Eduvest – Journal of Universal Studies
Volume 4 Number 06, June, 2024
p- ISSN 2775-3735- e-ISSN 2775-3727

EVALUATION AND SELECTION OF OPTIMAL DEEP LEARNING ARCHITECTURE FOR PREDICTING THE ENDPOINT IN HIGH SHEAR WET GRANULATION FOR ANTACID TABLET PRODUCTION

Irvan Maulana[1], Arry Yanuar[1]*, Sutriyo[1], Alhadi Bustamam[2]
[1] Faculty of Pharmacy, Universitas Indonesia, Jawa Barat, Indonesia.
[2] Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Indonesia, Jawa Barat, Indonesia.
Email: arry.yanuar@ui.ac.id

ABSTRACT

Objective: The purpose of this research was to evaluate and select the best architecture among native convolutional neural network (CNN), MobileNetV2, ResNet50V2, and EfficientNetB0 for predicting the endpoint of the high shear wet granulation process, with accuracy as the main evaluation metric. Methods: The dataset was captured from an industrial camera using static image analysis and was manually labeled as "NOT READY" and "READY" according to the traditional endpoint method based on the mixer's ampere point in the granulator. The dataset contained a total of 180 images, which were split between training and validation sets. Native CNN and TensorFlow Keras application programming interface (API) were utilized with MobileNetV2, EfficientNetB0, and ResNet50V2 as base feature encoders. Hyperparameters, such as final Fully Connected (FC) layer width, dropout rate, and learning rate, were optimized for binary classification using Keras hyper tuning. Results: The best was the native CNN, it was also the fastest among the three other models, taking only 20-30 ms per step for inference during runtime, though it requires 9000 ms time for training, the longest time among the models. It achieved an accuracy of 98%, and a validation accuracy of 97%. Conclusion: The system was able to determine when a wet granulation process has reached its endpoint based on live images from a camera after being trained on previously labeled data. The native CNN was the best model, offering the fastest runtime performance and the highest accuracy.

KEYWORDS *Wet Granulation, Image Processing, Deep Learning, Image-based inspection, MobileNetV2, EfficientNetB0, ResNet50V2.*

How to cite:	Irvan Maulana et al (2024). Evaluation And Selection Of Optimal Deep Learning Architecture For Predicting The Endpoint In High Shear Wet Granulation For Antacid Tablet Production. *Journal Eduvest*. *4* (6): 5422-5435
E-ISSN:	2775-3727
Published by:	https://greenpublisher.id/

INTRODUCTION

In the pharmaceutical industry, optimizing and ensuring the reproducibility of critical and sensitive processes is a top priority to guarantee product quality, consistency, and efficiency. Among these processes, wet granulation is a crucial step in the production of many solid tablet products. Wet granulation is a critical and challenging step in the solid tablet manufacturing process, particularly difficult to reproduce and scale up from research and development environments (Singh, Shirazian, Ranade, Walker, & Kumar, 2022). Most granulation processes are controlled through a fixed process condition such as the mixer current or mixer torque and are ultimately inspected manually through visual assessment by the operator. Image Processing is one of twelve methods available for inline monitoring in wet granulation. It is the most direct technique to monitor granule growth during the granulation process, however, a significant challenge of this method is frequent fouling of image capture equipment during the granulation (B. Liu et al., 2021).

The advancement in computational capabilities of image processing with deep learning algorithm has opened new opportunities in pharmaceutical manufacturing process, especially for analyzing granule growth or granule size distribution during granulation or milling process (Lou, Lian, & Hageman, 2021; Madarász, Mészáros, Köte, Farkas, & Nagy, 2023; Millen, Kovačević, Djuriš, & Ibrić, 2020; Yu et al., 2015). Image processing can be a versatile tool in the manufacturing and quality control methodologies in pharmaceutical industry, with various use cases, including controlling granule growth during granulation process (Farkas, Madarász, Nagy, Antal, & Kállai-Szabó, 2021; Mäki-Lohiluoma et al., 2021). The main challenge of applying deep learning-based image processing in granulation process is computational speed, as image capture and processing to predict endpoints must occur in real-time. Therefore, it is crucial to choose the optimal deep learning architecture for real-time performance. Native Convolutional Neural Network (CNN) architecture is widely used with good result for the abstract images (Amalia, Bustamam, & Sarwinda, 2021; Z. Liu et al., 2021; Sudarsono, Bustamam, & Tampubolon, 2020; Zhao et al., 2019). Three well-known models available in TensorFlow Keras Application Programming Interface (API) can handle resource-constrained environment with good performance: MobileNetV2 (Sandler, Howard, Zhu, Zhmoginov, & Chen, 2018), ResNet50V2 (Bustamam et al., 2021; He, Zhang, Ren, & Sun, 2016a; Sarwinda, Paradisa, Bustamam, & Anggia, 2021; Triyadi, Bustamam, & Anki, 2022), and EfficientNetB0 (Razi, Bustamam, & Latifah, 2023; Tan & Le, 2019) .

