Integration of Capsule Network with CNN for Plant Leaf Disease Classification

Aekkarat Suksukont[1], Ekachai Naowanich[2]
Department of Digital Media Technology, Faculty of Science and Technology, Rajamangala University of Technology Suvarnabhumi, Thailand[1]
Department of Digital Technology, Faculty of Science and Technology, Rajamangala University of Technology Suvarnabhumi, Thailand [2]
E-mail: 166490431006@rmutsb.ac.th [1]
E-mail: ekachai.n@rmutsb.ac.th[2]

Received: February 20, 2025; Revised: April 2, 2025; Accepted: April 11, 2025

ABSTRACT

Plant leaf disease classification is challenging due to the wide variation in disease symptoms and the diverse morphological characteristics of plant leaves. These variations complicate model training and hinder classification accuracy. This study proposed a hybrid deep learning (DL) model for leaf disease training and classification. The proposed model integrates Capsule Networks (CN) for spatial relationship retention, SE-Residual blocks improve feature extraction while minimizing information loss, and CN capture spatial relationships with reduced dependency on large datasets, and Long Short-Term Memory (LSTM) to enhance training efficiency. The proposed model was trained and evaluated using the Rice Leaf Disease Dataset (RLDD). Its performance was compared with existing state-of-the-art models. The experimental results showed that the proposed model achieved the highest training accuracy of 96.01%, classification results 75.67% for bacterial leaf blight, 80.43% for brown spot, 86.67% for healthy, 76.52% for leaf blight, 98.96% for leaf scald, and 93.18% for narrow brown spot. These results highlight the effectiveness of the proposed model in achieving high accuracy for plant leaf disease classification.

KEYWORDS: Plant leaf disease classification, Integration networks, Convolutional neural network, Deep learning.

Introduction

Plant diseases reduce the quality and quantity of crops in agriculture worldwide (Ristaino et al., 2021; Karthickmanoj et al., 2024; Gai et al., 2024). These diseases not only cause economic losses but also threaten global food security, making their identification and management essential to minimize their impact. Traditional methods for plant disease identification rely on visual inspection or diagnosis by experienced personnel, which is both time-consuming and dependent on specialized expertise. These methods often fail to provide the speed and accuracy required in modern agriculture. In resource-limited settings, the reliance on manual monitoring further complicates disease management, resulting in delays in diagnosis and preventive actions. These challenges highlight the need for new technologies and methods to identify plant diseases more quickly, accurately, and efficiently. DL has gained significant attention for its potential to address these limitations by reducing reliance on manual inspection and improving the accuracy of plant disease diagnosis (Heng et al., 2024; Sarkar et al., 2023).

In recent years, DL, especially CNN, has gained much attention for its ability to learn and extract spatial features, making it suitable for analyzing plant leaf diseases. (Singh et al., 2017) (Mahadevan et al., 2024). This has led to significant progress in developing techniques for leaf disease detection and classification, such as (Thaseentaj et al., 2023) proposed a customized deep CNN with a deeper structure and the ability to learn complex features from data to address issues affecting yield and quality in mango leaf disease detection and classification. (Paul et al., 2023) proposed a web and android application has been designed to assist farmers with real-time classification of tomato leaf diseases. The system integrates state-of-the-art VGG16 and VGG19 networks, trained using transfer learning, and emphasizes data augmentation to improve model accuracy. (Hessane et al., 2023) proposed image analysis combined with machine learning, focusing on feature extraction based on 80 gray level co-occurrence matrix and HSV features. It is tested with support vector machine, k-nearest neighbors, random forest, and light gradient boosting machine to detect and classify disease outbreak levels. (Ta ji et al., 2024) ensemble DL by combining CNN and Local Binary Pattern with binary dragonfly, ant colony, and moth flame optimization for plant leaf disease classification. (Muthusamy et al, 2024) ensemble learning with CNN by tuning parameters in dense layers and combining multiple networks using an averaging strategy. This approach aims to enhance classification efficiency through experimentation with three types of networks. These studies highlight the potential of CNN in plant disease analysis and classification. However, refining and optimizing their structure remains crucial for addressing challenges in agricultural applications.

Despite the advancements in CNN-based models, existing methods struggle with image variations, leading to misclassification. This study aims to address these challenges by integrating advanced DL components, including CN, SE-residual networks, and LSTM to improve the model's accuracy and learning capability. SE-residual address the efficiency degradation in deep networks by preserving critical information during learning, while CN capture complex structures and spatial relationships in data, reducing the dependency on large datasets and artificial data augmentation. By combining these techniques, the model can be adapted to more complex data, making it particularly effective for agricultural applications.

