Spices Identification in Essential Oil Producers using Comparasion of KNN and Naïve Bayes Classifier

Fifin Ayu Mufarroha [1], Achmad Zain Nur [1], Mohammad Rizal Rahabillah [1], Achmad Jauhari [1], Devie Rosa Anamisa [1], and Mulaab [1]

[1] University of Trunojoyo Madura, Bangkalan, Indonesia
fifin.mufarroha@trunojoyo.ac.id

Abstract. Indonesia is a spice-growing country, providing a variety of spices with numerous health advantages. Aside from being a producer, Indonesia is the world's largest supplier of spices. Spices have a wide range of usage, including food ingredients, herbal medicines, and essential oils. Essential oils are generally used as binders in the aromatherapy, perfume, cosmetic and pharmaceutical manufacturing industries. With so many types of essential oil production, it is necessary to know which spices are ingredients in the production of the appropriate types of essential oils, so that a classification system for types of spices is desirable. Machine learning was utilized in this study to analyze spice's image. Machine learning's K-NN and Naive Bayes algorithms were selected as classification techniques. The goal of this study is to identify spices using a machine learning method, which is projected to develop into a system that assists farmers and the larger community in growing or creating essential oils that are suited by employing the right spices. The K-NN method achieved better accuracy with a value of K = 3 obtaining an accuracy of 100%, while Naïve Bayes achieved 96% accuracy. This research highlights the need for a classification system to improve the quality of essential oils for farmers and communities.

Keywords: Spices, Essential Oils, Classification, Identification of spices, machine learning.

1 Introduction

Indonesia has a very diverse biodiversity and produces various types of spice families because it is a tropical country. Tropical or tropical climate in the sense that Indonesia has 2 seasons, namely rainy and dry season [1][2]. The diversity of types of spice plants is estimated to number 150 – 200 species which consist of various plant families such as Pinaceae, Labiateae, Compositae, Lauraceae, Myrtaceae, and Umbelliferaceae. The spice market share is very broad, this can be seen from the attractiveness of spices in the European region. It is known that spice plants have characteristics, namely a unique aroma and taste in each type. Based on these characteristics, spices are used in a variety of products, ranging from food preservatives, food and beverage flavours, herbal medicines, and essential oils and their derivatives [3]. Aromatherapy offers a wide range of products including essential oils, candles, spas, body lotions, perfumes

© The Author(s) 2023
M. Hartono et al. (eds.), *Proceedings of the 4th International Conference on Informatics, Technology and Engineering 2023 (InCITE 2023)*, Atlantis Highlights in Engineering 21,
https://doi.org/10.2991/978-94-6463-288-0_51

and other health remedies. In addition, it can also be used in processed foods that function as preservatives and herbal medicines as natural remedies. This spice is usually mixed with green tea, white rice, ginger, and vetiver to create a more refreshing scent than perfume. Spices are used as herbal remedies [4] such as ginger is used for colic and diarrhea [5] , while turmeric can help with digestive problems and cumin can significantly reduce stomach ache and gas [6] and many other types of spices. This herb is commonly used as a cooking spice. Because spices have strong aroma or taste characteristics. Essential oils can be derived from the entire plant, including the fruit, stems, bark, roots, flowers, leaves, rhizomes, and seeds[7][8][9].

Essential oils are scents found in plants [10][11]. This oil is also called essential oil because it easily evaporates at room temperature. The term essential oil is used to describe the aroma of plant it comes from. Essential oils are generally used as binders in pharmaceutical, cosmetic, perfume, fragrance or perfume manufacturing industries [12]. Essential oils are also used as spices such as cloves and cubes. Almost all volatile substances are antibacterial. Some essential oils that can be used as preservatives also have more specific functions, such as citronella oil which is known for its anti-mosquito function, nutmeg oil as an anti-inflammatory, eucalyptus oil as an anti-irritant, ginger as a stimulant, analgesic, anti-inflammatory, kaffir lime oil as anti-depression and gaharu as anti-rheumatism.