MobileNetV2 known for its efficiency, is particularly suitable for resource-constrained environments. ResNet50V2 with its deeper architecture, is renowned for its ability to handle complex hierarchical features. EfficientNetB0 leveraging compound scaling to achieve superior performance by balancing model depth, width, and resolution.

The purpose of this research was to evaluate and select the best architecture among native CNN, MobileNetV2, ResNet50V2, and EfficientNetB0 for predicting the endpoint of the high shear wet granulation process, with accuracy as the primary evaluation metric. To address the challenge of data acquisition related to fouling during image capturing in the granulation process, an in-process sampling method will be used to capture images and investigate the evolution of granule formation (Mahdi, Hassanpour, & Muller, 2018). This approach aims to prevent bias in quality of image processed with the deep learning architecture and ensure the accuracy of the predictions.

RESEARCH METHOD

Materials and Tools

This research utilized commercially available antacid tablet and the raw materials for these tablets which included Magnesium Hydroxide, Magnesium Aluminum Hydroxycarbonate, Activated Dimethylpolysiloxane, and several excipients such as binder, glidant, and fillers, comprising the mixture composition. All material used in this research were generously provided by Kalbe Farma, Indonesia.

The machine and tools used in this study included a Diosna P60 granulator and a 38-megapixel Relief M-12 camera for image capture. The processing computer was equipped with an Intel Core i7 13th Gen, 16 GB RAM and an NVIDIA GeForce RTX 4060 8GB graphics card.

Data Collection:

The granulation process was performed using a Diosna P60 granulation machine, and samples were taken every 1 minute for a total runtime 11 minutes. Each sample was then imaged using a camera with a resolution of 720 x 540 pixels with 72 dot per inch. As the samples were taken, the ampere values were also recorded. Then the images then labeled as “NOT READY” or “READY” based on the recorded ampere value.

Fig. 1. Sample collection and image capturing mechanism.

Model Architecture Selection:

The model first resized the images to 224 x 224 and rescaled the values to be between 1 to 0. For the base image features encoders, pretrained models were used. Before encoding, the images passed through a data augmentation layer (only during training) and a rescaling layer, depending on the encoder model. After encoding, the images were flattened and then passed through a fully connected layer, the width of which was optimized using hyperparameter tuning. Finally, the images passed through a final single-width fully connected layer. Dropout was used in some of the layers, with the dropout rate also being optimized through hyperparameter tuning.

Transfer Learning:

Transfer learning flow was implemented for MobileNetV2, ResNet50V2, and EfficientNetB0, and used the standard setup for transfer learning, only using the frozen convolution layers from the pre-trained models and discarding the last fully connected layer and adding native fully connected layers to train.

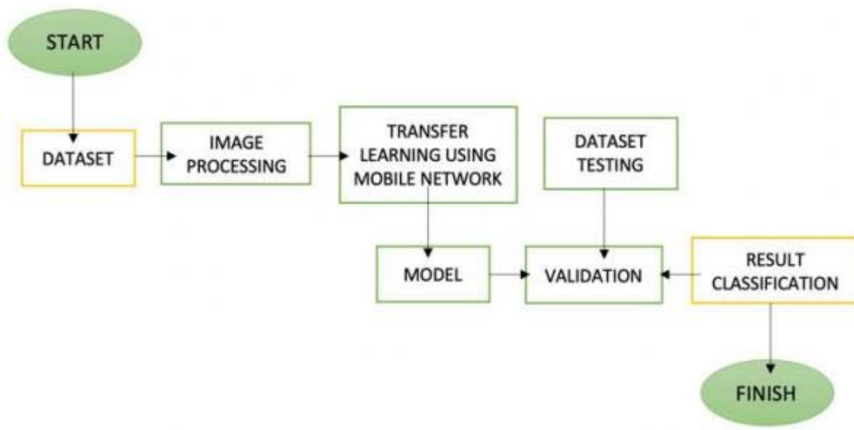

Fig. 2 Transfer learning flow

Model Architectures:

