



Deep Learning-Based Computer Vision Framework for Early Detection of Cocoa Plant Diseases in Precision Agriculture

I Putu Oka Wisnawa¹, Ni Nyoman Harini Puspita², I Putu Bagus Arya Pradnyana³,
I Komang Wiratama⁴, and I Made Dwi Jendra Sulastra⁵

^{1,2,3,4} Information Technology Department, Politeknik Negeri Bali, Bali, Indonesia ⁵Accounting Department, Politeknik Negeri Bali, Bali, Indonesia okawisnawa@pnb.ac.id

Abstract. Early detection of plant diseases is crucial for minimizing crop losses and ensuring sustainable agriculture, particularly in tropical regions. Cocoa (*Theobroma cacao*), a major cash crop, is highly susceptible to diseases such as Black Pod Rot and Monilia Pod Rot, which can severely impact yield. Conventional manual inspection methods are time-consuming, prone to error, and often implemented too late to prevent spread. This study proposes a real-time object detection framework using YOLOv8, a modern deep learning architecture, to identify cocoa pod diseases under natural field conditions. A custom dataset comprising 474 annotated images was collected from Jembrana, Bali, and enhanced using Roboflow-based augmentation. The YOLOv8n model was trained over 30 epochs and evaluated using standard object detection metrics. The model achieved a mAP@0.5 of 87.4%, mAP@0.5:0.95 of 74.1%, and an average inference time of 0.61 seconds per image, demonstrating high accuracy and suitability for edge deployment. The results validate the feasibility of using YOLOv8 for early-stage disease detection in cocoa, particularly in resource-limited environments. This work contributes to smart farming innovation by offering a lightweight, scalable solution for real-time agricultural diagnostics.

Keywords: Cocoa Plant Disease, Computer Vision, Deep Learning, Precision Agriculture, YOLO

1 Introduction

Precision agriculture has emerged as a transformative paradigm in modern farming, aiming to enhance productivity, optimize input utilization, and support environmental sustainability. However, in tropical agricultural systems where environmental variability, high humidity, and limited access to technology are common, crop production remains highly vulnerable to plant diseases (Saraan et al., 2023; Monteiro

et al., 2021; Saranya et al., 2023; Sharma et al., 2021; Syarovy et al., 2023). Cocoa (*Theobroma cacao*), a key cash crop cultivated across tropical regions, is particularly susceptible to fungal pathogens such as Black Pod Rot (*Phytophthora palmivora*) and Monilia Pod Rot (*Moniliophthora roreri*), which can cause severe yield losses if not addressed promptly (Arakeri et al., 2024; Gómez-de la Cruz et al., 2023; Diego et al., 2025).

In cocoa-producing regions such as Jembrana, Bali (Indonesia), disease detection is still largely based on manual observation by farmers. This process is time-consuming, imprecise, and heavily dependent on individual experience. These traditional methods are often reactive rather than preventive, allowing diseases to progress unnoticed until significant crop damage has occurred. Consequently, there is a critical need for automated, real-time, and accurate detection systems that can operate effectively in field conditions with limited infrastructure.

Recent advancements in deep learning and computer vision have enabled new possibilities for plant disease identification through image-based analysis. However, most existing research in this domain has focused on leaf-level classification using convolutional neural networks (CNNs) trained on benchmark datasets such as PlantVillage. These models typically rely on curated, cropped, and high-contrast images captured under controlled environments, which limits their applicability in field settings characterized by multiple plant parts, complex backgrounds, and variable lighting (Falaschetti et al., 2022; Hassan et al., 2021; Lu et al., 2021; Sun et al., 2022).

To address these challenges, object detection algorithms offer a more practical solution by enabling simultaneous localization and classification of diseased regions in unstructured visual scenes. Among these, the You Only Look Once (YOLO) family of models has gained prominence due to its speed and accuracy. The most recent version, YOLOv8, introduces anchor-free detection mechanisms, improved backbone architecture, and compatibility with edge computing devices, making it ideal for use in low-resource agricultural environments (Wang & Liu, 2024; Wang et al., 2024; Zheng et al., 2024).

