه‌های عصبی عمیق ساده

کیرسان[4] و همکارانش در سال ۲۰۱۲ نشان دادند که شبکه‌های عصبی عمیق ساده با لایه‌های پیچشی و تمام‌متصل می‌توانند دقت بالایی در تشخیص اعداد دست‌نویس MNIST (لکان و همکاران، ۲۰۱۰) داشته باشند. این مطالعه بر اهمیت عمق شبکه و تکنیک‌های آموزشی مختلف مانند پس‌انتشار تأکید داشت و نشان داد که افزایش عمق شبکه منجر به بهبود دقت تشخیص می‌شود، Accuracy این مدل ۹۵ درصد بود (کیرسان و همکاران، ۲۰۱۰).

۵- VGGNet

مدل VGGNet را سیمونیان و زیسرمان[5] در سال ۲۰۱۴ توسعه دادند. این مدل از فیلترهای پیچشی کوچک ۳*۳ استفاده می‌کند و دارای عمق بسیار زیادی است. VGGNet با ساختار ساده و کارایی بالای خود توانست دقت بسیار بالایی در تشخیص اعداد دست‌نویس مجموعه داده MNIST (لکان و همکاران، ۲۰۱۰) نشان دهد. این مدل به دلیل سادگی در طراحی و کارایی بالا به‌طور گسترده‌ای در یادگیری انتقالی نیز مورد استفاده قرار گرفت، Accuracy این مدل ۷۵ درصد بود (سیمونیان و زیسرمان، ۲۰۱۴).

---

[1] Convolutional Layers
[2] Subsampling Layers
[3] Zeiler & Fergus
[4] Cireşan
[5] Simonyan & Zisserman

۶- ALBERT

ALBERT یک مدل مبتنی بر Transformer است که با هدف کاهش پیچیدگی محاسباتی و حافظه موردنیاز مدل BERT (Bidirectional Encoder Representations from Transformers) معرفی شده است. این مدل با استفاده از تکنیک‌های مختلف مانند عامل‌سازی ماتریس توجه و به‌اشتراک‌گذاری پارامترها، به‌طور قابل‌توجهی مصرف حافظه و زمان محاسبه را کاهش می‌دهد. علی‌رغم این کاهش پیچیدگی، ALBERT به نرخ صحت ۹۹.۸۱ درصد در تشخیص اعداد دست‌نویس روی MNIST (لکان و همکاران، ۲۰۱۰) دست می‌یابد که نسبت به مدل BERT با Accuracy، ۹۹.۷۷ درصد عملکرد بهتری دارد (لان و همکاران، ۲۰۱۹)[1].

۷- EfficientNet

EfficientNet یک خانواده از مدل‌های شبکه عصبی کانولوشن (ConvNet) است که با هدف بهبود کارایی و دقت طراحی شده‌اند. این مدل‌ها با استفاده از تکنیک‌های مختلف مانند جستجوی خودکار معماری و ترکیب بلوک‌های مختلف ConvNet، به تعادل بین دقت و کارایی دست می‌یابند. EfficientNet با Accuracy، ۹۹.۵۳ درصد در تشخیص اعداد دست‌نویس روی MNIST عملکرد قابل‌قبولی دارد (تان و لی، ۲۰۱۹)[2].

۸- MixNet

MixNet یک خانواده دیگر از مدل‌های ConvNet است که با هدف بهبود کارایی و دقت طراحی شده‌اند. این مدل‌ها با استفاده از تکنیک‌های مختلف مانند ترکیب بلوک‌های مختلف ConvNet و استفاده از روش‌های فشرده‌سازی، به تعادل بین دقت و کارایی دست می‌یابند. MixNet با Accuracy، ۹۹.۵۶ درصد در تشخیص اعداد دست‌نویس روی MNIST (لکان و همکاران، ۲۰۱۰) عملکرد قابل‌قبولی دارد (تان و لی، ۲۰۱۹)[3].

۹- CutMix

CutMix یک روش تنظیم‌سازی برای مدل‌های یادگیری عمیق است که با هدف بهبود دقت و کارایی معرفی شده است. این روش با ترکیب تصادفی دو تصویر و آموزش مدل با استفاده از برچسب تصویر اصلی، مدل را به یادگیری ویژگی‌های بصری قوی‌تر تشویق می‌کند. CutMix در ترکیب با سایر روش‌ها مانند مدل‌های ConvNet می‌تواند به نرخ صحت ۹۹.۸۴ درصد در تشخیص اعداد دست‌نویس روی MNIST (لکان و همکاران، ۲۰۱۰) دست یابد (یون و همکاران، ۲۰۱۹)[4].