This research used native CNN, MobileNetV2, ResNet50V2, and EfficientNetB0 with architecture explain below:

MobileNetV2

MobileNetV2 is a convolutional neural network (CNN) architecture designed for mobile and edge devices. It introduces inverted residuals with linear bottlenecks and employs depthwise separable convolutions to reduce computational complexity. The network's basic building block is the inverted bottleneck block, composed of a 1x1 convolution, a depthwise separable convolution, and another 1x1 convolution (Sandler et al., 2018). Skip connections facilitate information flow, and global average pooling replaces fully connected layers at the end. The architecture allows for customization with width and resolution multipliers, providing flexibility in balancing model size and computational cost. Overall, MobileNetV2 is tailored for efficient on-device inference, making it suitable for resource-constrained environments.

Table 1 The MobileNetV2 architecture

Input	Operator	*t*	*c*	*n*	*s*
224^2 x 3	Conv2d	-	32	1	2
112^2 x 32	Bottleneck	1	16	1	1
112^2 x 16	Bottleneck	6	24	2	2
56^2 x 24	Bottleneck	6	32	3	2
28^2 x 32	Bottleneck	6	64	4	2
14^2 x 64	Bottleneck	6	96	3	1
14^2 x 96	Bottleneck	6	160	3	2
7^2 x 160	Bottleneck	6	320	1	1
7^2 x 320	Conv2d 1x1	-	1280	1	1
7^2 x 1280	Avgpool 7 x 7	-	-	1	-
1 x 1 x 1280	Conv2d 1x1	-	k	-	

ResNet50V2

ResNet50V2 is a 50-layer variant of the Residual Network architecture (He, Zhang, Ren, & Sun, 2016b). It utilizes residual blocks with three convolutional layers, including 1x1 projection shortcuts and 3x3 bottleneck convolutions. Residual skip connections are present between convolutional layers to facilitate an alternative path for data to flow. Batch normalization is applied after each convolutional layer to aid training convergence. The architecture ends with global average pooling and a dense layer for final predictions. ResNet50V2's design allows for the training of deep networks, addressing issues like vanishing gradients and promoting efficient information flow during both training and inference.

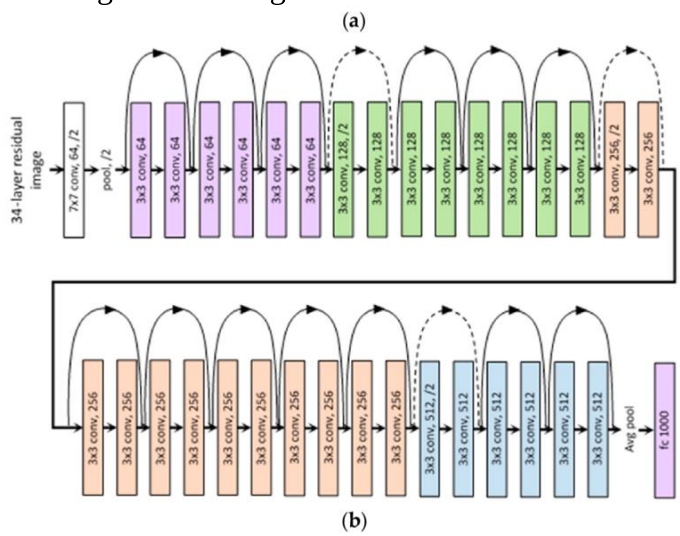

Fig. 3 The ResNet50V2 architecture

EfficientNetB0

EfficientNet optimizes CNN efficiency through compound scaling, adjusting depth, width, and resolution. It utilizes mobile inverted bottleneck convolutions (MBConvs) with depth wise separable convolutions and linear bottlenecks as building blocks. A unique compound scaling coefficient balances model size and computational cost, resulting in superior performance across resource constraints. The architecture, characterized by multiple blocks with different scaling factors (Tan & Le, 2019), has gained prominence for its efficiency and effectiveness in various computer vision tasks.