With development of a very modern era, the way to identify type of essential oil production no longer requires conventional methods, searching for information via internet or asking experts but can use the help of the system. Classification is an area of machine learning that performs class or group division procedures based on predefined rules [13]. The field of artificial intelligence known as "machine learning" focuses on creating computer programs that can access data and use it for self-learning[14] [15]. The benefit of doing this process is that it helps in studying the diversity of spices in the production of essential oils better and more accurately. The combination of information technology, artificial intelligence and knowledge of spices produces a system that is able to meet the needs of automatically identifying and classifying suitable spices in the production of suitable essential oils. The hope is that by applying the machine learning method to build a system with great accuracy for identifying spices so that the system can help farmers and the public understand types of spices for production of the right essential oils.

2 Literature Review

Spices have therapeutic effects and are widely used in cooking. Many ways have been developed to classify spices due to their diversity and identity. Herbs and spices are defined and classified, as well as their commerce and applications, are discussed by Peter et al [16]. Herbs are the dried leaves of aromatic plants used to flavor and odor foods, whereas spices are the dried sections of aromatic plants that do not include the leaves. The taxonomic taxonomy of spices is shown, as well as illustrations of the various plant parts utilized to generate herbs and spices. The essay also emphasizes the significance of growing high-quality, clean spices while using as few artificial

fertilizers and pesticides as possible. With rising consumer demand in importing countries for more unusual, ethnic flavours in food, the worldwide trade in spices is likely to grow. Machine learning is used by Manojkumar et al to detect and classify different spices based on their scents [17]. The traditional method of detecting odors using human noses has disadvantages, such as being time-consuming and imprecise. To circumvent these restrictions, electronic noses (E-Noses) have been developed. The experimental setup and procedures used to train and evaluate the Random Forest classifier are thoroughly described in the study.

3 Research Method

The built system aims to identify essential oil production using machine learning methods. Classification was carried out on 5 types of spices including Galangal, Curcuma, Aromatic Ginger, Ginger, and Turmeric. Figure 1 displays the dataset of the five types of spices used in this study.

(a) (b) (c) (d) (e)

Fig. 1. Spices: (a) Galangal, (b) Curcuma, (c) Aromatic ginger, (d) Ginger, (e) Turmeric

This research is divided into 2 processes, namely model development and model implementation. Figure 2 depicts the model building procedure. The procedure starts by separating the test and training groups in the data. From the size of the data obtained has high resolution, then the image size is changed. Referring to the learning data, after the image size has been reduced, the feature extraction stage is carried out to obtain the image value that characterizes the spice. Previously, image conversion was carried out by converting RGB images to grayscale images with equation 1:

$$I = (0.298 \times Re) + (0.5870 \times Gr) + (0.1140 \times Bl) \quad (1)$$

Based on the type of spice, the way to get the features of the spice is by utilizing light intensity. The classification stage is carried out by applying machine learning methods including K-NN and Naïve Bayes. The use of two machine learning methods to discover which method produces the best accuracy results. The K-NN method is a simple method by utilizing the value of K as a determinant of nearest neighbors number in the classification process. Calculation of shortest distance in this method uses the Euclidean distance. The distance d between two points, x and y, is known as the Euclidean distance. The distance calculation equation can be seen in general in Equation 2 [19]:

$$EDist = \sqrt{(X_1 - Y_1)^2 + (X_2 - Y_2)^2 + \cdots + (X_n - Y_n)^2} == \sqrt{\sum_{i=1}^{n}(X_i - Y_i)^2} \quad (2)$$