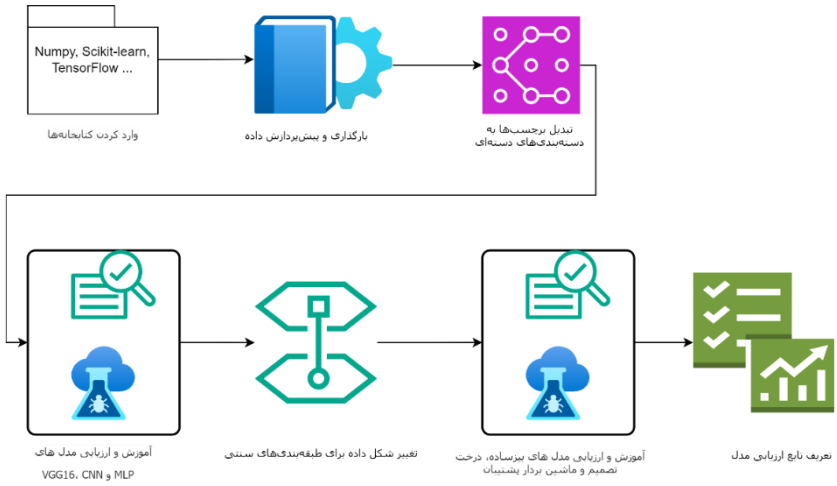
کارهای انجام‌شده با استفاده از الگوریتم‌های CNN برای تشخیص اعداد دست‌نویس MNIST (لکان و همکاران، ۲۰۱۰) نشان می‌دهند که شبکه‌های عصبی پیچشی با لایه‌های عمیق و تکنیک‌های بهینه‌سازی مناسب می‌توانند دقت بسیار بالایی در این زمینه داشته باشند. این مطالعات پایه‌ای برای توسعه مدل‌های پیچیده‌تر و استفاده از یادگیری انتقالی در مسائل مرتبط با بینایی ماشین فراهم کرده‌اند. مرور پیشینه نشان می‌دهد که شبکه‌های عصبی پیچشی (CNN) و یادگیری انتقالی ازجمله روش‌های مؤثر برای تشخیص اعداد دست‌نویس هستند. استفاده از مدل‌های ازپیش آموزش‌دیده مانند VGG16 به عنوان یک روش یادگیری انتقالی، می‌تواند بهبود قابل‌توجهی در دقت و کارایی مدل‌ها به همراه داشته باشد. تحقیقات نشان داده‌اند که این روش‌ها نه‌تنها نیاز به داده‌های آموزشی بزرگ را کاهش می‌دهند بلکه زمان و منابع

[1] Lan
[2] Tan & Le
[3] Tan & Le
[4] Yun

محاسباتی موردنیاز برای آموزش مدل‌ها را نیز بهبود می‌بخشند. در این پژوهش، با تمرکز بر استفاده از یادگیری انتقالی و مدل VGG16 برای تشخیص اعداد دست‌نویس، به بررسی و ارزیابی این روش‌ها پرداخته خواهد شد.

روش‌شناسی

برای اجرای رویکرد پیشنهادی از زبان برنامه‌نویسی پایتون ۳.۱۲.۳ استفاده شد. شکل ۴، مراحل رویکرد پیشنهادی را نشان می‌دهد. هریک از مراحل در ادامه توضیح داده شده‌اند.



شکل ۴. مراحل رویکرد پیشنهادی.

۱- **واردکردن کتابخانه‌ها:** در اولین مرحله، کتابخانه‌های ضروری برای اجرای کد وارد می‌شوند. کتابخانه‌هایی مانند NumPy برای انجام عملیات عددی و TensorFlow برای ساخت و آموزش مدل‌های یادگیری عمیق به‌کار گرفته می‌شوند. علاوه بر این، از کتابخانه‌های مرتبط با یادگیری ماشین مانند scikit-learn برای اجرای الگوریتم‌های سنتی یادگیری ماشین استفاده می‌شود.

۲- **بارگذاری و پیش‌پردازش داده‌ها:** در این بخش، مجموعه داده MNIST (لکان و همکاران، ۲۰۱۰) که شامل تصاویر دست‌نویس اعداد است، بارگیری می‌شود. سپس داده‌ها به دو بخش آموزش و آزمایش تقسیم می‌شوند. پس از آن، تصاویر نرمال‌سازی می‌شوند تا مقیاس پیکسل‌های آنها یکسان شود و به شکل مناسبی برای ورودی مدل تغییر شکل داده می‌شوند.

۳- **تبدیل برچسب‌ها به دسته‌بندی‌های دسته‌ای:** برچسب‌های مرتبط با تصاویر که اعداد ۰ تا ۹ را نشان می‌دهند، به بردارهای one-hot تبدیل می‌شوند. این تبدیل، ضروری است زیرا مدل‌های یادگیری عمیق در مسائل طبقه‌بندی چندکلاسه نیاز به برچسب‌هایی به این شکل دارند تا بتوانند به‌درستی آموزش ببینند.

۴- **آموزش و ارزیابی مدل‌های VGG16، CNN و MLP:** در این مرحله، ابتدا مدل VGG16 با وزن‌های ازپیش آموزش‌داده‌شده بر روی مجموعه داده ImageNet استفاده می‌شود. لایه‌های پایه این مدل قفل می‌شوند و فقط لایه‌های نهایی با داده‌های جدید MNIST (لکان و همکاران، ۲۰۱۰) آموزش داده می‌شوند. سپس مدل‌های شبکه عصبی پیچشی و شبکه عصبی پیش‌خور ساخته و با داده‌های آموزشی آموزش داده می‌شوند.