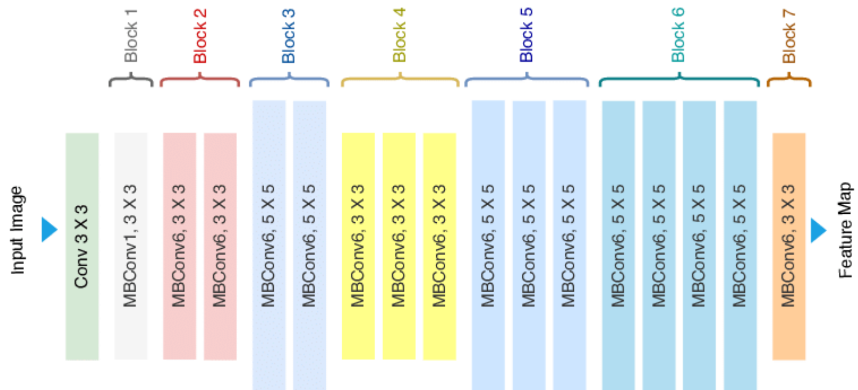

Fig. 4 The EfficientNetB0 architecture

Native CNN

A small standard architecture network was used for the native CNN, featuring only 2 convolution layers, each followed by max pooling. The output was then flattened and passed through 2 fully connected layers before reaching the final binary output layer. The filter sizes and counts were mostly chosen arbitrarily, while adhering to known rules for CNNs, which favor depth over width (Li, 2024).

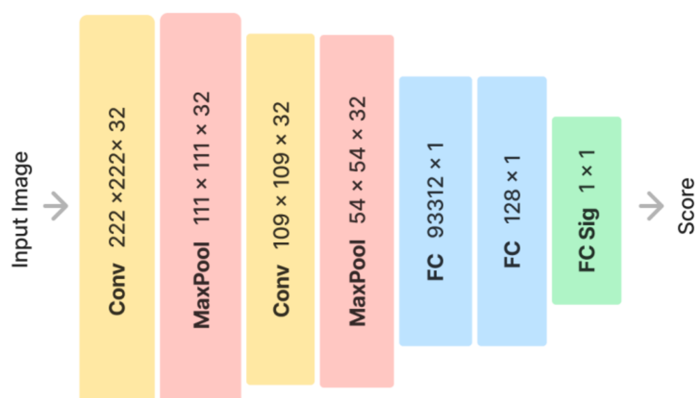

Figure 5: The native CNN architecture used in this research for binary classification of granulation process images.

Model Training:

Before the models underwent transfer learning, optimal hyperparameters were first searched using the Keras Tuner. All the pretrained models were frozen to act only as encoders and were not optimized; only the final fully connected layers were optimized using the Adam algorithm. The training was conducted over 100 epochs.

To improve model generalization, data augmentation techniques such as rotation, flipping, and zooming were applied during the training phase. This helped the models learn diverse features and perform better on new data. Hyperparameter tuning was applied to the transfer learning models (except the native CNN) using the Keras Tuner library, which includes multiple tools for hyperparameter tuning, such as the Hyperband algorithm and more. Three hyperparameters of the last two dense layers of the transfer learning models were tuned: dropout rate, dense layer units, and optimizer learning rate.

Model Evaluation:

To ensure the robustness of the models, a cross-validation approach was considered, splitting the dataset into two sets for training and validation. This process helped validate the consistency of model performance across different subsets of the dataset. The methodology described above formed the foundation for employing CNNs in determining the endpoint of a wet granulation process using machine vision. The utilization of transfer learning and hyperparameter optimization contributed to the development of a reliable and accurate model for granulation endpoint prediction.

After training, the images in the test set were used to evaluate the model's performance, with accuracy being the primary metric to compare the models and decide on the best one. The model's effectiveness in distinguishing between "NOT READY" and "READY" granulation states was assessed. Model predictions were categorized into true positive (TP), true negative (TN), false positive (FP), and false negative (FN) outcomes. Accuracy was calculated using the following formula:

$$\text{Accuracy} = (TP + TN) / (TP + FP + TN + FN) \quad (1)$$

RESULT AND DISCUSSION

Data Collection

A total of 180 clean images were collected, with 15 images sampled per minute over an 11-minute granulation process. The images had a resolution of 7120 x 5144 pixels. A sample image captured during the process can be seen in Figure 6 below.