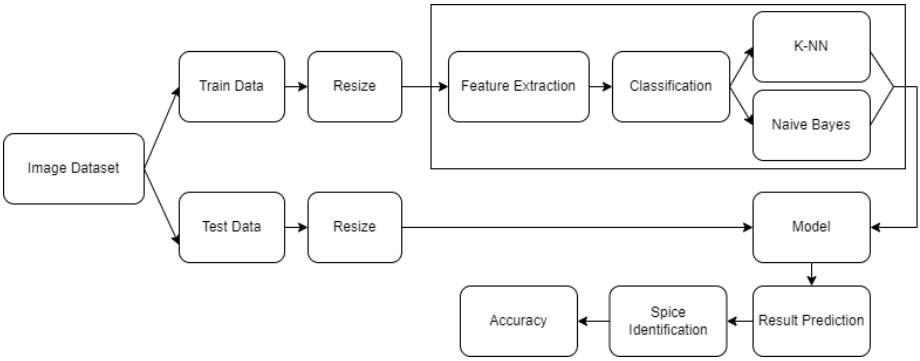


Fig. 2. Classification process model

The Naive Bayes classifier is a probabilistic machine learning model that is used to classify data. The Bayes theorem underpins the classifier's core. The Bayes' theorem equation is described in Equation 3. Using Bayes' theorem, we may calculate the likelihood that A will occur if B has already occurred. In this case, B is the proof, while A is the hypothesis. In this situation, the predictors/features are supposed to be independent. This implies that the presence of one attribute has no effect on the presence of the others.

$$XP(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{3}$$

KNN and Naive Bayes classifiers are straightforward algorithms that are simple to comprehend and apply. They do not use sophisticated neural network topologies or deep learning frameworks. If the problem can be satisfactorily solved with a less complex model, then simplicity can be useful. Deep learning models can need large computational resources, particularly during training. KNN and Naive Bayes are less computationally intensive and may be applied to smaller hardware configurations, making them appropriate for resource-constrained applications. When the feature extraction method is simple and does not require sophisticated learnt features, KNN and Naive Bayes classifiers can function effectively in some circumstances[20]. Deep learning models excel in learning detailed features from raw data, but if the task can be completed with simpler feature extraction approaches, such as KNN or Naive Bayes, the latter may suffice. Some of the benefits listed above are motivations for implementing the two ways.

The model that has been formed will be tested with test data. As a result, the prediction of the spice class between the test data and the pre-processed training data. In addition, the results obtained from the identification of spices in essential oil production are based on the classification results. The K-NN and Naïve Bayes evaluation models were used to determine the classification method's performance. The evaluation stage seeks to test the accuracy of the created classification system. Using the convolution matrix was chosen as the classification calculation method. The Confusion Matrix depicts a description of the prediction outcomes categorization problem,

with the number of correct and incorrect predictions totaled and separated by class[21]–[24].

The second stage is implementing the system by creating a Graphical User Interface (GUI). Utilization of GUI to facilitate the implementation of pre-designed process model and used for communication between user and device. GUI with interactive visual advantages can present image objects and represent information on classification results and spices. The system is developed on a desktop platform with Python as the programming language of choice.

4 Result and Discussion

It has been described in the previous session that the creation of a classification system is divided into 2 processes. In this session will be discussed one by one how the completion and results of the two processes.

4.1 Model Development

The dataset in this study comes from https://www.kaggle.com/code/kerneler/starter-spice-173499e2-4/data with a total data set of 150 images. The dataset used in this study had 100 training data and 50 test data. Each class has 20 details in a set of 100 training data. The number of classes reflects to the number of spices used in the study. As previously stated, the spices chosen are Galangal, Curcuma, Aromatic Ginger, Ginger, and Turmeric, resulting in a total of 5 classes (class 1 through class 5). Figure 1 depicts the dataset image circumstances, with the image background set to white. In image processing, having a white background has various advantages, including minimizing distractions and making it easier to choose and focus on the subject of the image. White backgrounds also allow images to be saved in lesser file sizes, which can have a significant impact on computation.

The data will be reduced from the original image size of 1773×2374. This process is a process of resizing the size of the original image to facilitate model formation and speed up computation time. Before getting image characteristics, the image is transformed to grayscale. By utilizing the intensity of light in the image, features are obtained. The process is continued by carrying out classification using K-NN and Naïve Bayes methods.

In the K-NN method, 3 scenarios were applied using K = 1, 3, and 6. The value of K is significant since it influences the model's performance. When K is too low, the model becomes overly specialized and fails to generalize well; it is also susceptible to noise. If K is too large, the model becomes too generic and may fail to describe the data's local structure. The values 1, 3, and 6 are widely utilized since they offer a variety of alternatives for balancing these aspects. The best value of K, on the other hand, is determined by the dataset and the issue being solved. The K value is an important parameter in the KNN method, and must select an acceptable value of K to produce the greatest model performance. Table 1 shows the classification results of the K-NN and Nave Bayes methods.