۵- **تغییر شکل داده برای طبقه‌بندی‌های سنتی:** داده‌های تصویری برای استفاده در الگوریتم‌های سنتی یادگیری ماشین مانند بیز ساده، درخت تصمیم و ماشین بردار پشتیبان، به شکل‌های یک‌بعدی تغییر شکل داده می‌شوند. این تغییر شکل به الگوریتم‌های سنتی امکان می‌دهد که با داده‌های تصویر کار کنند.

۶- **آموزش و ارزیابی مدل بیز ساده، درخت تصمیم و ماشین بردار پشتیبان:** در این مرحله، مدل‌های سنتی یادگیری ماشین مانند بیز ساده، درخت تصمیم و ماشین بردار پشتیبان آموزش داده می‌شوند. این مدل‌ها با داده‌های آموزشی تمرین داده شده و سپس با استفاده از داده‌های آزمایشی ارزیابی می‌شوند.

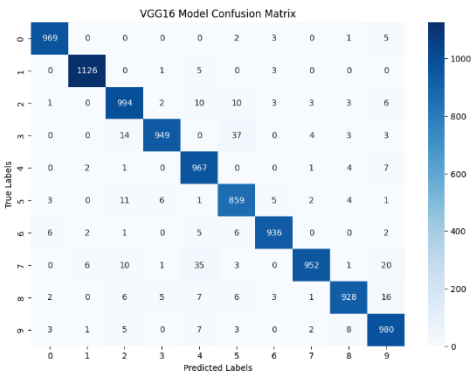
۷- **تعریف تابع ارزیابی مدل:** در این مرحله، یک تابع ارزیابی تعریف می‌شود که عملکرد مدل‌های توسعه‌داده‌شده را با استفاده از معیارهای مختلف ارزیابی می‌کند. این معیارها شامل نرخ دقت (Accuracy)، نرخ دقیق‌بودن (Precision)، نرخ کامل‌بودن (Recall) و نرخ میانگین هارمونیک (F1-score) هستند. این توابع کمک می‌کنند تا کیفیت و کارایی مدل‌ها به‌دقت سنجیده و مدل‌های بهینه شناسایی شوند.

این فرایند، مراحل مختلف از آماده‌سازی داده تا ارزیابی مدل‌ها را به‌طور کامل پوشش می‌دهد و در نهایت به مقایسه کارایی مدل‌های مختلف برای تشخیص اعداد دست‌نویس می‌پردازد.

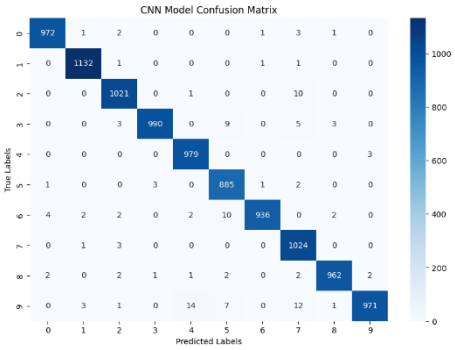
در این پژوهش، نوآوری اصلی در استفاده از یادگیری انتقالی برای بهبود دقت تشخیص اعداد دست‌نویس نهفته است. به‌جای آموزش مدل‌ها از ابتدا، از مدل VGG16 که قبلاً بر روی مجموعه داده ImageNet آموزش دیده بود، استفاده شد. با این روش، زمان آموزش به‌طور قابل‌توجهی کاهش یافت و دقت مدل به ۹۹ درصد رسید. این مدل با تنظیم دقیق لایه‌های انتهایی و استفاده از تکنیک‌هایی مانند افزایش داده، توانست عملکرد بهتری نسبت به مدل‌های سنتی و حتی برخی از شبکه‌های عصبی پیچشی به‌دست آورد. استفاده از این رویکرد، راهی مؤثر برای بهبود عملکرد مدل‌های یادگیری عمیق با صرفه‌جویی در منابع محاسباتی است.

نتایج و بحث

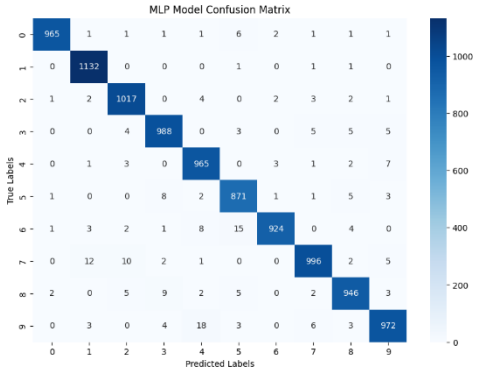
شکل‌های ۵، ۶ و ۷ به‌ترتیب ماتریس درهم‌ریختگی ایجادشده از اجرای مدل‌های CNN، MLP و VGG16 را هنگام تشخیص اعداد دست‌نویس نشان می‌دهند.



