Enhanced Fashion-MNIST Classification Using a Hybrid VGG-16 DenseNet121 Architecture

[1]Gheri Febri Ananda, [2] Risfendra, [3] Eko Wahyudi
[1,3]Department of Electrical and Informatics Engineering, Universitas Gadjah Mada, Indonesia
[2]Department of Electrical Engineering, Universitas Negeri Padang, Indonesia
Email: [1]gherifebriananda1998@mail.ugm.ac.id, [2]risfendra@ft.unp.ac.id, [3]ekowahyudi@mail.ugm.ac.id

Article Info
Article history:
Received Aug 18th, 2024
Revised Oct 17th, 2024
Accepted Nov 4th, 2024

Keyword:
Comparative Architecture
DenseNet121
Enhanced Fashion
Fashion MNIST
VGG16

ABSTRACT

This study aims to explore the effectiveness of a hybrid model combining the VGG16 and DenseNet121 architectures for image classification tasks on the Fashion MNIST dataset. This model is designed to leverage the advantages of both architectures to produce richer feature representations. In this study, the performance of the hybrid model is compared with several other architectures, including LeNet-5, VGG-16, ResNet-20, ResNet-50, EfficientNet-B0, and DenseNet-121, using various optimizers such as Adam, RMSProp, AdaDelta, AdaGrad and SGD. The test results indicate that the Adam and SGD optimizers deliver excellent results. The VGG16 + DenseNet121 hybrid model achieved perfect training accuracy 100%, the highest validation accuracy 94.65%, and excellent test accuracy 94.16%. Confusion matrix analysis confirms that this model is capable of correctly classifying the majority of images, although there is some confusion between classes with visual similarities. These findings affirm that a hybrid approach and the appropriate selection of optimizers can significantly enhance model performance in image classification tasks.

Corresponding Author:
Gheri Febri Ananda,
Departement of Electrical and Information Engineering,
Universitas Gadjah Mada,
Jl. Grafika No. 2 Kampus UGM, Yogyakarta, Indonesia
Email: gherifebriananda1998@mail.ugm.ac.id

DOI: http://dx.doi.org/10.24014/ijaidm.v8i1.32225

1. INTRODUCTION

In recent decades, Convolutional Neural Networks (CNNs) have become fundamental in image classification and processing, particularly when dealing with complex datasets like Fashion-MNIST. This dataset, developed by Zalando Research, includes 70,000 clothing images classified into 10 categories, providing a more realistic and challenging alternative to the traditional MNIST dataset. The primary aim of Fashion-MNIST is to address the limitations of the MNIST dataset in representing contemporary image classification challenges by offering greater variation and complexity, which are more representative of real-world applications [1], [2].

In response to these challenges, various CNN architectures have been developed. LeNet-5, one of the earliest CNN architectures, uses simple yet efficient convolutional and pooling layers to extract features from images [3]. Despite being an older design, LeNet-5 remains relevant for simple image classification applications due to its efficiency. Other architectures, such as VGG16, introduced the use of sequential 3x3 convolutional layers to extract deeper and more complex features, demonstrating impressive performance on large-scale image classification tasks like ImageNet [4].

Further innovations in architecture, such as ResNet, introduced the concept of residual connections, enabling the training of deeper networks without encountering vanishing gradient issues and enhancing the model's ability to handle more complex image classification tasks [5], [6]. Recent approaches like EfficientNet integrate model scaling by simultaneously considering depth, width, and input resolution, effectively reducing the number of parameters required without sacrificing accuracy, making it ideal for applications requiring high computational efficiency [7], [8]. DenseNet, with its novel approach to layer connections, allows for effective reuse of features and reduces the vanishing gradient problem, demonstrating superior results in various image classification tasks [9].

Although CNN architectures such as VGG, ResNet, and DenseNet have been widely used in various image classification applications, this research focuses specifically on the Fashion-MNIST dataset. These architectures have demonstrated their success in numerous cases beyond Fashion-MNIST. For instance, VGG and ResNet have been evaluated in studies comparing their performance in image classification across diverse datasets, achieving significant accuracy levels in various application domains, including medical image recognition and product classification [10]. DenseNet and EfficientNet, designed to enhance efficiency and effectiveness in parameter usage and data processing, have been successfully implemented in disease detection and classification tasks in guava plants, proving their capability in handling complex image classification tasks [11].

However, despite advancements in CNN architectures, there are still limitations in performance, especially on complex datasets like Fashion-MNIST. Research using the VGG-11 model for Fashion-MNIST classification found that while adding batch normalization layers improved performance, the achieved accuracy was only 91.5% [12]. Other studies have shown that while some CNN architectures achieve high accuracy on the MNIST dataset, their performance on the Fashion-MNIST dataset still indicates lower accuracy levels [13].