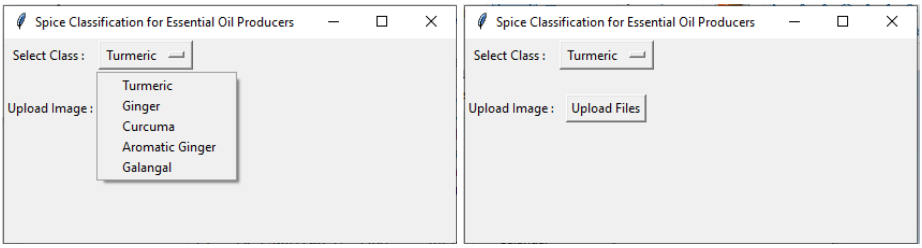
Table 1. Result Classification

Method	Accuracy (%)	Precision	Recall	f-1 Score
KNN (K =1)	96%	96%	100%	98%
KNN (K =3)	100%	100%	100%	100%
KNN (K =6)	95%	95%	100%	97%
Naïve Bayes	96%	97%	96%	96%

4.2 Model Implementation

The implementation stage is carried out by building a classification system to identify spices for essential oil production. The system created classifies identification types of spices with input in the form of a spice image. A desktop-based system built with several features provided, including (a) class determination feature, (b) image upload feature, (c) K-NN and Naïve Bayes classification feature, this feature functions for user can choose to classify the K-NN model or the Naïve Bayes model (Fig. 5), (d) Result features

The features for determining class are shown in Figure 3a, information is presented namely class determination. Class determination is intended for users can choose the class that they want to assign in the classification process. Later, choosen image will be compared with uploading images. The system will access the local drive where user can search and select the image to be tested (Figure 3b). It makes easy for users to select the test images. If user has specified a test image from the local drive, the system displays that image. With the representation of test images in the system, the user can reassure that the test images are in accordance with the user's intent. As in the model design, user can carry out the classification process shown in Figure 5.



(a) (b)

Fig. 3. Displays: (a) Class determination, (b) Upload image

Fig. 4. View test data and classification

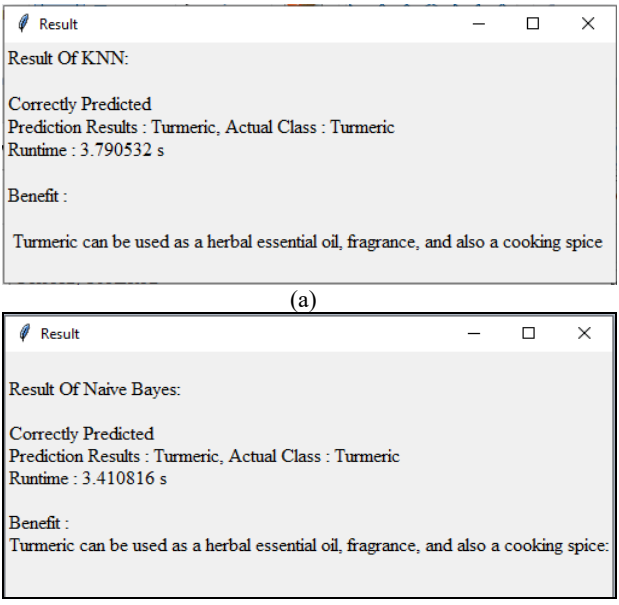


Fig. 5. Results: (a) K-NN (b) Naïve Bayes

In Figure 4, there are 2 buttons representing 2 classification methods, namely KNN method classification or naïve Bayes method classification. When user has made a choice of which method to use in the classification process, the system will display the results which can be seen in Figure 5. This feature functions to display the classification results, both on the K-NN model and the Naïve Bayes model. The user can see the predicted results of the classification method chosen (explanation of Figure 4) based on the predetermined class (explanation of Figure 3). In addition, the system will show the system computation time on the method and when the input image of

spices has produced a predicted class, information on essential oil production will also be displayed.

Based on the results of the classification and identification of spices, it can be estimated that the computation time for the naïve Bayes method has a faster computation time. The K-NN method processes 0.3 seconds longer with a time of 3.790532 seconds compared to Naïve Bayes with a computational time of 3.410816 seconds.