Objectives

1. To study DL networks, specifically CNN and Capsule Networks, for plant disease classification.

2. To integrate Capsule Network with CNN to improve the accuracy of plant leaf disease classification.

Hypothesis

1. Noisy leaf characteristics interfere with the network's ability to effectively learn and classify leaf disease characteristics.

2. The integration of Capsule Networks with CNN improves training efficiency and increases the accuracy of leaf disease classification compared to individual networks like CNN or other common models.

Expected Benefits

1. Enhance the efficiency of DL models for accurate plant disease classification.

2. Provide guidance for developing mobile applications for plant leaf disease classification systems.

Experimental Method

Dataset

This study utilized a disease dataset, with the rice leaf diseases dataset (RLDD) (Singh et al., 2020) was used for experimentation. It contains 2,628 rice leaf images categorized into six types of diseases: bacterial leaf blight (BLB), brown Spot (BS), leaf blast (LB), leaf scald (LS), narrow brown spot (NBS), and healthy (HE), with each category consisting of 428 images, are shown in figure 2. The images were organized and categorized for training and testing the DL. The dataset was divided into a training set comprising 80% of the data (60% for training and 20% for validation) and 20% for testing set.

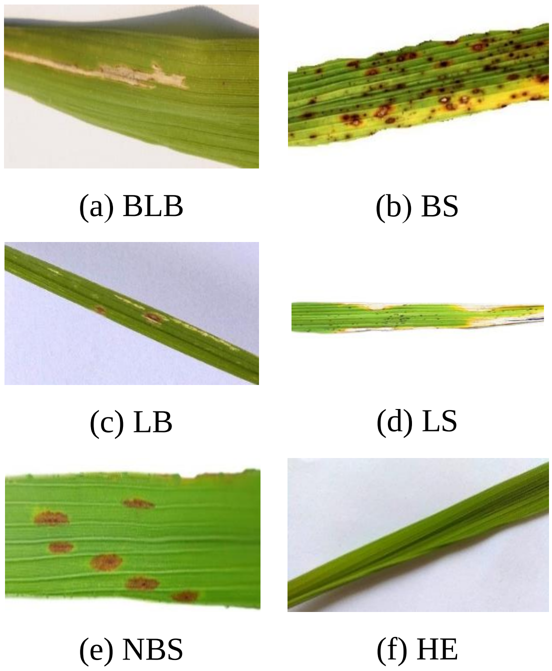


Figure 1: RLDD dataset.

Deep Learning Network Design

The experimental VGG19 model is modified and enhanced to improve its processing and learning efficiency, (referred to as Modified VGG19), in combination with the design approach presented (Zhang et al., 2024). SE-SK-CapResNet is proposed as an artificial neural network (ANN) that integrates residual blocks and CN to enhance data collection and processing capabilities. This architecture is designed to improve model performance across all dimensions.

Figure 2 illustrates the Modified VGG19 model implemented this approach by defining convolutional layers, as described in Equation 1, with filter sizes of 64, 128, and 256, increasing progressively in each block to capture data features, using 3×3 filters with SAME padding to preserve image dimensions during computation. For spatial dimensionality reduction, a 2×2 MaxPooling2D layer with a stride of 2 is used to highlight significant features.

The output is then passed through the SE-Residual block, which applies filters of sizes 64, 128, and 256 with a stride of 2 to enhance the efficiency of feature aggregation across multiple channels. The results from the SE-Residual block are forwarded to a CN layer, configured with 32 capsules and a capsule dimension of 8, utilizing 3×3 filters with a stride of 2 to capture detailed spatial relationships.

Subsequently, the data is passed into an LSTM layer with 128 units, incorporating a dropout rate of 0.2 to prevent overfitting and optimize the learning potential of the data sequence. Finally, the output is processed by a SoftMax layer to convert the results into probabilities for classifying the data into corresponding categories.

$$y_{i,j} = \sum_{m} \sum_{n} x_{(i+m),(j+n)} \cdot k_{m,n} + b \quad (1)$$

When $x_{(i+m),(j+n)}$ input at position *(i+m, j+n)*, $y_{i,j}$ is the output at pixel *(i, j)*, $k_{m,n}$ denotes the filter at position *(m, n)*, and *b* is the bias term.