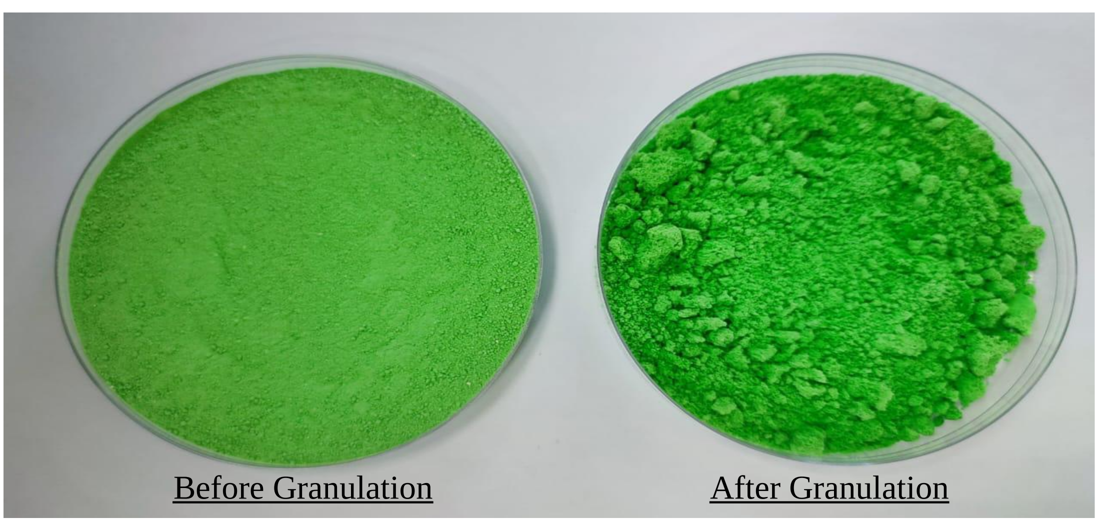

Fig. 6: Sample of images captured before and after the granulation process.

As the images were evaluated, the formation of granules during the granulation process could be observed using the in-process sampling method in the high shear granulator [20]. This method successfully addressed the data acquisition challenges inherent in the image processing methodology, allowing for a more thorough evaluation of the granulation process and the models' performance.

During the sampling, ampere values were recorded every minute. The recorded ampere data is presented in Figure 7 below.

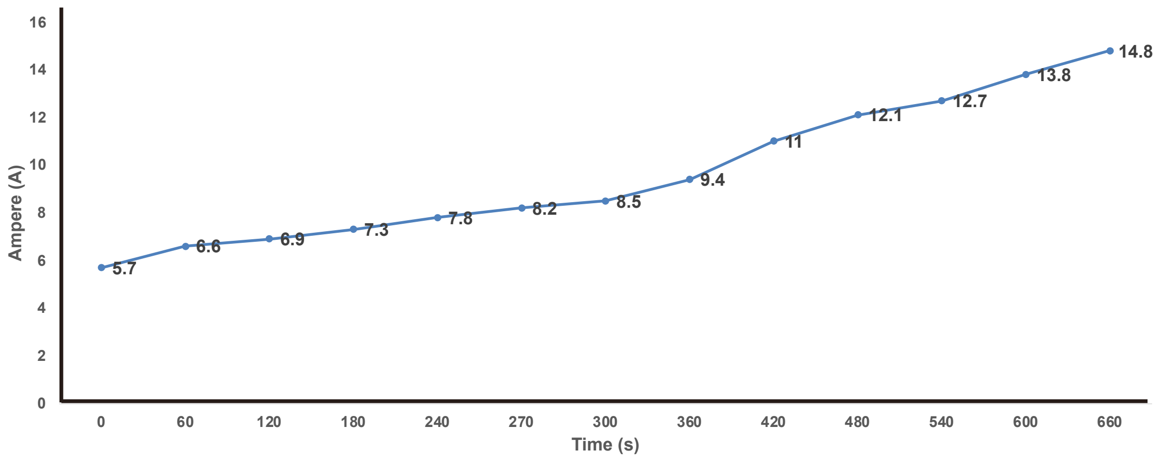

Fig. 7 : Ampere recorded during the granulation process.