شکل ۵. ماتریس درهم‌ریختگی ایجادشده از الگوریتم VGG16

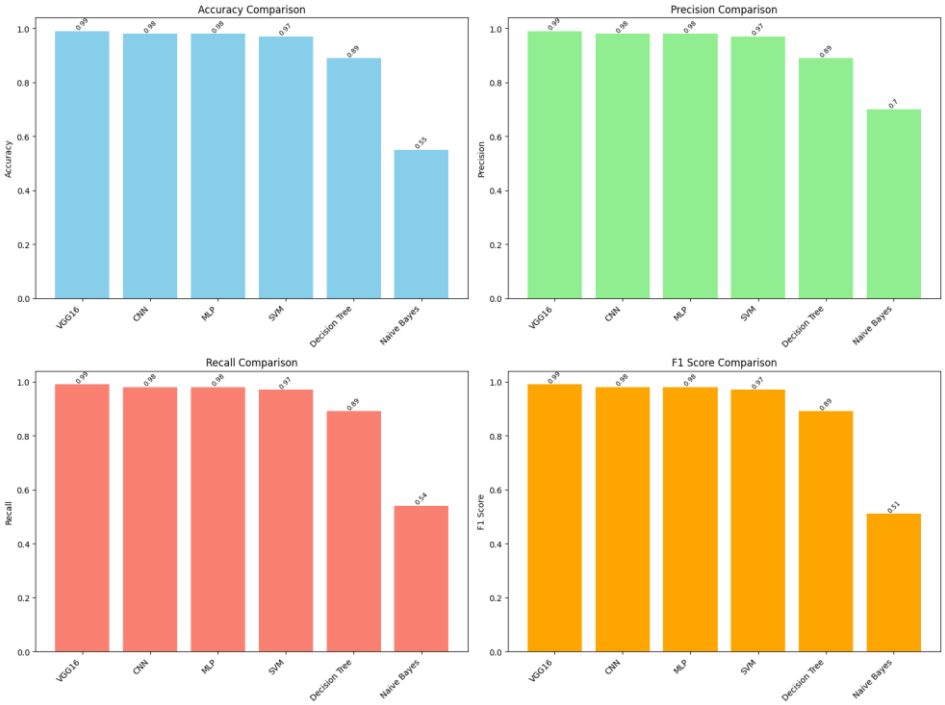


شکل ۶. ماتریس درهم‌ریختگی ایجادشده از الگوریتم CNN



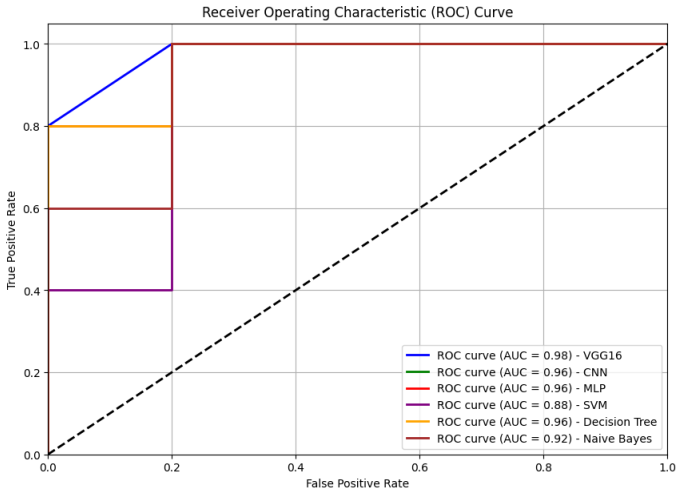
شکل ۷. ماتریس درهم‌ریختگی ایجادشده از الگوریتم MLP

در شکل ۸ عملکرد مدل VGG16 با استفاده از Accuracy، Precision، Recall، F1 score و نتایج آن با سایر الگوریتم‌های یادگیری عمیق مانند شبکه‌های عصبی پیچشی، شبکه‌های عصبی پیش‌خور و الگوریتم‌های یادگیری ماشین سنتی مقایسه شده است.



شکل ۸. معیارهایی Accuracy، Precision، Recall و F1 Score الگوریتم‌های یادگیری عمیق، یادگیری سنتی و مدل VGG16

با توجه به شکل ۸، مدل VGG16 با Accuracy، ۹۹ درصد بهترین عملکرد را دارد. مدل‌های CNN و MLP با Accuracy، ۹۸ درصد در رتبه‌های بعدی قرار دارند. همچنین، مدل‌های یادگیری ماشین سنتی مانند ماشین بردار پشتیبان، درخت تصمیم و بیز ساده به‌ترتیب با مقادیر ۹۷، ۸۹ و ۵۵ درصد، پایین‌ترین Accuracy را در تشخیص تصاویر دست‌نویس نشان می‌دهند. شکل ۹ منحنی ROC هریک از مدل‌های توسعه داده‌شده را نشان می‌دهد.



شکل ۹. معیارهایی Accuracy، Precision، Recall و F1 Score الگوریتم‌های یادگیری عمیق، یادگیری سنتی و مدل VGG16.

براساس منحنی‌های ROC نشان داده‌شده در شکل ۹، مدل VGG16 بهترین عملکرد را با ۰.۹۸ دارد. مدل‌های CNN و MLP نیز عملکرد خوبی با ۰.۹۶ دارند. مدل‌های ماشین بردار پشتیبان، درخت تصمیم و بیز ساده عملکرد ضعیف‌تری به‌ترتیب با مقادیر ۰.۸۸، ۰.۹۶ و ۰.۹۲ می‌باشد.

مقایسه

در جدول ۱، نتایج پژوهش انجام‌شده با سایر کارهای مرتبط مقایسه شده است:

جدول ۱. مقایسه نتایج پژوهش انجام‌شده با سایر کارهای مرتبط.