This research proposes a hybrid approach by combining the VGG-16 and DenseNet121 architectures. The primary reason for choosing this combination is that both architectures have complementary strengths that can significantly impact data modeling. VGG-16 is known for its capability to extract deep features from images through the use of sequential 3x3 convolutional layers, enabling the model to learn highly detailed and complex representations from image data [14]. On the other hand, DenseNet121 introduces an innovative approach to connecting its layers, allowing for more effective feature reuse and addressing the vanishing gradient problem [15]. This results in reduced parameter requirements and improved convergence speed.

Overall, comprehensive research comparing various CNN architectures and optimization techniques on complex datasets like Fashion-MNIST is still limited. Discussions on the effectiveness of different optimizers highlight the importance of optimization in enhancing model performance [16], [17]. Optimizers such as Adam, RMSprop, AdaDelta, Adagrad, and SGD have characteristics that can significantly influence model convergence and performance [18], [19].

Therefore, this study aims to make two main contributions in the context of Fashion-MNIST classification, with a focus on improving accuracy. First, we compare several advanced architectures in deep learning such as VGG16, LeNet-5, ResNet-20, ResNet-50, EfficientNetB0, and DenseNet121. We propose a hybrid approach that combines the strengths of VGG-16 and DenseNet121 to improve classification accuracy. This approach, named Vgg16 + DenseNet121, is designed to improve classification accuracy by integrating the feature extraction capabilities of VGG-16 with the efficiency of DenseNet121 in solving the problem of vanishing gradients. The second contribution is to provide valuable insights into selecting the most effective optimizer for similar image classification challenges, offering practical guidance for practitioners in searching for the best solutions for real-world applications.

2. RESEARCH METHOD

2.1. Fashion MNIST Dataset

The Fashion-MNIST dataset developed by Zalando Research serves as a standard benchmark in image classification, especially for the fashion and clothing categories. This dataset includes 70,000 grayscale images distributed across 10 different classes, each representing a unique type of clothing or accessory. Fashion-MNIST is introduced as a more challenging and realistic alternative to traditional MNIST datasets, which feature complex images that pose greater challenges for image classification algorithms. Each image is 28 x 28 pixels, similar in size to the original MNIST, but offering more diversity in terms of patterns, textures and shapes.

This dataset is divided into two main subsets: 60,000 images for training and 10,000 images for testing. Additionally, during training, the training data is further divided with a 0.1 proportion for validation, aiming to evaluate the model's performance more accurately during the training process. This partitioning allows for effective evaluation of the trained classification model's performance. A detailed description of the Fashion-MNIST dataset is presented in Figure 1.

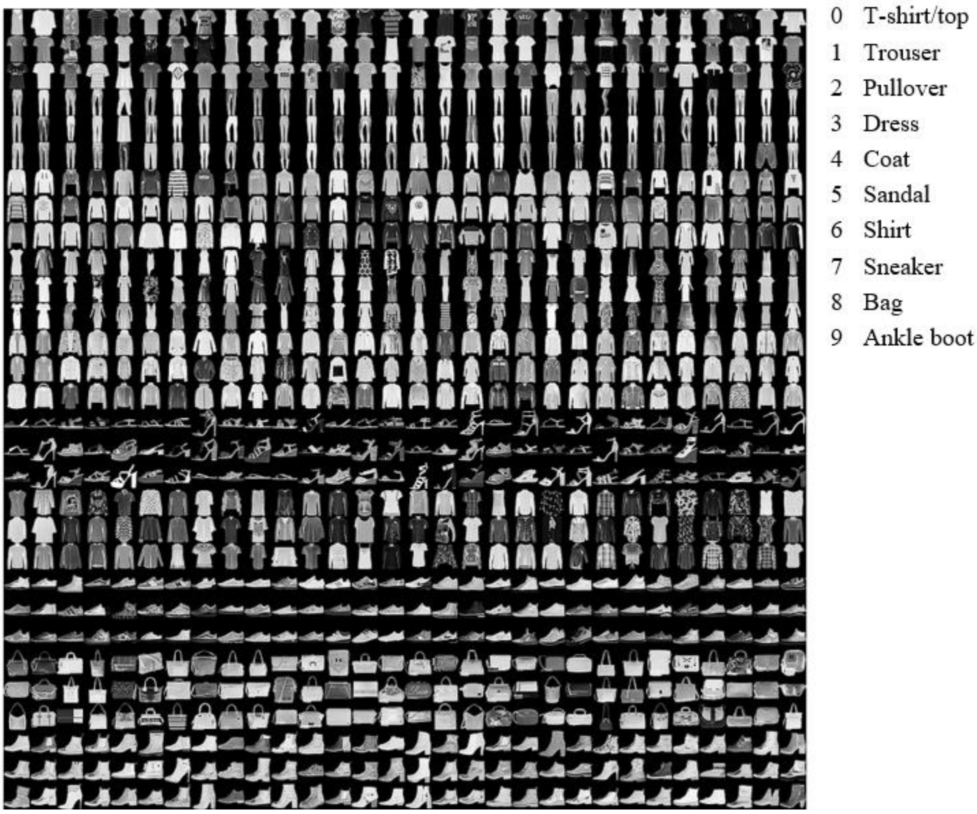


Figure 1. Examples Of Fashion Mnist Dataset

In this research, Fashion-MNIST is selected over the standard MNIST dataset because it offers a more complex set of challenges, featuring diverse clothing and accessory categories. This enables a thorough testing of various CNN architectures and optimization techniques. The use of Fashion-MNIST aids in evaluating model performance more deeply and helps develop effective solutions for complex visual object recognition, with potential applications in e-commerce, manufacturing, and visual security industries.