Since the traditional endpoint of the granulation process is 8.5 amperes, all images captured before this value is reached are labeled as "NOT READY," indicating that the granulation process has not yet ended. Conversely, all images captured after this value is reached are labeled as "READY," indicating that the granulation process has reached its endpoint.

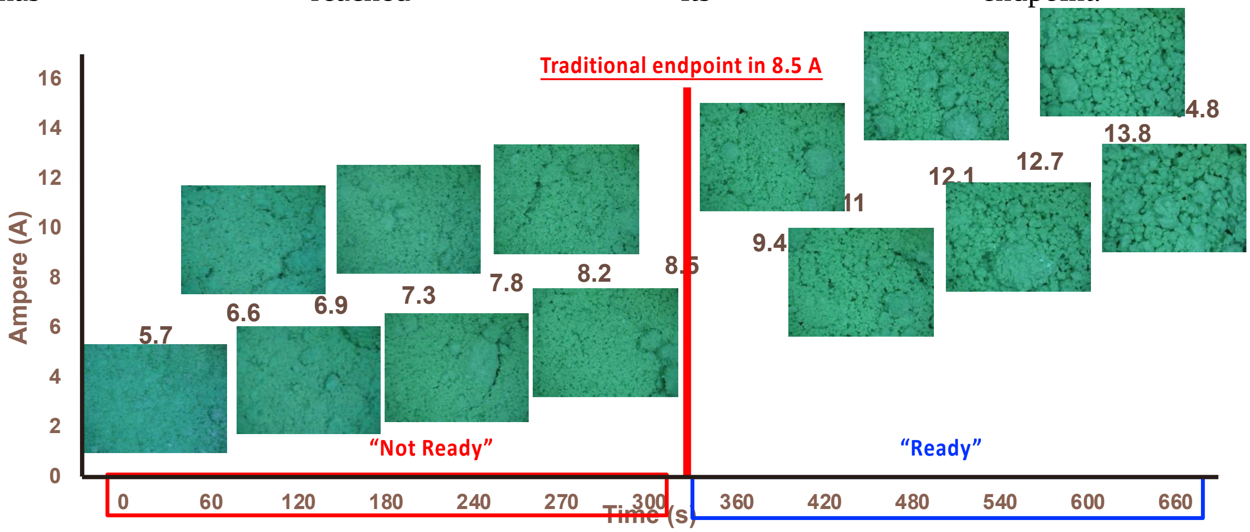

Fig. 8: Image captured and labelled based on ampere value on traditional endpoint

As seen in Figure 8, the final cleaned dataset contains a total of 180 images, split between training (139 images) and validation (41 images) sets. The training set consists of 69 images labeled as "NOT READY" and 70 images labeled as "READY," while the validation set consists of 20 images labeled as "NOT READY" and 21 images labeled as "READY."

Model Training

During the model training and hyperparameter tuning process, several restarts were necessary due to issues with plateauing accuracy values and errors. By the end of the hyperparameter tuning, the following optimal hyperparameters for each of the transfer learning models were identified.

Table 2: Optimal Hyperparameter

Model Name	Dropout rate	Layer unit count	Learning rate
MobileNetV2 224x224	0.21%	18	0.001
ResNet50V2 224x224	0.34%	32	0.001
EfficientNet 224x224	0.35%	40	0.001

After hyperparameter tuning was complete, each of the transfer learning models was fully trained with 100 epochs. Since the native CNN model did not undergo hyperparameter tuning, its hyperparameters remained fixed.

Model Evaluation

As seen in Table 3, the native CNN showed excellent performance with a final accuracy of 98% and a final validation accuracy of 97%. It had a very fast processing speed of only 20–30 ms per inference step but required 9000 ms per training step (this may include the validation step, so the actual training step speed may be faster). It was followed by MobileNetV2, which achieved a final accuracy of 91% and a final validation accuracy of 90%, with slightly slower processing speeds of 34–40 ms per inference step and only 650–750 ms per training step.

ResNet50V2 ranked third, showing a decent validation accuracy of 80% but was significantly slower, with 70–80 ms per inference step and 900–1000 ms per training step. It experienced some difficulty during the hyperparameter search. EfficientNet was particularly challenging during training and hyperparameter tuning, taking significantly more time to converge on optimal hyperparameters and parameters while also having a speed similar to that of ResNet50V2.