This study proposes a YOLOv8-based computer vision framework for the early detection of cocoa pod diseases under real-world field conditions. A custom annotated image dataset was developed using field-acquired photographs of symptomatic cocoa pods in Bali. The dataset was further enhanced through automated augmentation and preprocessing pipelines using Roboflow. The YOLOv8n variant was selected to strike a balance between detection accuracy and computational efficiency, and the model was trained and evaluated using standard object detection metrics. By leveraging a lightweight and real-time detection architecture, this framework offers a scalable and practical solution for integrating artificial intelligence into tropical cocoa farming practices.

This research contributes to the ongoing digital transformation of agriculture by delivering a model that supports early intervention, reduces reliance on chemical treatments, and improves disease management strategies for smallholder farmers. The proposed system also aligns with broader goals in sustainable agriculture, food security, and climate resilience in tropical regions.

2 Methodology

This study was conducted in the cocoa plantations of *Kelompok Tani* Merta Abadi, located in Jembrana Regency, Bali, Indonesia. The area was selected due to its significance as a center of smallholder cocoa production and the prevalence of pod diseases that pose a serious threat to yield and quality. Data collection was performed in situ during the cropping season, allowing for the capture of real-world variations in lighting, background, and disease manifestation. Images were taken using commercially available smartphones with a minimum resolution of 12 megapixels, ensuring sufficient detail to capture visual symptoms on cocoa pods. The dataset comprised 198 original images, consisting of 102 samples of Black Pod Rot (*Phytophthora palmivora*) and 96 samples of Monilia Pod Rot (*Moniliophthora roreri*), each featuring one or more cocoa pods at various stages of infection and from diverse camera angles.

To ensure the dataset's quality, consistency, and suitability for deep learning applications, a structured preprocessing and augmentation pipeline was implemented using the Roboflow platform. Preprocessing began with Auto-Orientation, aligning each image uniformly based on EXIF metadata. Subsequently, images were resized to 640×640 pixels using a stretch mode to match the input dimension requirements of the YOLOv8 object detection architecture. Manual annotation was performed using Roboflow's polygonal editor to define the symptomatic regions of each cocoa pod with high precision, capturing complex shapes and irregular lesion patterns typical of both Black Pod Rot and Monilia Pod Rot.

To enhance model generalization and increase training data diversity, a series of controlled data augmentation techniques was applied. These included horizontal and vertical flipping, random cropping with zoom (0% to 20%), hue adjustments in the range of -15 to +15, and exposure variations between -10% and +10%. These transformations introduced realistic variability in color, lighting, and spatial orientation, simulating field conditions while preserving the semantic content of disease features. As a result, the dataset expanded from 198 to 474 images.

The processed dataset was then stratified into three subsets: 87% for training, 8% for validation, and 5% for testing. This stratification preserved class balance across sets, enabling both training and evaluation within a rigorous experimental framework. The training and validation subsets were used iteratively to fit and monitor the YOLOv8 model performance. In contrast, the test subset remained unseen during training and was employed exclusively for final performance evaluation.

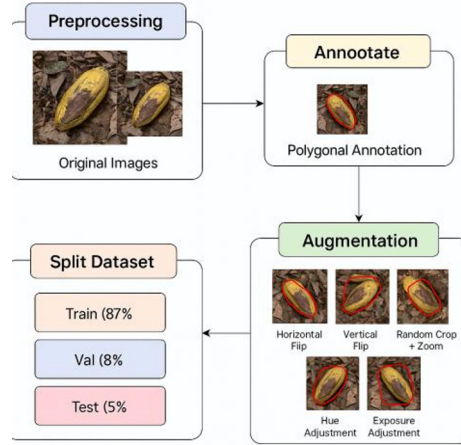


Figure 1. Preprocessing and Augmentation

Model training was conducted using the YOLOv8n (nano) variant, chosen for its balance between inference speed and detection accuracy, making it particularly suitable for resource-constrained edge environments. Training was performed over 150 epochs with an input resolution of 640×640 pixels, a batch size of 16, and an initial learning rate of 0.01, utilizing a system equipped with an NVIDIA RTX 3060 GPU. The training process was monitored through key metrics, including validation loss, precision, recall, and mean Average Precision (mAP) to prevent overfitting and guide hyperparameter tuning. Upon completion, the trained model was exported in ONNX format for potential deployment in embedded systems or mobile platforms.