2.2. Model Architecture

In this study, we explore and compare several convolutional neural network (CNN) architectures that have proven effective in image classification tasks, including LeNet-5, VGG-16, ResNet, DenseNet, and EfficientNet To further enhance performance, we propose a hybrid method that combines the strengths of VGG-16 and DenseNet. This hybrid approach leverages VGG-16's capability to extract detailed and complex features from images, alongside DenseNet's efficiency in parameter usage and its effectiveness in managing very deep network architectures [20]. By integrating these two architectures, we aim to create a more robust solution tailored to the challenges posed by fashion image classification tasks, particularly within the Fashion-MNIST dataset.

The Fashion-MNIST dataset presents unique classification challenges due to its wide variety of clothing items and intricate visual details. Our hybrid model is specifically designed to address these complexities, improving accuracy in distinguishing between various fashion items. The proposed architecture, shown in Figure 2, illustrates the integration of VGG-16 and DenseNet components, effectively combining their complementary strengths to deliver a more optimal solution for fashion image classification.

Based on Figure 2, it can be seen that this study proposes a hybrid architectural model using two input layers, each with a shape of (None, 32, 32, 3), representing a batch of 32x32 pixel images with 3 color channels (RGB). The model consists of two main branches: VGG16 and DenseNet121. The first branch is the VGG16 model, adapted to accept input without the top layers (fully connected layers), producing an output with dimensions (None, 1, 1, 512). The second branch is the DenseNet121 model, also adapted to extract features from the same images, producing an output with dimensions (None, 1, 1, 1024). Both outputs are then processed through a Global Average Pooling layer, reducing the spatial dimensions to (None, 512) for VGG16 and (None, 1024) for DenseNet121, respectively.