Table 3: Metrics result of models.

Model Name	Final Accuracy	Final Val Accuracy	Final Sensitivity	Final Val Sensitivity	Final F1 Score	Final Val F1 Score	Runtimes live (ms)	Runtime training (ms)
MobileNetV2 224x224	91%	90%	0.75	0.99	0.85	0.95	34-40	650-750
ResNet50V2 224x224	84%	80%	0.87	0.99	0.90	0.91	70-80	900-1000
EfficientNet 224x224	73%	80%	0.66	0.95	0.79	0.95	70-80	800-1000
CNN Native 224x224	98%	97%	0.76	0.95	0.84	0.98	20-30	9000

From the data in Table 2, a hypothesis can be drawn as to why the native CNN outperformed any transfer learning model. The transfer learning method only trains the final fully connected layers of the model while using pre-trained weights for the bulk of the feature extraction in the prior CNN layers. These layers are trained on the ImageNet dataset, which mostly contains natural images and labels significantly different from our dataset of granule images. This difference may induce a bias in the model, causing it to have difficulty classifying the images accurately. In contrast, the native CNN learns features from scratch, leading to a better fit to our specific dataset.

CONCLUSION

The native CNN, EfficientNet, ResNet50V2, and MobileNetV2 all exhibited decent capabilities, with the native CNN standing out for its overall performance across multiple hyperparameter tuning trials and its processing speed. The choice of model may depend on the specific requirements of the target task, computational resources, and dataset characteristics. For this research, the native CNN was the optimal choice because resources for the model were relatively limited, and the model needed to have low latency to process the live feed from the camera and prevent late stopping of the granulation machine.

In summary, this research advanced endpoint determination in granulation processes and provided a practical foundation for integrating CNNs into real-time, dynamic image analysis systems in industrial settings. The promising results encourage continued efforts to enhance manufacturing processes through efficient and adaptive technologies.

REFERENCES

Amalia, R., Bustamam, A., & Sarwinda, D. (2021). Detection and description generation of diabetic retinopathy using convolutional neural network and long short-term memory. *Journal of Physics: Conference Series*, *1722*(1). IOP Publishing Ltd. https://doi.org/10.1088/1742-6596/1722/1/012010

Bustamam, A., Sarwinda, D., Paradisa, R. H., Victor, A. A., Yudantha, A. R., & Siswantining, T. (2021). Evaluation of convolutional neural network variants for diagnosis of diabetic retinopathy. *Commun. Math. Biol. Neurosci.*, *2021*(0), Article ID 42. https://doi.org/10.28919/CMBN/5660

Farkas, D., Madarász, L., Nagy, Z. K., Antal, I., & Kállai-Szabó, N. (2021, May 1). Image analysis: A versatile tool in the manufacturing and quality control of pharmaceutical dosage forms. *Pharmaceutics*, Vol. 13. MDPI AG. https://doi.org/10.3390/pharmaceutics13050685

He, K., Zhang, X., Ren, S., & Sun, J. (2016a). Deep residual learning for image recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2016-December*, 770–778. https://doi.org/10.1109/CVPR.2016.90

He, K., Zhang, X., Ren, S., & Sun, J. (2016b). Deep residual learning for image recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2016-December*, 770–778. IEEE Computer Society. https://doi.org/10.1109/CVPR.2016.90

Li, X. (2024). Exploring the effect of depth and width of CNN models on binary classification of dogs and cats. *Applied and Computational Engineering*, *47*(1), 147–158. https://doi.org/10.54254/2755-2721/47/20241274

Liu, B., Wang, J., Zeng, J., Zhao, L., Wang, Y., Feng, Y., & Du, R. (2021). A review of high shear wet granulation for better process understanding, control and product development. *Powder Technology*, *381*, 204–223. https://doi.org/10.1016/J.POWTEC.2020.11.051

Liu, Z., Agu, E., Pedersen, P., Lindsay, C., Tulu, B., & Strong, D. (2021). Comprehensive Assessment of Fine-Grained Wound Images Using a Patch-Based CNN with Context-Preserving Attention. *IEEE Open Journal of Engineering in Medicine and Biology*, *2*, 224–234. https://doi.org/10.1109/OJEMB.2021.3092207