مدل	دقت (درصد)	ویژگی‌های کلیدی
VGG16 (پژوهش انجام‌شده)	۹۹	معماری عمیق، استفاده از انتقال یادگیری
LeNet-5 (لکان و همکاران، ۱۹۹۸)	۹۵	یکی از اولین مدل‌های CNN، ساختار ساده
AlexNet (کریژوسکی و همکاران، ۲۰۱۲)	۸۰	معماری عمیق‌تر از LeNet-5، استفاده از ReLU و Dropout
ZFNet (زیلر و فرگاس، ۲۰۱۴)	۷۰	بهبود معماری AlexNet با استفاده از تکنیک‌های بصری‌سازی
شبکه‌های عصبی عمیق ساده	۹۵	تأکید بر اهمیت عمق شبکه

مدل	دقت (درصد)	ویژگی‌های کلیدی
(کیرسان و همکاران، ۲۰۱۰)		
VGGNet (سیمونیان و زیسرمان، ۲۰۱۴)	۷۵	معماری ساده با فیلترهای کوچک x3۳
ALBERT (لان و همکاران، ۲۰۱۹)	۹۹/۸۱	مبتنی بر Transformer، کاهش پیچیدگی محاسباتی
EfficientNet (تان و لی، ۲۰۱۹)	۹۹/۵۳	بهبود کارایی و دقت
MixNet (تان و لی، ۲۰۱۹)	۹۹/۵۶	بهبود کارایی و دقت
CutMix (یون و همکاران، ۲۰۱۹)	۹۹/۸۴	تکنیک تقویت داده

در مقایسه با سایر کارهای انجام‌شده در حوزه تشخیص اعداد دست‌نویس، مشاهده می‌شود که مدلLeNet-5 ، که یکی از اولین کاربردهای موفق شبکه‌های عصبی پیچشی در این زمینه است، با دقت ۹۵ درصد، عملکرد قابل‌قبولی داشت اما نسبت به مدل VGG16 در این پژوهش، دقت کمتری ارائه می‌دهد. همچنین مدلAlexNet با دقت ۸۰ درصد و مدل ZFNet با دقت ۷۰ درصد، هر دو با معماری‌های عمیق‌تر و پیچیده‌تر از LeNet-5 توسعه یافته‌اند اما باز هم نتوانستند به دقتی برابر با VGG16 دست یابند.

از طرف دیگر، مدل‌های جدیدتری مانند ALBERT با دقت ۹۹.۸۱ درصد و CutMix با دقت ۹۹.۸۴ درصد عملکرد بهتری از VGG16 در این پژوهش نشان داده‌اند اما باید توجه داشت که این مدل‌ها از تکنیک‌های پیچیده‌تری استفاده می‌کنند و برای کاربردهای خاصی بهینه‌سازی شده‌اند. مدل‌های EfficientNet و MixNet نیز با دقت‌های ۹۹.۵۳ و ۹۹.۵۶ درصد به‌ترتیب، عملکردی نزدیک به VGG16 داشتند اما VGG16 همچنان با توجه به سادگی نسبی خود، توانسته است دقت بسیار خوبی ارائه دهد.

به‌طور کلی، مقایسه نتایج این پژوهش با سایر کارهای انجام‌شده نشان می‌دهد که استفاده از مدل‌های پیش‌آموزش‌دیده و یادگیری انتقالی می‌تواند بهبود قابل‌توجهی در دقت و کارایی مدل‌های تشخیص اعداد دست‌نویس ایجاد کند، در حالی که نیاز به منابع محاسباتی و داده‌های آموزشی گسترده را کاهش می‌دهد. این رویکرد می‌تواند به عنوان یک راهکار مؤثر در مسائل مشابه در بینایی ماشین مورد استفاده قرار گیرد.

نتیجه‌گیری

تحقیقات انجام‌شده در زمینه تشخیص اعداد دست‌نویس با استفاده از شبکه‌های عصبی پیچشی نشان‌دهنده توانایی بالای این الگوریتم‌ها در استخراج و شناسایی ویژگی‌های پیچیده تصاویر است (برنگارت، ۲۰۲۳). از اولین مدل‌های معرفی‌شده مانند LeNet-5 گرفته تا مدل‌های پیچیده‌تری همچون AlexNet، ZFNet و VGGNet همگی بهبودهای قابل‌توجهی در دقت و کارایی تشخیص اعداد دست‌نویس در مجموعه دادهMNIST (لکان و همکاران، ۲۰۱۰) را نشان داده‌اند.

مدل LeNet-5 به‌عنوان نقطه شروعی برای استفاده از CNN در تشخیص اعداد دست‌نویس مطرح شد و اصول اساسی استفاده از لایه‌های پیچشی و تجمیع را معرفی کرد. به دنبال آن، AlexNet با معماری عمیق‌تر و استفاده از تکنیک‌های بهینه‌سازی جدید، عملکرد چشمگیری را در تشخیص تصاویر به نمایش گذاشت. ZFNet با بهبود معماری AlexNetو استفاده از تکنیک‌های بصری‌سازی توانست به درک بهتری از عملکرد لایه‌های پیچشی دست یابد و

ویژگی‌های مهم‌تری از تصاویر را استخراج کند. در نهایت، مدل VGGNet با استفاده از فیلترهای کوچک و عمق بیشتر، دقت بسیار بالایی در تشخیص اعداد دست‌نویس نشان داد و به‌طور گسترده‌ای در یادگیری انتقالی مورد استفاده قرار گرفت (برنگارت، ۲۰۲۳).