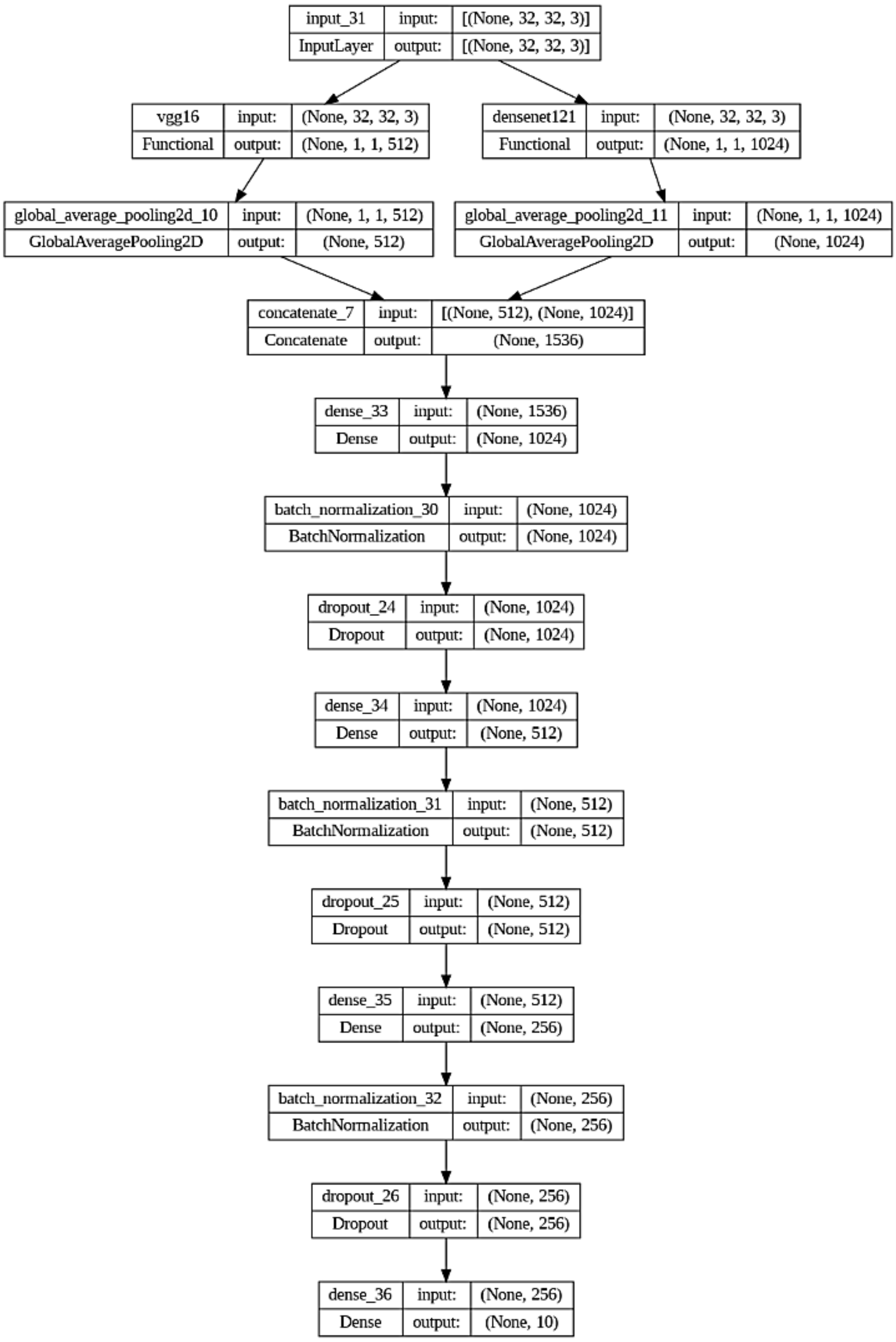


Figure 2. Proposed hybrid VGG16 + DenseNet121 architecture

After these features are extracted by both CNN architectures, they are combined through a concatenation layer, resulting in a tensor with the shape (None, 1536), which is a combination of the features learned by both models. This tensor is then passed through a series of fully connected layers. The first Dense layer has 1024 units with a ReLU activation function, followed by Batch Normalization and Dropout with a rate of 0.5 to reduce the risk of overfitting. The second Dense layer consists of 512 units, also with ReLU activation, and is followed by the same batch normalization and dropout process. The third Dense layer has 256 units and follows consistent normalization and dropout processes like the previous layers.

At the end of this architecture, there is an output Dense layer with 10 units representing the 10 classes in the Fashion-MNIST dataset. This layer uses the softmax activation function to produce a probability distribution for each class, enabling the model to accurately classify the input images. The combination of these two models is expected to enhance accuracy and efficiency in classifying the complex and diverse fashion images.

2.3. Training Models

During the model training phase, we conducted an experimental approach to assess the efficacy of various optimizers in the context of Fashion-MNIST, employing the hybrid VGG16-DenseNet121 architecture. Additionally, several different architectural models were experimented with to compare their performance. All models underwent training for 50 epochs to evaluate convergence and long-term performance in image classification. To ensure methodological rigor and comparability, all models were trained under uniform conditions, including consistent epoch numbers and batch sizes. The evaluated optimizers encompassed Adam, RMSprop, SGD, Adadelta, and Adagrad, selected based on their effectiveness and established usage across diverse machine learning tasks.

Each optimizer was configured with default parameters to achieve optimal outcomes, as per relevant scholarly literature. This study aims to investigate how the selection of optimizer impacts model performance on intricate image datasets such as Fashion-MNIST, with a primary objective of discerning the most efficacious strategies to enhance classification accuracy for practical applications.

2.4. Evaluations Models

To evaluate the effectiveness of the proposed hybrid VGG16-DenseNet model, we utilized three key metrics: accuracy, precision, and recall. Accuracy measures the overall correctness of the model and is calculated as the ratio of correctly predicted instances to the total instances. Precision assesses the correctness of positive predictions and is computed as the ratio of true positive predictions to the sum of true positive and false positive predictions. Recall, also known as sensitivity, gauges the model's ability to identify all relevant instances and is calculated as the ratio of true positive predictions to the sum of true positive and false negative predictions [21], [22]. Evaluation was conducted on the training, validation, and test datasets to ensure the model's generalizability to previously unseen data.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3}$$

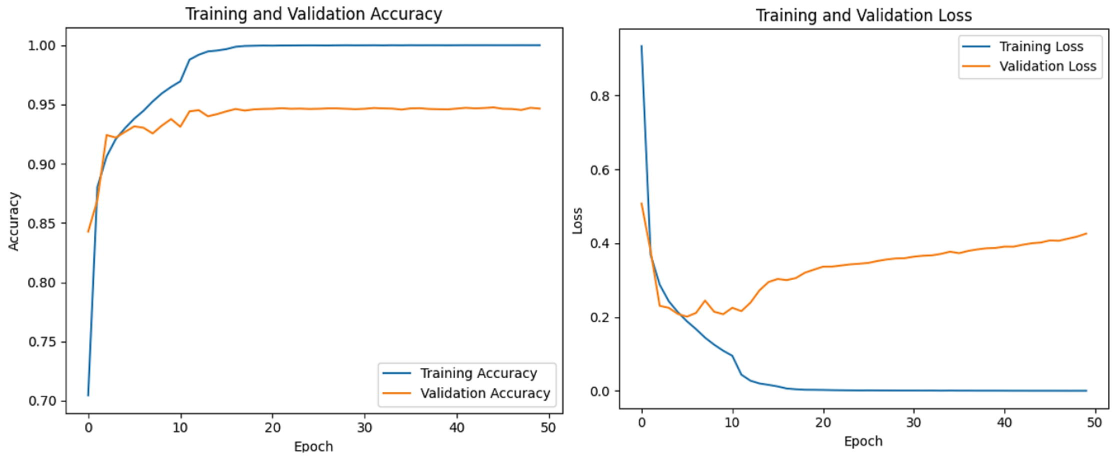
Specifically for the proposed hybrid model, confusion matrix analysis was used to provide deep insights into classification performance for each class. The confusion matrix helps identify error patterns and classes that may require further adjustment. Additionally, the performance of various optimizers used during training (Adam, RMSprop, SGD, Adadelta, and Adagrad) was compared to determine their impact on model convergence efficiency. Observing changes in validation loss during each epoch aids in assessing the risk of overfitting or underfitting. This evaluation is expected to verify the effectiveness of the proposed hybrid approach and provide practical guidance for further optimization, thereby contributing to the development of more robust and accurate image classification models.