5 Conclusion

The urgency of this research is that farmers and the community need a classification system to identify spices in the manufacture of essential oils. The quality or quality of essential oils is determined by the natural characteristics of the constituent materials and other ingredients that are mixed. Determination of the right material will determine the quality of the oil. With the help of advanced technology, a classification system has been built that helps in identifying types of spices. Methods in machine learning namely K-NN and Naïve Bayes have been chosen as classification methods. The classification results were obtained using the K-NN approach with three scenarios, namely with a K value of 1, 3, and 6. A value of K = 1 provides an accuracy of 96%, a value of K = 3 provides an accuracy of 100%, and a value of K = 6 provides an accuracy of 95%. The Nave Bayes approach produced classification results with an accuracy of 96%. Comparison of the accuracy results obtained for the two classification methods can be concluded that the K-NN method is better than the accuracy results of Naïve Bayes.

Acknowledgment

Thank you for the opportunity to conduct research with contract number 414/UN46.4.1/PT.01.03/2022 at the University of Trunojoyo Madura. The author also acknowledges the Sumenep office of the Department of Agriculture and Horticulture for their cooperation in supplying data for this research.

References

[1] A. Efendi, A. Nugraha, and R. Baharta, “Manufacturing of Electrical Dryer Machine for Food and Fruit Products,” *IOP Conf Ser Mater Sci Eng*, vol. 692, no. 1, p. 12006, 2019, doi: 10.1088/1757-899X/692/1/012006.

[2] S. B. Kurniawan and M. F. Imron, “Seasonal variation of plastic debris accumulation in the estuary of Wonorejo River, Surabaya, Indonesia,” *Environ Technol Innov*, vol. 16, p. 100490, 2019, doi: https://doi.org/10.1016/j.eti.2019.100490.

[3] M. Karmakar *et al.*, “Immunostimulant Properties of Some Commonly Used Indian Spices and Herbs with Special Reference to Region-Specific Cuisines BT - Plants and Phytomolecules for Immunomodulation: Recent Trends and Advances,” N. S.

Sangwan, M. A. Farag, and L. V. Modolo, Eds., Singapore: Springer Nature Singapore, 2022, pp. 191–249. doi: 10.1007/978-981-16-8117-2_8.

[4] R. Pavela *et al.*, "Traditional herbal remedies and dietary spices from Cameroon as novel sources of larvicides against filariasis mosquitoes?," *Parasitol Res*, vol. 115, no. 12, pp. 4617–4626, 2016, doi: 10.1007/s00436-016-5254-4.

[5] M. N. Ghayur and A. H. Gilani, "Pharmacological Basis for the Medicinal Use of Ginger in Gastrointestinal Disorders," *Dig Dis Sci*, vol. 50, no. 10, pp. 1889–1897, 2005, doi: 10.1007/s10620-005-2957-2.

[6] S. Agah, A. M. Taleb, R. Moeini, N. Gorji, and H. Nikbakht, "Cumin extract for symptom control in patients with irritable bowel syndrome: a case series.," *Middle East J Dig Dis*, vol. 5, no. 4, pp. 217–222, Oct. 2013.

[7] J. A. Jamal, Z. Abd. Ghafar, and K. Husain, "Medicinal Plants used for Postnatal Care in Malay Traditional Medicine in the Peninsular Malaysia," *Pharmacognosy Journal*, vol. 3, no. 24, pp. 15–24, 2011, doi: https://doi.org/10.5530/pj.2011.24.4.

[8] G. M. W. Lengai, J. W. Muthomi, and E. R. Mbega, "Phytochemical activity and role of botanical pesticides in pest management for sustainable agricultural crop production," *Sci Afr*, vol. 7, p. e00239, 2020, doi: https://doi.org/10.1016/j.sciaf.2019.e00239.

[9] L. Felix-Cuencas *et al.*, "Chapter 8 - Bioactivity characterization of herbal molecules," S. C. Mandal, A. K. Nayak, and A. K. B. T.-H. B. in H. A. Dhara, Eds., Academic Press, 2022, pp. 145–183. doi: https://doi.org/10.1016/B978-0-323-85852-6.00007-X.

[10] P. S. Vankar, "Essential oils and fragrances from natural sources," *Resonance*, vol. 9, no. 4, pp. 30–41, 2004, doi: 10.1007/BF02834854.