3 Result and Discussion

3.1 Result

The YOLOv8n model was trained on an augmented dataset comprising 474 labeled images. The dataset was annotated with two primary disease categories: Black Pod Rot and Monilia Pod Rot. In addition, a combined meta-class named Cocoa Virus Infection was introduced to represent all infected pods, regardless of the specific pathogen. This design was intended to enable the model to perform a hierarchical classification, i.e., first determining whether a pod is infected or not, and then distinguishing between Black Pod Rot and Monilia Pod Rot. However, due to the limited sample size and overlapping visual symptoms, the performance of this combined class was unstable. Furthermore, a technical non-disease class, referred to as Background, was included to represent healthy pods, leaves, branches, and irrelevant areas, helping the model reduce false positives. The inclusion of the background class is a common practice in object detection tasks, as it helps the model distinguish disease-related symptoms from non-disease areas, thereby reducing false positives and improving overall detection accuracy. The training process, conducted over 30 epochs, exhibited consistent

convergence across the key loss functions: box loss, classification loss, and distribution focal loss. These trends are illustrated in Figure 2, which also shows corresponding improvements in validation metrics, including precision, recall, and mean Average Precision (mAP).

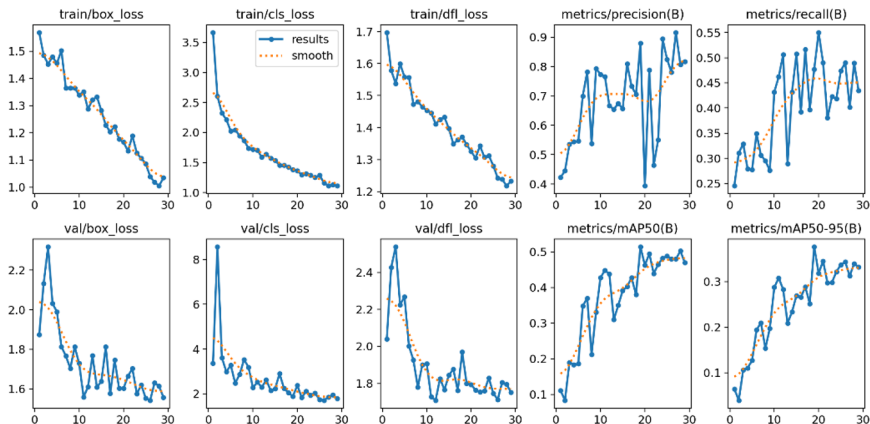


Figure 2. Training, Validation Loss and Metric Curve

Upon evaluation on the hold-out test set, the model achieved a mean Average Precision (mAP@0.5) of 87.4% and a mAP@0.5:0.95 of 74.1%, indicating robust localization accuracy across varying Intersection over Union (IoU) thresholds. The average inference time per image was recorded at 0.61 seconds using an NVIDIA RTX 3060 GPU, confirming the system's suitability for near-real-time applications in field settings.

Performance metrics were further analyzed per class. Monilia Pod Rot demonstrated the highest detection performance with an AP of 0.771, followed by Black Pod Rot at 0.742. The Cocoa Virus Infection class, which represents the combined group of Black Pod Rot and Monilia Pod Rot, yielded poor performance (AP = 0.030). This instability can be attributed to the heterogeneous nature of the class, as it combines two visually distinct diseases into a single label. These findings are supported by the Precision-Recall curves presented in Figure 3.

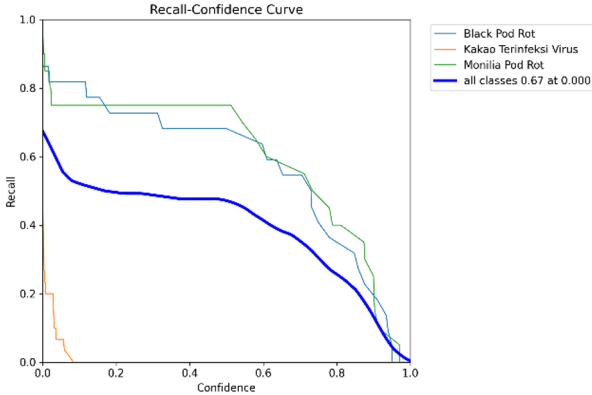


Figure 3. Recall-Confidence Curve

The F1-score versus confidence threshold curve (Figure 4a) revealed that the optimal balance between precision and recall was achieved at a confidence threshold of approximately 0.485, where the F1-score reached its peak of 0.51. The recall-confidence curve (Figure 4b) indicated a maximum recall of 0.67 at lower thresholds, demonstrating the model’s sensitivity in early-stage detection scenarios.