3. RESULTS AND ANALYSIS

3.1. Results

This study employs a sophisticated approach by integrating the VGG16 and DenseNet121 architectures to perform classification tasks on the Fashion MNIST dataset, encompassing diverse categories such as t-shirts, trousers, pullovers, dresses, coats, sandals, shirts, sneakers, bags, and ankle boots. The methodology leverages transfer learning, initially freezing the layers of VGG16 and DenseNet121 to preserve previously acquired features, followed by fine-tuning the top layers with the Fashion MNIST dataset. This strategic amalgamation capitalizes on the unique strengths of each model, thereby significantly boosting the accuracy and robustness of classification outcomes. The training process's progression in terms of accuracy and loss is depicted in Figure 3, providing insights into the model's convergence and performance refinement over epochs.

The evaluation results from the training process indicate that the hybrid VGG16 and DenseNet121 model achieved high testing accuracy with low loss values. The training and validation accuracy graphs, as shown in Figure 2 above, illustrate a significant increase in accuracy over 50 epochs. The accuracy graph in Figure 3(a) shows that the training accuracy increased rapidly, reaching over 95% within the first 10 epochs, and approached 100% by the end of the training period. The validation accuracy also showed stable results, ranging between 90% and 94%, indicating good model performance on unseen data during training.



(a) (b)

Figure 3. Plot of (a) accuracy and (b) loss function during the training proces

The loss graph in Figure 3(b) demonstrates a significant decrease in training loss, approaching zero, indicating that the model successfully adapted to the training data. Meanwhile, the validation loss initially showed a declining pattern and then stabilized with minor fluctuations after the 20th epoch, indicating that the model has a good capacity to handle validation data. Overall, the hybrid VGG16 and DenseNet121 model exhibited excellent and consistent training performance on the Fashion MNIST dataset, demonstrating promising results in classifying fashion items and showing good generalization capability on validation data. For a clearer understanding of the model's performance, the confusion matrix results are shown in Figure 4.

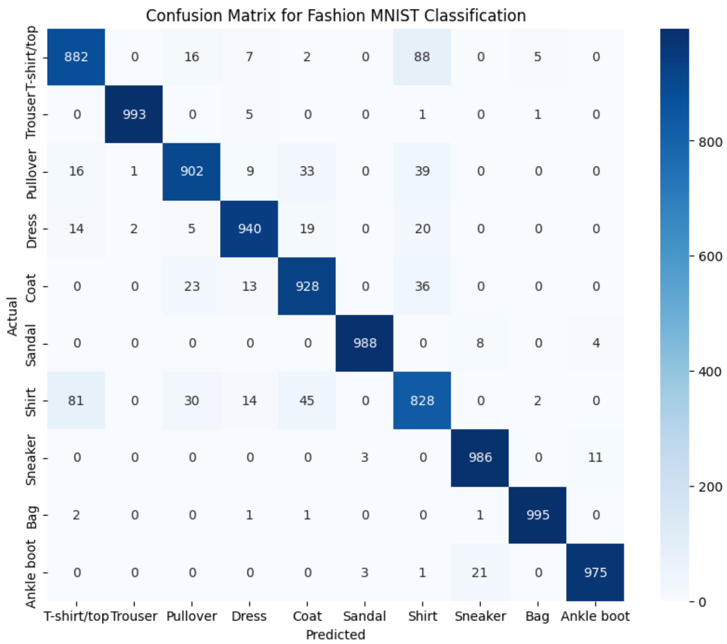


Figure 4. Coffusion Matriks Result

The confusion matrix above shows the performance of the hybrid VGG16 + DenseNet121 model in classifying the Fashion MNIST dataset. This matrix displays the number of correct and incorrect predictions for each class, encompassing ten categories: T-shirt/top, Trouser, Pullover, Dress, Coat, Sandal, Shirt, Sneaker, Bag, and Ankle boot. A summary table of the evaluation for each class can be seen in Figure 5.