[11] H. Loza-Tavera, "Monoterpenes in Essential Oils BT - Chemicals via Higher Plant Bioengineering," F. Shahidi, P. Kolodziejczyk, J. R. Whitaker, A. L. Munguia, and G. Fuller, Eds., Boston, MA: Springer US, 1999, pp. 49–62. doi: 10.1007/978-1-4615-4729-7_5.

[12] L. Lesage-Meessen, M. Bou, J.-C. Sigoillot, C. B. Faulds, and A. Lomascolo, "Essential oils and distilled straws of lavender and lavandin: a review of current use and potential application in white biotechnology," *Appl Microbiol Biotechnol*, vol. 99, no. 8, pp. 3375–3385, 2015, doi: 10.1007/s00253-015-6511-7.

[13] A. Fernández, S. García, J. Luengo, E. Bernadó-Mansilla, and F. Herrera, "Genetics-Based Machine Learning for Rule Induction: State of the Art, Taxonomy, and Comparative Study," *IEEE Transactions on Evolutionary Computation*, vol. 14, no. 6, pp. 913–941, 2010, doi: 10.1109/TEVC.2009.2039140.

[14] J. G. Carbonell, R. S. Michalski, and T. M. Mitchell, "1 - AN OVERVIEW OF MACHINE LEARNING," R. S. Michalski, J. G. Carbonell, and T. M. B. T.-M. L. Mitchell, Eds., San Francisco (CA): Morgan Kaufmann, 1983, pp. 3–23. doi: https://doi.org/10.1016/B978-0-08-051054-5.50005-4.

[15] J. Shabbir and T. Anwer, "Artificial Intelligence and its Role in Near Future," vol. 14, no. 8, pp. 1–11, 2018.

[16] K. V. Peter and M. R. Shylaja, "Introduction to herbs and spices: Definitions, trade and applications," in *Handbook of Herbs and Spices: Second Edition*, Elsevier Inc., 2012, pp. 1–24. doi: 10.1533/9780857095671.1.

[17] M. Kukade, T. Karve, and D. Gharpure, "Identification and classification of spices by machine learning," in *Proceedings - 2019 IEEE International Conference on Intelligent Systems and Green Technology, ICISGT 2019*, Institute of Electrical and Electronics Engineers Inc., Jun. 2019, pp. 1–4. doi: 10.1109/ICISGT44072.2019.00015.

[18] J. Pages-Rebull, C. Pérez-Ràfols, N. Serrano, M. del Valle, and J. M. Díaz-Cruz, "Classification and authentication of spices and aromatic herbs by means of HPLC-UV and chemometrics," *Food Biosci*, vol. 52, Apr. 2023, doi: 10.1016/j.fbio.2023.102401.

[19] F. A. Mufarroha, D. R. Anamisa, and A. G. Hapsani, "Content Based Image Retrieval Using Two Color Feature Extraction," *J Phys Conf Ser*, vol. 1569, p. 32072, 2020, doi: 10.1088/1742-6596/1569/3/032072.

[20] S. D. Jadhav and H. P. Channe, "Comparative Study of K-NN, Naive Bayes and Decision Tree Classification Techniques," 2013. [Online]. Available: www.ijsr.net

[21] Ž. Vujović, "Classification model evaluation metrics," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 6, pp. 599–606, 2021.

[22] A. Tayal, J. Gupta, A. Solanki, K. Bisht, A. Nayyar, and M. Masud, "DL-CNN-based approach with image processing techniques for diagnosis of retinal diseases," *Multimed Syst*, pp. 1–22, 2021.

[23] Q. A. Al-Haija and A. A. Alsulami, "High performance classification model to identify ransomware payments for heterogeneous bitcoin networks," *Electronics (Basel)*, vol. 10, no. 17, p. 2113, 2021.

[24] A. Asokan, J. Anitha, M. Ciobanu, A. Gabor, A. Naaji, and D. J. Hemanth, "Image Processing Techniques for Analysis of Satellite Images for Historical Maps Classification—An Overview," *Applied Sciences*, vol. 10, no. 12, 2020, doi: 10.3390/app10124207.

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.