استفاده از یادگیری انتقالی با مدل‌های ازپیش‌آموزش‌دیده، مانند VGG16، روشی مؤثر برای کاهش نیاز به منابع محاسباتی و داده‌های بزرگ در آموزش مدل‌ها بوده است. این روش‌ها با انتقال ویژگی‌های یادگرفته‌شده از مجموعه داده‌های بزرگ به مجموعه داده‌های کوچک‌تر مانند MNIST (لکان و همکاران، ۲۰۱۰)، نه‌تنها دقت تشخیص را بهبود بخشیده‌اند بلکه زمان و هزینه آموزش مدل‌ها را نیز کاهش داده‌اند.

در مجموع، نتایج این تحقیقات نشان می‌دهند که شبکه‌های عصبی پیچشی و یادگیری انتقالی ابزارهای قدرتمندی برای تشخیص اعداد دست‌نویس هستند. این تکنیک‌ها با استفاده از ویژگی‌های محلی تصاویر و بهینه‌سازی معماری مدل‌ها، به دقت بالایی در طبقه‌بندی اعداد دست‌یافته‌اند. با ادامه پیشرفت‌ها در این زمینه، می‌توان انتظار داشت که مدل‌های یادگیری عمیق با دقت و کارایی بیشتری در مسائل مختلف بینایی ماشین به‌کار گرفته شوند.

این پژوهش با هدف بررسی و ارزیابی استفاده از یادگیری انتقالی و مدل VGG16 برای تشخیص اعداد دست‌نویس انجام شده است. ضرورت این تحقیق از آن‌جا ناشی می‌شود که با وجود دقت بالای مدل‌های CNN، نیاز به منابع محاسباتی و داده‌های آموزشی بزرگ همچنان معضلی جدی است. استفاده از یادگیری انتقالی می‌تواند این معضل را به میزان قابل‌توجهی کاهش دهد و مدل‌های دقیق‌تری را با منابع کمتری فراهم کند. هدف نهایی این پژوهش، بهبود دقت و کارایی مدل‌های تشخیص اعداد دست‌نویس با استفاده از روش‌های نوین یادگیری انتقالی است.

پیشنهادها

استفاده از مدل‌های پیشرفته‌تر: CNN با توجه به نتایج مثبت حاصل از استفاده از مدل‌های CNN برای تشخیص اعداد دست‌نویس، پیشنهاد می‌شود مدل‌های پیچیده‌تر و پیشرفته‌تری مانند ResNet و DenseNet نیز بررسی شوند. این مدل‌ها با عمق بیشتر و معماری‌های نوین می‌توانند دقت و کارایی بیشتری در تشخیص اعداد دست‌نویس ارائه دهند.

۱- **افزایش داده‌های آموزشی با استفاده از داده‌افزایی**[1]: یکی از روش‌های مؤثر برای بهبود دقت مدل‌ها، استفاده از تکنیک‌های داده‌افزایی است. با اعمال تغییرات مختلف مانند چرخش، تغییر مقیاس و جابه‌جایی به تصاویر ورودی، می‌توان تعداد نمونه‌های آموزشی را افزایش داد و مدل را به شکلی مقاوم‌تر در برابر تنوع داده‌ها آموزش داد.

۲- **بهینه‌سازی معماری شبکه:** بررسی و بهینه‌سازی معماری شبکه‌های عصبی می‌تواند منجر به بهبود عملکرد مدل‌ها شود. برای مثال، استفاده از تکنیک‌هایی مانند جستجوی معماری شبکه[2] می‌تواند معماری‌های بهینه‌تری را برای تشخیص اعداد دست‌نویس پیدا کند.

۳- **استفاده از تکنیک‌های منظم‌سازی**[3]**:** به‌منظور جلوگیری از بیش‌برازش[4] مدل‌ها به داده‌های آموزشی، استفاده از تکنیک‌های منظم‌سازی مانند Dropout و Batch Normalization توصیه می‌شود. این تکنیک‌ها می‌توانند عملکرد مدل‌ها را در مواجهه با داده‌های جدید بهبود بخشند.

۴- **یادگیری انتقالی با استفاده از مدل‌های پیش‌آموزش‌دیده جدیدتر:** استفاده از مدل‌های پیش‌آموزش‌دیده جدیدتر و قدرتمندتر مانند EfficientNet می‌تواند منجر به بهبود عملکرد مدل‌های

[1] Data Augmentation
[2] Neural Architecture Search
[3] Regularization
[4] Overfitting

تشخیص اعداد دست‌نویس شود. این مدل‌ها با بهره‌گیری از معماری‌های بهینه‌تر، می‌توانند دقت بالاتری را ارائه دهند.