The classification model applied to the Fashion-MNIST dataset demonstrates high performance with an overall accuracy of 94%. The analysis of precision, recall, and F1-score metrics shows excellent performance, particularly in the Trouser category, which achieved perfect precision and recall of 1.00 and an F1-score of 0.99. However, the Shirt category showed slightly lower performance with a precision and F1-score of 0.82, indicating challenges in classifying this type of clothing. The macro and weighted averages for precision, recall, and F1-score are all 0.94, indicating that the model has consistent performance and is not biased towards any particular class. These findings suggest that the model can effectively classify various types of clothing in this dataset, although special attention is needed for the Shirt category to achieve further improvements.

	precision	recall	f1-score	support
T-shirt/top	0.89	0.88	0.88	1000
Trouser	1.00	0.99	0.99	1000
Pullover	0.92	0.90	0.91	1000
Dress	0.95	0.94	0.95	1000
Coat	0.90	0.93	0.92	1000
Sandal	0.99	0.99	0.99	1000
Shirt	0.82	0.83	0.82	1000
Sneaker	0.97	0.99	0.98	1000
Bag	0.99	0.99	0.99	1000
Ankle boot	0.98	0.97	0.98	1000
accuracy			0.94	10000
macro avg	0.94	0.94	0.94	10000
weighted avg	0.94	0.94	0.94	10000

Figure 5. Metric Summary

3.2. Comparison of Several Architecture

In this study, various convolutional neural network (CNN) architectures were compared to evaluate their performance in classifying the Fashion-MNIST dataset. The architectures tested include LeNet-5, VGG-16, ResNet-20, ResNet-50, EfficientNet-B0, DenseNet-121, and the proposed hybrid model VGG16+DenseNet121. The evaluation metrics include training accuracy, training loss, validation accuracy, validation loss, test accuracy, and training time. The table below summarizes the results of these architectures when trained for 50 epochs.

Table 1. Testing some Architectures

Architecture	Train Accuracy	Train Loss	Val Accuracy	Val loss	Test Accuracy	Time (Minute)
LeNet-5	0.972	0.072	0.901	0.451	0.903	4
VGG-16	1.00	0.0004	0.942	0.867	0.9314	28
ResNet-20	0.992	0.022	0.914	0.435	0.913	25
ResNet-50	0.999	0.0004	0.940	0.536	0.932	38
EfficientNet-B0	0.978	0.059	0.915	0.348	0.899	34
DenseNet-121	0.999	0.002	0.936	0.512	0.930	56
Vgg16 + DenseNet 121 (Our)	1.000	0.003	0.9465	0.426	0.9416	59

The table 1 shows the performance comparison of various CNN architectures on the Fashion-MNIST dataset. LeNet-5, a simple architecture, demonstrated good results with a test accuracy of 90.35% and very short training time. VGG-16 and ResNet-50 achieved perfect training accuracy; however, VGG-16 experienced relatively high validation loss, indicating challenges in generalization. ResNet-20 and ResNet-50 showed a balance between accuracy and training time, with ResNet-50 slightly superior in validation and test accuracy.

EfficientNet-B0, despite having good training accuracy, showed slightly lower test accuracy compared to other architectures. DenseNet-121 displayed excellent performance with a validation accuracy of 93.63% and test accuracy of 93.04%, although it required a longer training time. The proposed hybrid model VGG16 + DenseNet121 showed the best results with the highest validation accuracy of 94.65% and test accuracy of 94.16%, despite the longest training time. Overall, the hybrid VGG16 + DenseNet121 model demonstrated the most promising results in terms of validation and test accuracy, despite requiring longer training time. This architecture offers a good balance between accuracy and generalization capability, making it suitable for complex image classification applications.

3.3. Comparison of Several Optimizers on the Proposed Architectural Model

In the advanced stage of this research, we conducted an in-depth analysis of several optimizers used to train the hybrid architecture VGG16 + DenseNet121 on the Fashion-MNIST dataset. The evaluated optimizers include Adam, SGD, RMSProp, AdaDelta, and AdaGrad, each employing unique optimization approaches during the model training process. This analysis aims to gain a deep understanding of how each optimizer impacts the model's performance in understanding and generalizing patterns from the Fashion-MNIST dataset. The results of this comparison are expected to identify the most optimal optimizer for the proposed method on the Fashion-MNIST dataset. Table 2 below presents the comparison results of these optimizers based on key metrics such as training accuracy, training loss, validation accuracy, validation loss, and test accuracy.

Table 2. Testing Several Optimization Methods

Architecture	Train Accuracy	Train Loss	Val Accuracy	Val_loss	Test Accuracy
Adam	1.0000	0.0003	0.9465	0.4261	0.9416
SGD	0.9916	0.0313	0.9377	0.2437	0.9347
RMSProp	0.7913	0.5248	0.8788	0.4043	0.8783
AdaDelta	0.7907	0.5279	0.8837	0.3976	0.8837
AdaGrad	0.9497	0.1554	0.2109	0.9300	0.9286

The table above shows the performance comparison of various optimizers in training the hybrid VGG16 + DenseNet121 architecture on the Fashion-MNIST dataset. The Adam optimizer showed the best results with a training accuracy of 100%, validation accuracy of 94.65%, and test accuracy of 94.16%. This indicates that Adam can optimize the model very well. SGD also demonstrated high performance with a training accuracy of 99.16%, validation accuracy of 93.77%, and test accuracy of 93.47%, maintaining a good balance between training and validation. RMSProp and AdaDelta showed fairly good results with validation and test accuracy around 88%, but lower compared to Adam and SGD. AdaGrad had a training accuracy of 94.97% and validation accuracy of 93.00%, but the test accuracy was slightly lower (92.86%). Overall, Adam and SGD are the most effective optimizers for the hybrid VGG16 + DenseNet121 model on the Fashion-MNIST dataset, with Adam being the best.