Lou, H., Lian, B., & Hageman, M. J. (2021). Applications of Machine Learning in Solid Oral Dosage Form Development. *Journal of Pharmaceutical Sciences*, *110*(9), 3150–3165. https://doi.org/10.1016/j.xphs.2021.04.013

Madarász, L., Mészáros, L. A., Köte, Á., Farkas, A., & Nagy, Z. K. (2023). AI-based analysis of in-line process endoscope images for real-time particle size measurement in a continuous pharmaceutical milling process. *International Journal of Pharmaceutics, 641*. https://doi.org/10.1016/j.ijpharm.2023.123060

Mahdi, F., Hassanpour, A., & Muller, F. (2018). An investigation on the evolution of granule formation by in-process sampling of a high shear granulator. *Chemical Engineering Research and Design*, *129*, 403–411. https://doi.org/10.1016/j.cherd.2017.10.038

Mäki-Lohiluoma, E., Säkkinen, N., Palomäki, M., Winberg, O., Ta, H. X., Heikkinen, T., … Kauppinen, A. (2021). Use of machine learning in prediction of granule particle size distribution and tablet tensile strength in commercial pharmaceutical manufacturing. *International Journal of Pharmaceutics*, *609*, 121146. https://doi.org/10.1016/J.IJPHARM.2021.121146

Millen, N., Kovačević, A., Djuriš, J., & Ibrić, S. (2020). Machine Learning Modeling of Wet Granulation Scale-up Using Particle Size Distribution Characterization Parameters. *Journal of Pharmaceutical Innovation*, *15*(4), 535–546. https://doi.org/10.1007/s12247-019-09398-0

Razi, F. A., Bustamam, A., & Latifah, A. L. (2023). Development of Efficient Brain Tumor Classification on MRI Image Results Using EfficientNet. *2023 International Seminar on Intelligent Technology and Its Applications: Leveraging Intelligent Systems to Achieve Sustainable Development Goals, ISITIA 2023 - Proceeding*, 575–580. https://doi.org/10.1109/ISITIA59021.2023.10221186

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). MobileNetV2: Inverted Residuals and Linear Bottlenecks. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 4510–4520. IEEE Computer Society. https://doi.org/10.1109/CVPR.2018.00474

Sarwinda, D., Paradisa, R. H., Bustamam, A., & Anggia, P. (2021). Deep Learning in Image Classification using Residual Network (ResNet) Variants for Detection of Colorectal Cancer. *Procedia Computer Science*, *179*, 423–431. Elsevier B.V. https://doi.org/10.1016/j.procs.2021.01.025

Singh, M., Shirazian, S., Ranade, V., Walker, G. M., & Kumar, A. (2022). Challenges and opportunities in modelling wet granulation in pharmaceutical industry – A critical review. *Powder Technology*, *403*, 117380. https://doi.org/10.1016/J.POWTEC.2022.117380

Sudarsono, E., Bustamam, A., & Tampubolon, P. P. (2020). An optimized convolutional neural network using diffgrad for cataract image classification. *AIP Conference Proceedings*, *2296*. American Institute of Physics Inc. https://doi.org/10.1063/5.0030746

Tan, M., & Le, Q. V. (2019). *EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks*. Retrieved from http://arxiv.org/abs/1905.11946

Triyadi, A. B., Bustamam, A., & Anki, P. (2022). Deep Learning in Image Classification using VGG-19 and Residual Networks for Cataract Detection. *Proceedings - 2022 2nd International Conference on Information Technology and Education, ICIT and E 2022*, 293–297. https://doi.org/10.1109/ICITE54466.2022.9759886

Yu, H., Fu, J., Dang, L., Cheong, Y., Tan, H., & Wei, H. (2015). Prediction of the Particle Size Distribution Parameters in a High Shear Granulation Process Using a Key Parameter Definition Combined Artificial Neural Network Model. *Industrial and Engineering Chemistry Research*, *54*(43), 10825–10834. https://doi.org/10.1021/acs.iecr.5b02679

Zhao, X., Liu, Z., Agu, E., Wagh, A., Jain, S., Lindsay, C., … Kan, J. (2019). Fine-Grained Diabetic Wound Depth and Granulation Tissue Amount Assessment Using Bilinear Convolutional Neural Network. *IEEE Access*, *7*, 179151–179162. https://doi.org/10.1109/ACCESS.2019.2959027