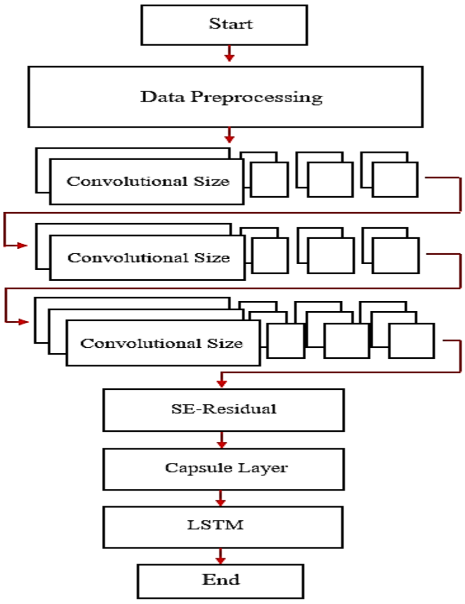


Figure 2: Proposed method.

In this experiment, the researchers performed data cleaning (DC) to change the characteristics of the experimental images by creating a function to detect excessive background color levels and calculating the ratio of pixels with intensity values lower than 30 to pixels in the image and comparing it to a threshold value set at 0.8 of the images, which sets the ratio exceeding the threshold value to zero. The image after DC shown in figure 3.

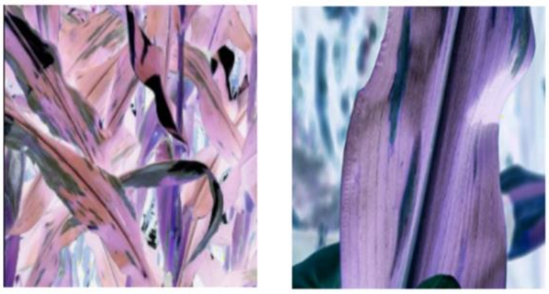


Figure 3: Image transformed using thresholding.

The proposed network was trained using the parameters summarized in Table 1.

Table 1: Parameter for training model.

Parameter	**Value**
Image size	224x224x3
Learning rate	10^{-3}
Epoch	50
Batch size	64
Loss function	Categorical crossentropy
Optimization	Adam

Evaluation

Accuracy is used to evaluate training efficiency by calculating the ratio of correctly predicted samples to the total number of samples as:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (2)$$

Where true positive (TP) refers to the number of correct samples accurately predicted by the model, true negative (TN) represents the number of incorrect samples correctly identified as negative, false positive (FP) indicates the number of incorrect samples mistakenly predicted as positive, and false negative (FN) denotes the number of correct samples that the model incorrectly predicts as negative.

The confusion matrix was utilized to analyze and present the classification results of the model, with the outcomes displayed in a tabular format for better interpretation (Krstinić et al., 2024).

Result

Training Performance

Figure 4 illustrates the comparative performance of models trained on the RLDD dataset. The Modified VGG19 model achieves the highest performance with an accuracy of 89.93%. In contrast, VGG19 achieves the next highest accuracy at 88.12%.

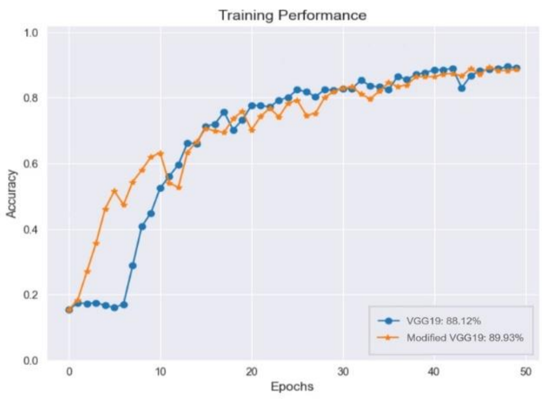


Figure 4: Training performance of RLDD.

Figure 5 illustrates the comparative performance of models trained with DC on the RLDD dataset. The Modified VGG19 model achieves the highest performance with an accuracy of 96.01%, demonstrating consistent training progress from the initial epochs (0–10) and maintaining stability throughout the training process. In contrast, VGG19 achieves the next highest accuracy at 93.98%.