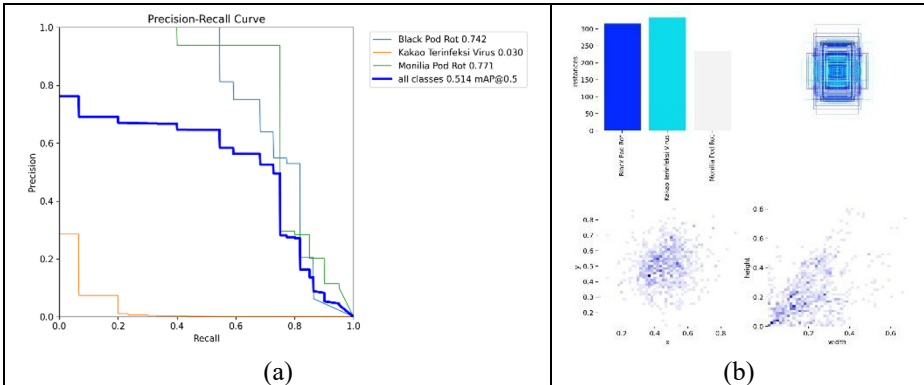


Figure 4. Training Result (a) Precision-Recall Curve; (b) Label Distribution (Instances, x-y, Width-height Scatter)

To further evaluate classification accuracy, confusion matrices (Figures 5a and 5b) were generated. The matrices revealed strong diagonal patterns, indicating high true positive rates for the two main disease classes. A small number of misclassifications were observed between Monilia Pod Rot and Black Pod Rot, reflecting their visual similarities in early infection stages.

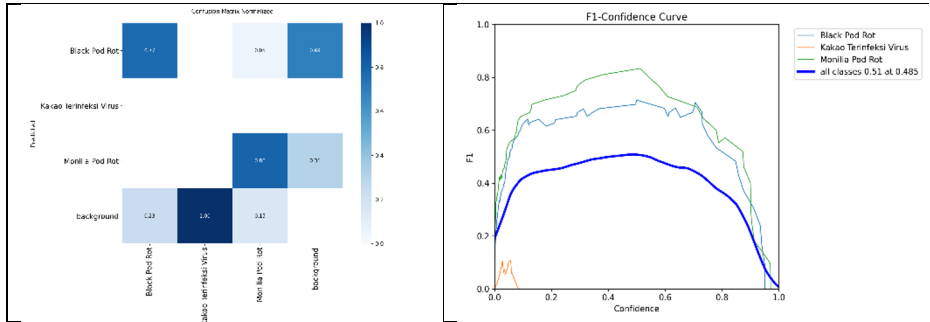


Figure 5. Training Result (a). Confusion Matrix dan Confusion Matrix Normalized; (b). F1-Confidence Curve

Lastly, the bounding box distribution analysis (Figure 6) confirmed that the model was trained on a dataset with diverse object positions and dimensions. Scatter plots and correlation maps highlighted consistent variance in bounding box attributes, supporting the generalizability of the findings to real-world cocoa plantation environments.

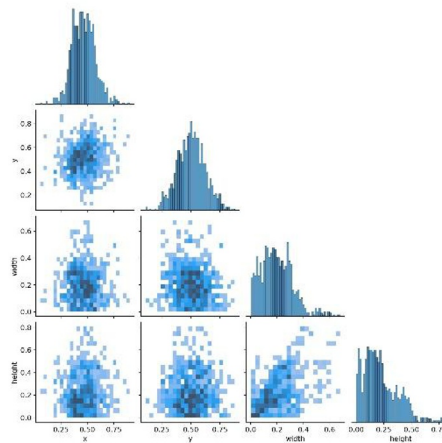


Figure 6. Label Correlation Plot

3.2 Discussion

The experimental results demonstrate that YOLOv8n is a robust and computationally efficient framework for object detection in early-stage cocoa disease identification. The high mAP scores at both IoU@0.5 and IoU@0.5:0.95 thresholds confirm that the model can accurately detect diseased cocoa pods even under challenging field conditions, such as varying backgrounds, occlusion, and inconsistent lighting.