4. CONCLUSION AND FUTURE WORK

This study demonstrates that the hybrid VGG16 + DenseNet121 model with the Adam optimizer provides the best performance in image classification on the Fashion MNIST dataset compared to several other architectural models. This model achieved perfect training accuracy (100%), the highest validation accuracy (94.65%), and excellent test accuracy (94.16%). The confusion matrix analysis confirms that this model can classify most images correctly, although there is some confusion between classes with visual similarities. Overall, using a hybrid architecture that combines the strengths of two different CNN models has proven effective in improving classification accuracy. The Adam optimizer has proven to be the most efficient in achieving these results, showing superior stability and optimization capability. These results provide strong evidence that a hybrid approach and the appropriate choice of optimizer can significantly enhance model performance in image classification tasks.

Although the hybrid VGG16 + DenseNet121 model shows excellent performance, several challenges remain. Significant confusion between classes with visual similarities, such as T-shirt/top and Shirt, or Pullover and Coat, indicates that the model can still be improved in terms of generalization across similar classes. Additionally, the longer training time for this hybrid model can be a hindrance in practical implementation, especially on hardware with limited resources. To address these challenges, it is recommended that future investigations focus on adopting advanced data augmentation methods and applying more rigorous regularization approaches to enhance the model's generalization abilities. Additionally, exploring various optimizers and optimizing hyperparameter configurations could significantly improve performance. Assessing the model on larger and more varied datasets is crucial to ensure its robustness in diverse real-world conditions. Furthermore, future research should consider deploying the model in environments with constrained hardware resources to evaluate its practical performance and efficiency.

REFERENCES

[1] H. Xiao, K. Rasul, and R. Vollgraf, "Fashion-MNIST: a Novel Image Dataset for Benchmarking Machine Learning Algorithms," Sep. 15, 2017, *arXiv*: arXiv:1708.07747. Accessed: Jul. 02, 2024. [Online]. Available: http://arxiv.org/abs/1708.07747

[2] M. Kayed, A. Anter, and H. Mohamed, "Classification of Garments from Fashion MNIST Dataset Using CNN LeNet-5 Architecture," in *2020 International Conference on Innovative Trends in Communication and Computer Engineering (ITCE)*, Aswan, Egypt: IEEE, Feb. 2020, pp. 238–243. doi: 10.1109/ITCE48509.2020.9047776.

[3] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998, doi: 10.1109/5.726791.

[4] T. Kaur and T. K. Gandhi, "Automated Brain Image Classification Based on VGG-16 and Transfer Learning," in *2019 International Conference on Information Technology (ICIT)*, Dec. 2019, pp. 94–98. doi: 10.1109/ICIT48102.2019.00023.

[5] S. Targ, D. Almeida, and K. Lyman, "Resnet in Resnet: Generalizing Residual Architectures," Mar. 25, 2016, *arXiv*: arXiv:1603.08029. doi: 10.48550/arXiv.1603.08029.

[6] S. A. Hasanah, A. A. Pravitasari, A. S. Abdullah, I. N. Yulita, and M. H. Asnawi, “A Deep Learning Review of ResNet Architecture for Lung Disease Identification in CXR Image,” *Appl. Sci.*, vol. 13, no. 24, p. 13111, Dec. 2023, doi: 10.3390/app132413111.

[7] M. Tan and Q. Le, “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks,” in *Proceedings of the 36th International Conference on Machine Learning*, PMLR, May 2019, pp. 6105–6114. Accessed: Jul. 02, 2024. [Online]. Available: https://proceedings.mlr.press/v97/tan19a.html

[8] T. Ahmad *et al.*, “EfficientNet-Based Robust Recognition of Peach Plant Diseases in Field Images,” *Comput. Mater. Contin.*, vol. 71, no. 1, pp. 2073–2089, 2022, doi: 10.32604/cmc.2022.018961.

[9] N. Hasan, Y. Bao, A. Shawon, and Y. Huang, “DenseNet Convolutional Neural Networks Application for Predicting COVID-19 Using CT Image,” *SN Comput. Sci.*, vol. 2, no. 5, p. 389, Jul. 2021, doi: 10.1007/s42979-021-00782-7.