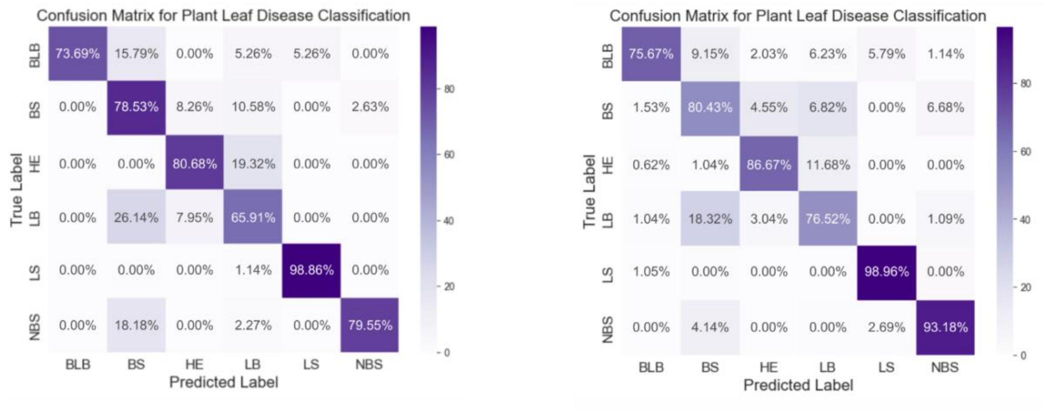
Figure 5: Training performance with DC of RLDD.

Classification Performance

After training, the network was tested for its ability to classify plant leaf diseases, and the results are presented in the confusion matrices shown in figure 4.

Figure 6 compares the performance of VGG19, and Modified VGG19 models in classifying leaf diseases using the RLDD dataset, as represented by their respective confusion matrices. The Modified VGG19, while reducing some errors, still exhibited significant challenges in classification. It performed relatively well in certain groups, such as LS at 98.96%, and BLB at 93.18%, but showed considerable confusion in the HE, BS, LB, and BLB groups, limiting its overall effectiveness. VGG19 achieved the second-highest accuracy, performing well in the LS groups with a TP rate of 98.86%.

However, it struggled in classifying similar data, as reflected by lower TP rates in the HE at 80.68%, NBS at 79.55%, BS at 78.53, and BLB at 73.69% groups, but showed considerable confusion in the HE, NBS, BS, and BLB groups, limiting its overall effectiveness highlighting its limitations in distinguishing overlapping features. These results demonstrate the outstanding performance of the Modified VGG19 model in handling complex data within high-performance training sets; however, improvements are still needed in the classification component of the dataset to achieve even higher accuracy.



(a) CNN (VGG19)

(b) Modified VGG19

Figure 6: Classification results for the RLDD.

Comparison Performance

In this experiment, state-of-the-art techniques were evaluated for comparison with the proposed method. These included ResNet50 (Adnan et al., 2023), ResNet101 (Sethy et al., 2024), DenseNet121 (Huang et al., 2017), and InceptionV3 (Szegedy et al., 2016), all trained using standardized parameters outlined in table 1. The training performance is illustrated in figure 5.

Figure 7 shows the training results on the RLDD dataset. InceptionV3 achieved the highest efficiency at 80.87%, with rapid initial learning and stable performance. DenseNet121 followed at 73.76%, showing steady improvement. In contrast, ResNet101 and ResNet50 had the lowest efficiencies, at 25.08% and 24.05%.

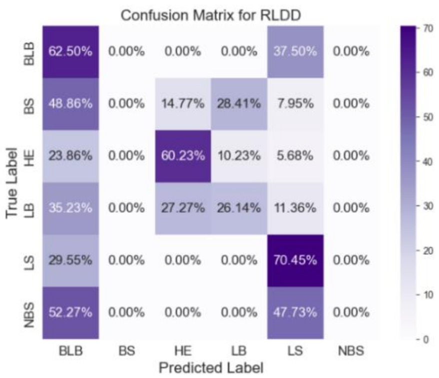
Figure 7: Training performance of DL with the RLDD.

The comparison performance training results shown in table 2.

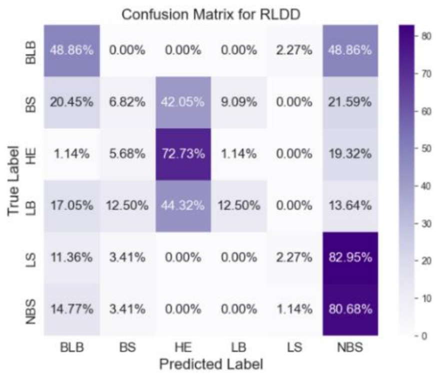
Table 2: Comparison performance.