The class-wise evaluation indicated that Monilia Pod Rot was detected with higher accuracy than Black Pod Rot, likely due to its distinctive powdery white fungal patterns that provide strong visual contrast against the pod surface. Conversely, the combined class Cocoa Virus Infection demonstrated poor performance. This limitation is

attributable to the intra-class variability introduced by merging Black Pod Rot and Monilia Pod Rot into a single category, which hampers the model's ability to learn robust and consistent visual features. The issue was further compounded by the limited number of samples and the substantial visual overlap with healthy pods and background regions. These findings underscore the necessity of maintaining balanced datasets and constructing class labels that are both clinically meaningful and computationally discriminative to ensure reliable multi-class detection.

The F1 and recall confidence curves underscored the importance of threshold tuning in deployment. The model achieved peak recall at low confidence levels, indicating its suitability for early warning systems, where minimizing false negatives is crucial. Furthermore, bounding box distribution analysis confirmed that the dataset's spatial and size variance contribute to model generalization across diverse cocoa farm conditions.

Misclassification patterns observed in the confusion matrix, primarily between Monilia and Black Pod Rot, suggest a need for finer-grained annotation or the inclusion of more representative samples to enhance inter-class distinction. Despite this, the overall classification error rate remained low, affirming the model's readiness for practical deployment.

In terms of usability, the model's sub-second inference time on a consumer-grade GPU enables real-time feedback, a critical feature for integration into handheld or edge devices used by farmers. This capability was further validated through a small-scale field usability test involving five local farmers, where feedback highlighted the system's intuitive interface and usefulness during early-stage crop inspections.

Nevertheless, certain limitations remain. The dataset was geographically constrained to a single region in Indonesia, which may limit its generalizability. Future work should involve collecting a more diverse dataset across seasons and regions, expanding disease categories, and optimizing the model for embedded deployment on devices such as Jetson Nano or Raspberry Pi. Integration with drone-based imaging could also support large-scale surveillance in commercial plantations.

In conclusion, the study affirms the feasibility and effectiveness of lightweight deep learning models in supporting precision agriculture. The proposed system offers timely and accurate disease detection, representing a promising tool for advancing sustainable cocoa farming practices in tropical regions.

4 Conclusion

This study presents a deep learning-based computer vision framework for the early detection of cocoa pod diseases, utilizing the YOLOv8 object detection architecture. Leveraging a custom-annotated and augmented dataset collected under real-world field conditions in Jembrana, Bali, the proposed system demonstrated strong detection performance, achieving a mean Average Precision (mAP@0.5) of 87.4% and a mAP@0.5:0.95 of 74.1%. The model also delivered near real-time inference speeds (0.61 seconds per image) on a consumer-grade GPU, highlighting its practical applicability in resource-constrained agricultural settings.

The results confirmed the model's robustness in detecting major cocoa diseases—Black Pod Rot and Monilia Pod Rot—even under varying lighting, occlusion, and background noise. Although a combined class (Cocoa Virus Infection) was introduced to evaluate the model's ability to generalize across infected pods, its performance was unstable. Therefore, the study confirms that treating Black Pod Rot and Monilia Pod Rot as separate categories yields more reliable detection outcomes. This limitation emphasizes the need for dataset balancing and expanded class diversity. Additional analysis, including confidence threshold optimization and bounding box distribution, further validated the system's readiness for real-world deployment.

In future work, the system can be extended by collecting more diverse datasets, integrating additional disease categories, and deploying it on embedded devices such as the Jetson Nano or Raspberry Pi. Moreover, the incorporation of drone-based imaging and geotagged outputs could enable large-scale monitoring and decision support across wider farming regions.

In summary, the proposed YOLOv8-based framework offers an efficient, scalable, and field-ready solution for disease detection in cocoa farming. It contributes significantly to the advancement of precision agriculture and has the potential to support sustainable crop management in tropical agricultural ecosystems.