[10] N. Zakaria, F. Mohamed, R. Abdelghani, and K. Sundaraj, “VGG16, ResNet-50, and GoogLeNet Deep Learning Architecture for Breathing Sound Classification: A Comparative Study,” in *2021 International Conference on Artificial Intelligence for Cyber Security Systems and Privacy (AI-CSP)*, Nov. 2021, pp. 1–6. doi: 10.1109/AI-CSP52968.2021.9671124.

[11] S. Rana, M. Aeri, V. Kukreja, and R. Sharma, “Integrating EfficientNet and Fine-Tuned DenseNet Models for Advanced Detection and Classification of Guava Diseases,” in *2024 International Conference on Emerging Technologies in Computer Science for Interdisciplinary Applications (ICETCS)*, Apr. 2024, pp. 1–5. doi: 10.1109/ICETCS61022.2024.10543286.

[12] C. Duan, P. Yin, Y. Zhi, and X. Li, “Image Classification of Fashion-mnist Data Set Based on VGG Network”.

[13] S. S. Kadam, A. C. Adamuthe, and A. B. Patil, “CNN Model for Image Classification on MNIST and Fashion-MNIST Dataset,” *J. Sci. Res.*, vol. 64, no. 02, pp. 374–384, 2020, doi: 10.37398/JSR.2020.640251.

[14] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” Apr. 10, 2015, *arXiv*: arXiv:1409.1556. doi: 10.48550/arXiv.1409.1556.

[15] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, “Densely Connected Convolutional Networks,” presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4700–4708. Accessed: Sep. 18, 2024. [Online]. Available: https://openaccess.thecvf.com/content_cvpr_2017/html/Huang_Densely_Connected_Convolutional_CVPR_2017_paper.html

[16] N. S. Mahajaya, P. D. W. Ayu, and R. R. Huizen, “Pengaruh Optimizer Adam, AdamW, SGD, dan LAMB terhadap Model Vision Transformer pada Klasifikasi Penyakit Paru-paru,” *Semin. Has. Penelit. Inform. Dan Komput. SPINTER Inst. Teknol. Dan Bisnis STIKOM Bali*, pp. 818–823, Jun. 2024.

[17] C. Desai, “Comparative Analysis of Optimizers in Deep Neural Networks,” Oct. 2020.

[18] E. Okewu, P. Adewole, and O. Sennaike, “Experimental Comparison of Stochastic Optimizers in Deep Learning,” in *Computational Science and Its Applications – ICCSA 2019*, S. Misra, O. Gervasi, B. Murgante, E. Stankova, V. Korkhov, C. Torre, A. M. A. C. Rocha, D. Taniar, B. O. Apduhan, and E. Tarantino, Eds., Cham: Springer International Publishing, 2019, pp. 704–715. doi: 10.1007/978-3-030-24308-1_55.

[19] D. Irfan, T. S. Gunawan, and W. Wanayumini, “COMPARISON OF SGD, RMSProp, AND ADAM OPTIMATION IN ANIMAL CLASSIFICATION USING CNNs,” *Proceeding Int. Conf. Inf. Sci. Technol. Innov. ICoSTEC*, vol. 2, no. 1, Art. no. 1, Feb. 2023, doi: 10.35842/icostec.v2i1.32.

[20] “Classification and detection of COVID-19 X-Ray images based on DenseNet and VGG16 feature fusion,” *Biomed. Signal Process. Control*, vol. 77, p. 103772, Aug. 2022, doi: 10.1016/j.bspc.2022.103772.

[21] L. A. Andika, H. Pratiwi, and S. S. Handajani, “KLASIFIKASI PENYAKIT PNEUMONIA MENGGUNAKAN METODE CONVOLUTIONAL NEURAL NETWORK DENGAN OPTIMASI ADAPTIVE MOMENTUM,” *Indones. J. Stat. Its Appl.*, vol. 3, no. 3, pp. 331–340, Oct. 2019, doi: 10.29244/ijsa.v3i3.560.

[22] H. Hendriyana and Y. H. Maulana, “Identification of Types of Wood using Convolutional Neural Network with Mobilenet Architecture,” *J. RESTI Rekayasa Sist. Dan Teknol. Inf.*, vol. 4, no. 1, Art. no. 1, Feb. 2020, doi: 10.29207/resti.v4i1.1445.

BIBLIOGRAPHY OF AUTHORS

Gheri Febri Ananda is a master's student at Universitas Gadjah Mada, specializing in computer vision, deep learning, and intelligent control systems. His research interests focus on the development and application of advanced deep learning techniques and intelligent systems for solving complex problems in various domains.

Risfendra earned his Doctor of Philosophy (Ph.D.) from Southern Taiwan University of Science and Technology in 2017. He is currently a lecturer at Universitas Negeri Padang, specializing in *Industrial Automation* and *Intelligent Control Systems*. Risfendra's research focuses on the development and application of intelligent control systems in various industrial settings.

Eko Wahyudi is a master's student at Universitas Gadjah Mada, with expertise in computer vision, deep learning, and digital image processing. He is committed to advancing his skills in Python, AI, IoT, and robotics, focusing his research on applying advanced deep learning techniques and image processing across various fields.