Model	RLDD
VGG19	93.98%
Modified VGG19	96.01%
Resnet50	24.05%
Resnet101	25.08%
DenseNet121	73.76%
InceptionV3	80.87%

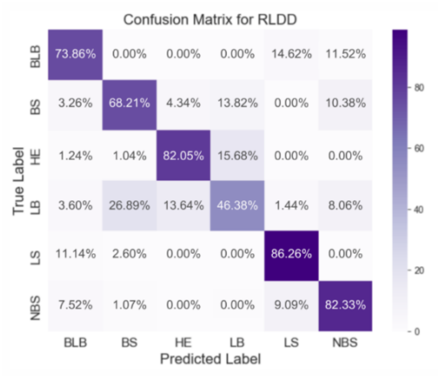
When the training model were tested for classification performance, shown in figure 8.



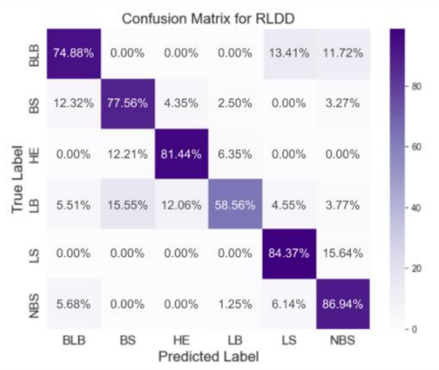
(a) ResNet50



(b) ResNet101



(c) DenseNet121



(d) InceptionV3

Figure 8: Classification results of DL with the RLDD.

Figure 8, the confusion matrix for the RLDD, showing that DenseNet121 and InceptionV3 outperform ResNet50 and ResNet101. DenseNet121 achieves the highest TP rates across groups, including LS at 86.26%, NBS at 82.33%, and HE at

82.05%, while InceptionV3 performs well in NBS at 86.94%, LS at 84.37% and HE at 81.44%. In contrast, ResNet50 and ResNet101 perform poorly, with ResNet50 achieving low TP rates in LS at 70.45% and BLB at 62.50%, and ResNet101 struggling in LS at 2.27%, BS at 6.82% and LB at 12.50%.

Discussion

Plant leaf diseases classification is challenging due to the variations in disease symptoms, which often lead to misclassification. This study presents a hybrid DL model for leaf disease training and classification. The proposed model integrates CN, SE-Residual Blocks, and LSTM to enhance training efficiency. The design employs an improved convolutional operator to efficiently extract features, followed by SE-Residual blocks to emphasize critical features and address the issue of information loss. CN are utilized to capture complex structures and spatial relationships. The researchers performed DC to change the characteristics of the experimental images by creating a function to detect excessive background color levels and calculating the ratio of pixels with intensity values lower than 30 to pixels in the image and comparing it to a threshold value set at 0.8 of the images, which sets the ratio exceeding the threshold value to zero. The experimental results show that this approach outperforms existing models, improving classification accuracy and overall efficiency. This design expands upon the study in (Zhang et al., 2024) to further enhance its capabilities.

Conclusion

This experiment focuses on advancing hybrid DL model for leaf disease training and classification. The proposed model integrates CN, SE-Residual Blocks, and LSTM to enhance training efficiency. The design employs an improved convolutional operator to efficiently extract features, followed by SE-Residual blocks to emphasize critical features and address the issue of information loss.

The proposed model was trained on the RLDD datasets, achieving maximum training accuracy of 96.01%. During testing, the model demonstrated outstanding classification accuracies of 75.67%, 80.43%, 86.67%, 76.52%, 98.96%, and 93.18% for BLB, BS, HE, LB, LS, and NBS. These results highlight the proposed method's superior accuracy and efficiency compared to previous studies, establishing it as a reliable solution for plant leaf disease classification across diverse environmental conditions.

In the future, research aim to development networks capable of classification datasets with higher accuracy, ultimately providing an efficient and accessible tool for plant leaf disease classification in agriculture.

References

Ristaino, J. B., Anderson, P. K., Bebber, D. P., Wei, Q., et al. (2021). The persistent threat of emerging plant disease pandemics to global food security. *Proceedings of the National Academy of Sciences of the United States of America, 118*(23), Art. No. e2022239118. https://doi.org/10.1073/pnas.2022239118.

Karthickmanoj, R., & Sasilatha, T. (2024). Development of plant disease detection for smart agriculture. *Multimedia Tools and Applications, 83,* 54391–54410. https://doi.org/10.1007/s11042-023-17687-7.

Gai, Y., & Wang, H. (2024). Plant disease: A growing threat to global food security. *Agronomy, 14*(8), Art. No. 1615. https://doi.org/10.3390/agronomy14081615.