۵- **استفاده از تکنیک‌های تفسیرپذیری مدل[1]:** با توجه به اهمیت تفسیرپذیری مدل‌های یادگیری عمیق، استفاده از تکنیک‌هایی مانند Grad-CAM و LIME برای بررسی و درک بهتر تصمیم‌گیری‌های مدل‌ها توصیه می‌شود. این تکنیک‌ها می‌توانند به شناسایی و رفع مشکلات مدل‌ها کمک کنند.

۶- **اجرای سیستم‌های ترکیبی:** ترکیب مدل‌های مختلف و ایجاد سیستم‌های ترکیبی[2] می‌تواند منجر به بهبود دقت و کارایی مدل‌ها شود. استفاده از روش‌های ترکیبی مانند Bagging و Boosting می‌تواند عملکرد مدل‌ها را بهبود بخشد.

۷- **استفاده از شبکه‌های مولد تخاصمی[3]:** برای افزایش داده‌های آموزشی و ایجاد نمونه‌های جدید و متنوع از اعداد دست‌نویس، استفاده از شبکه‌های مولد تخاصمی توصیه می‌شود. این روش می‌تواند به تولید داده‌های مصنوعی با کیفیت بالا کمک کند و مدل‌های تشخیص را بهبود بخشد.

۸- **توسعه سیستم‌های واقعی و اجرای عملی:** به‌منظور بررسی کاربردپذیری مدل‌های توسعه‌یافته در محیط‌های واقعی، اجرای عملی این مدل‌ها در سیستم‌های واقعی و ارزیابی عملکرد آنها در شرایط عملی توصیه می‌شود. این کار می‌تواند به شناسایی معضلات عملی و بهبود مدل‌ها کمک کند.

References

Aslani, S., & Jacob, J. (2023). Utilisation of deep learning for COVID-19 diagnosis. *Clinical Radiology*, *78*(2), 150-157. https://doi.org/10.1016/j.crad.2022.11.006

Azizi, S., Kornblith, S., Saharia, C., Norouzi, M., & Fleet, D. J. (2023). Synthetic data from diffusion models improves imagenet classification. *arXiv* 1-19. https://doi.org/10.48550/arXiv.2304.08466

Bakasa, W., & Viriri, S. (2023). VGG16 Feature Extractor with Extreme Gradient Boost Classifier for Pancreas Cancer Prediction. *Journal of Imaging*, *9*(7), 138. https://doi.org/10.3390/jimaging9070138

Berngardt, O. I. (2023). Improving Classification Neural Networks by using Absolute activation function (MNIST/LeNET-5 example). *arXiv*, 1-19. https://doi.org/10.48550/arXiv.2304.11758

Chandure, S., & Inamdar, V. (2023). Handwritten MODI Character Recognition Using Transfer Learning with Discriminant Feature Analysis. *Institution of Electronics and Telecommunication Engineers Journal of Research*, *69*(5), 2584-2594. https://doi.org/10.1080/03772063.2021.1902867

Chen, H., Luo, H., Huang, B., Jiang, B., & Kaynak, O. (2024). Transfer Learning-Motivated Intelligent Fault Diagnosis Designs: A Survey, Insights, and Perspectives. *Institute of Electrical and Electronics Engineers Transactions on Neural Networks and Learning Systems*, *35*(3), 2969-2983. https://doi.org/10.1109/TNNLS.2023.3290974

Cireşan, D. C., Meier, U., Gambardella, L. M., & Schmidhuber, J. (2010). Deep, Big, Simple Neural Nets for Handwritten Digit Recognition. *Neural Computation*, *22*(12), 3207-3220. https://doi.org/10.1162/NECO_a_00052

[1] Model Interpretability
[2] Ensemble Systems
[3] GANs

Cohen, G., Afshar, S., Tapson, J., & Schaik, A. V. (2017, May 14-19). *EMNIST: Extending MNIST to handwritten letters* [Conference session]. 2017 International Joint Conference on Neural Networks Anchorage, Alaska, USA. https://doi.org/10.1109/IJCNN.2017.7966217

Deng, J., Dong, W., Socher, R., Li, L. J., Kai, L., & Li, F-F. (2009, June 20-25). *ImageNet: A large-scale hierarchical image database* [Conference session]. 2009 Institute of Electrical and Electronics Engineers Conference on Computer Vision and Pattern Recognition, Miami, Florida, USA. https://doi.org/10.1109/CVPR.2009.5206848

Fateh, A., Fateh, M., & Abolghasemi, V. (2021). Multilingual handwritten numeral recognition using a robust deep network joint with transfer learning. *Information Sciences, 581*(3), 479-494. https://doi.org/10.1016/j.ins.2021.09.051

Ghaffarian, H., & Bamohabbat, A. R. (2023). Classification and Prediction of Customer Categories Using Combination of LRFM Method, Quartiles and Multi-class Data Mining Methods. *Quarterly Scientific Journal of Technical and Vocational University*, *20*(1), 511-532. https://doi.org/10.48301/kssa.2022.316104.1852

Hassan, E., Hossain, M. S., Saber, A., Elmougy, S., Ghoneim, A., & Muhammad, G. (2024). A quantum convolutional network and ResNet (50)-based classification architecture for the MNIST medical dataset. *Biomedical Signal Processing and Control*, *87*(7792), 105560. https://doi.org/10.1016/j.bspc.2023.105560