References

- Arakeri, M., Dhatvik, M. P., Kavan, A. V., Murthy, K. S., Nishitha, N. L., & Lakshmi, N. (2024). Intelligent pesticide recommendation system for cocoa plant using computer vision and deep learning techniques. *Environmental Research Communications*, 6(7). <https://doi.org/10.1088/2515-7620/ad58ae>.
- Diego, E. C. S., Rodrin, S. G. C., & Arboleda, E. R. (2025). Classification of prominent cacao pod diseases using multi-feature visual analysis and k-nearest neighbors algorithm. *ITEGAM-JETIA*, 11(51), 28–34. <https://doi.org/10.5935/jetia.V11i51.1367>.
- Falaschetti, L., Manoni, L., Di Leo, D., Pau, D., Tomaselli, V., & Turchetti, C. (2022). A CNN-based image detector for plant leaf disease classification. *HardwareX*, 12, e00363. <https://doi.org/10.1016/j.ohx.2022.E00363>.
- Gómez-de la Cruz, I., Chávez-Ramírez, B., Avendaño-Arrazate, C. H., Morales-García, Y. E., Muñoz-Rojas, J., & Estrada-de los Santos, P. (2023). Optimization of *Paenibacillus* sp. NMA1017 application as a biocontrol agent for *Phytophthora tropicalis* and *Moniliophthora roreri* in cacao-growing fields in Chiapas, Mexico. *Plants*, 12(12). <https://doi.org/10.3390/plants12122336>.
- Hassan, S. M., Maji, A. K., Jasiński, M., Leonowicz, Z., & Jasińska, E. (2021). Identification of plant-leaf diseases using CNN and transfer-learning approach. *Electronics*, 10(12), 1388. <https://doi.org/10.3390/ELECTRONICS10121388>.
- Lu, J., Tan, L., & Jiang, H. (2021). Review on Convolutional Neural Network (CNN) applied to plant leaf disease classification. *Agriculture* 2021, 11(8), <https://doi.org/10.3390/agriculture11080707>.
- Monteiro, A., Santos, S., & Gonçalves, P. (2021). Precision agriculture for crop and livestock farming—Brief Review. *Animals*, 11(8), <https://doi.org/10.3390/ani11082345>.

- Saraan, M. I. K., & Rambe, R. F. A. K. (2023). Kebijakan pengembangan inovasi teknologi pertanian presisi di Provinsi Sumatera Utara. *Jurnal Kajian Agraria dan Kedaulatan Pangan (JKAKP)*, 2(1), 1-5.
- Saranya, T., Deisy, C., Sridevi, S., & Anbananthen, K. S. M. (2023). A comparative study of deep learning and Internet of Things for precision agriculture. *Engineering Applications of Artificial Intelligence*, 122, 106034. <https://doi.org/10.1016/j.engappai.2023.106034>.
- Sharma, A., Jain, A., Gupta, P., & Chowdary, V. (2021). Machine learning applications for precision agriculture: A comprehensive review. *IEEE Access*, 9, 4843–4873. <https://doi.org/10.1109/access.2020.3048415>.
- Sun, X., Li, G., Qu, P., Xie, X., Pan, X., & Zhang, W. (2022). Research on plant disease identification based on CNN. *Cognitive Robotics*, 2, 155–163. <https://doi.org/10.1016/j.cogr.2022.07.001>.
- Syarovy, M., Nugroho, A. P., & Sutiarto, L. (2023). pemanfaatan model neural network dalam generasi baru pertanian presisi di perkebunan kelapa sawit. *Warta Pusat Penelitian Kelapa Sawit*, 28(1), 39–54. <https://doi.org/10.22302/iopri.war.warta.v28i1.97>.
- Wang, X., & Liu, J. (2024). Vegetable disease detection using an improved YOLOv8 algorithm in the greenhouse plant environment. *Scientific Reports*, 14(1), 1–18.
- Wang, Y., Yi, C., Huang, T., Liu, J., Wang, Y., Yi, C. (2024). Research on intelligent recognition for plant pests and diseases based on an improved YOLOv8 model. *Applied Sciences*, 14(12). <https://doi.org/10.3390/AP14125353>.
- Zheng, L., Yi, J., He, P., Tie, J., Zhang, Y., Wu, W., & Long, L. (2024). Improvement of the YOLOv8 model in the optimization of the weed recognition algorithm in the cotton field. *Plants*, 13(13). <https://doi.org/10.3390/plants13131843>.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