Heng, Q., Yu, S., & Zhang, Y. (2024). A new AI-based approach for automatic identification of tea leaf disease using deep neural network based on hybrid pooling. *Heliyon*. No. e26465. https://doi.org/10.1016/j.heliyon.2024.e26465.

Sarkar, C., Gupta, D., Gupta, U., & Hazarika, B. B. (2023). Leaf disease detection using machine learning and deep learning: Review and challenges. *Applied Soft Computing, 145,* Art. No. 110534. https://doi.org/10.1016/j.asoc.2023.110534.

Singh, V., & Misra, A. K. (2017). Detection of plant leaf diseases using image segmentation and soft computing techniques. *Information Processing in Agriculture, 4*(1), 41–49. https://doi.org/10.1016/j.inpa.2016.10.005.

Mahadevan, K., Punitha, A., & Suresh, J. (2024). Automatic recognition of rice plant leaf diseases detection using deep neural network with improved threshold neural network. E-Prime - Advances in Electrical Engineering, *Electronics and Energy*, 8, Art. No. 100534. https://doi.org/10.1016/j.prime.2024.100534.

Thaseentaj, S., & Ilango, S. S. (2023). Deep convolutional neural networks for South Indian mango leaf disease detection and classification. *Computers, Materials and Continua, 77*(3), 3593–3618. https://doi.org/10.32604/cmc.2023.042496.

Paul, S. G., Biswas, A. A., Saha, A., Zulfiker, M. S., Ritu, N. A., Zahan, I., Rahman, M., & Islam, M. A. (2023). A real-time application-based convolutional neural network approach for tomato leaf disease classification. *Array,* 19, Art. No. 100313. https://doi.org/10.1016/j.array.2023.100313.

Hessane, A., El Youssefi, A., Farhaoui, Y., Aghoutane, B., & Amounas, F. (2023). A machine learning-based framework for a stage-wise classification of date palm white scale disease. *Big Data Mining and Analytics, 6*(3), 263–272. https://doi.org/10.26599/BDMA.2022.9020022.

Taji, K., Sohail, A., Shahzad, T., Khan, B. S., Khan, M. A., & Ouahada, K. (2024). An ensemble hybrid framework: A comparative analysis of metaheuristic algorithms for ensemble hybrid CNN features for plants disease classification. *IEEE Access, 12*, 61886–61906. https://doi.org/10.1109/ACCESS.2024.3389648.

Muthusamy, S., & Ramu, S. P. (2024). IncepV3Dense: Deep ensemble-based average learning strategy for identification of micro-nutrient deficiency in banana crop. *IEEE Access*, 1273779–73792. https://doi.org/10.1109/ACCESS.2024.3405027.

Zhang, X., Mao, Y., Yang, Q., & Zhang, X. (2024). A plant leaf disease image classification method integrating capsule network and residual network. *IEEE Access, 12*, 44573–44585. https://doi.org/10.1109/ACCESS.2024.3377230.

Singh, D., Jain, N., Jain, P., Kayal, P., Kumawat, S., & Batra, N. (2020). PlantDoc: A dataset for visual plant disease detection. *In Proceedings of the 7th ACM IKDD CoDS and 25th COMAD*, 249–253. https://doi.org/10.1145/3371158.337119.

Krstinić, D., Kuzmanić Skelin, A., Slapničar, I., & Braović, M. (2024). Multi-label confusion tensor. *IEEE Access, 12*, 9860–9870. https://doi.org/10.1109/ACCESS.2024.3433495.

Adnan, M. M., Rahim, M. S. M., Khan, A. R., Alkhayyat, A., Alamri, F. S., & Saba, T. (2023). Automated image annotation with novel features based on deep ResNet50-SLT. *IEEE Access*, 11, 40258–40277. https://doi.org/10.1109/ACCESS.2023.3266296.

Sethy, P. K., Korada, L., Behera, S. K., Shirole, A., Amat, R., & Nanthaamornphong, A. (2024). Maximizing steel slice defect detection: Integrating ResNet101 deep features with SVM via Bayesian optimization. *Sensors and Actuators: A. Systems*, 6, Art. No. 200170. https://doi.org/10.1016/j.sasc.2024.200170.

Huang, G., Liu, Z., van der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (CVPR), 2261–2269. https://doi.org/10.1109/CVPR.2017.243

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (CVPR), 2818–2826. https://doi.org/10.1109/CVPR.2016.308.