Iman, M., Arabnia, H. R., & Rasheed, K. (2023). A Review of Deep Transfer Learning and Recent Advancements. *Technologies*, *11*(2), 40. https://doi.org/10.3390/technologies11020040

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, *25*(2), 1-9. https://doi.org/10.1145/3065386

Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., & Soricut, R. (2019, May 6-9). *Albert: A lite bert for self-supervised learning of language representations* [Conference session]. International Conference on Learning Representations, New Orleans, Louisiana, United States. https://doi.org/10.48550/arXiv.1909.11942

Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the Institute of Electrical and Electronics Engineers*, *86*(11), 2278-2324. https://doi.org/10.1109/5.726791

Lecun, Y., Cortes, C., & Burges, C. J. (2010). *MNIST handwritten digit database* [Data set]. AT&T Labs. http://yann.lecun.com/exdb/mnist

Lejeune, E. (2020). Mechanical MNIST: A benchmark dataset for mechanical metamodels. *Extreme Mechanics Letters*, *36*, 100659. https://doi.org/10.1016/j.eml.2020.100659

Namjouye Rad, A. A., & Dadgarpour, M. (2021). Detection of network penetration by data mining and using machine learning via SVM algorithm. *Quarterly Scientific Journal of Technical and Vocational University*, *17*(4), 13-34. https://doi.org/10.48301/kssa.2021.128393

Rudregowda, S., Patil Kulkarni, S., H L, G., Ravi, V., & Krichen, M. (2023). Visual Speech Recognition for Kannada Language Using VGG16 Convolutional Neural Network. *Acoustics*, *5*(1), 343-353. https://doi.org/10.3390/acoustics5010020

Salehi, A. W., Khan, S., Gupta, G., Alabduallah, B. I., Almjally, A., Alsolai, H., Siddiqui, T., & Mellit, A. (2023). A Study of CNN and Transfer Learning in Medical Imaging: Advantages, Challenges, Future Scope. *Sustainability*, *15*(7), 5930. https://doi.org/10.3390/su15075930

Shang, S., Shan, Z., Liu, G., Wang, L., Wang, X., Zhang, Z., & Zhang, J. (2024, February 20-27). *Resdiff: Combining Cnn and Diffusion Model for Image Super-resolution* [Conference session]. Proceedings of the Association for the Advancement of Artificial Intelligence Conference on Artificial Intelligence, Vancouver, Canada. http://dx.doi.org/10.13140/RG.2.2.22060.13444

Simonyan, K., & Zisserman, A. (2014, May 7-9). *Very deep convolutional networks for large-scale image recognition* [Conference session]. International Conference on Learning Representations, San Diego, California. https://doi.org/10.48550/arXiv.1409.1556

Taghandiki, K. (2023). Implementation of a Noisy Hyperlink Removal System: Using the Semantic and Relational Approach of the DBpedia Ontology. *Quarterly Scientific Journal of Technical and Vocational University*, *20*(3), 485-507. https://doi.org/10.48301/kssa.2023.382583.2426

Taghandiki, K., Ahmadi, M. H., & Ehsan, E. R. (2023). Automatic summarisation of Instagram social network posts Combining semantic and statistical approaches. *arXiv* 1-7. https://doi.org/10.48550/arXiv.2303.07957

Tan, M., & Le, Q. (2019, Jun 9-15). *Efficientnet: Rethinking model scaling for convolutional neural networks* [Conference session]. International conference on machine learning, Long Beach, California, USA. https://proceedings.mlr.press/v97/tan19a.html?ref=jina-ai-gmbh.ghost.io

Tan, M., & Le, Q. V. (2019). Mixconv: Mixed depthwise convolutional kernels. *arXiv*, 1-13. https://doi.org/10.48550/arXiv.1907.09595

Taye, M. M. (2023). Understanding of Machine Learning with Deep Learning: Architectures, Workflow, Applications and Future Directions. *Computers*, *12*(5), 91. https://doi.org/10.3390/computers12050091

Theodoris, C. V., Xiao, L., Chopra, A., Chaffin, M. D., Al Sayed, Z. R., Hill, M. C., Mantineo, H., Brydon, E. M., Zeng, Z., Liu, X. S., & Ellinor, P. T. (2023). Transfer learning enables predictions in network biology. *Nature*, *618*(7965), 616-624. https://doi.org/10.1038/s41586-023-06139-9

Yun, S., Han, D., Oh, S. J., Chun, S., Choe, J., & Yoo, Y. (2019, October 27- November 02). *Cutmix: Regularization strategy to train strong classifiers with localizable features* [Conference session]. Proceedings of the Institute of Electrical and Electronics Engineers/ International Conference on Computer Vision international conference on computer vision, Seoul, Korea (South). https://doi.org/10.1109/ICCV.2019.00612

Zeiler, M. D., & Fergus, R. (2014). Visualizing and Understanding Convolutional Networks. In D. Fleet, T. Pajdla, B. Schiele, & T. Tuytelaars (Eds.), *Computer Vision – European Conference on Computer Vision 2014* (pp. 818-833). Springer International Publishing. https://doi.org/10.1007/978-3-319-10590-1_53