INTL JOURNAL OF ELECTRONICS AND TELECOMMUNICATIONS, 2024, VOL. 70, NO. 4, PP. 849–853
Manuscript received January 30, 2024; revised October 2024. doi: 10.24425/ijet.2024.152069

MFCC-Based Sound Classification of Honey Bees

Urszula Libal, and Pawel Biernacki

***Abstract*—Smart beekeeping is a rapidly developing field. Automated detection and classification of honey bees opens many new opportunities for studies on their behavior. In this paper, we focus on distinguishing between two classes of bees: female workers and male drones. The classification is performed on mel-frequency cepstral coefficients obtained for audio recordings of their flights in a close proximity to an entrance to a beehive. We compare the classification accuracy for several classifiers. We investigate how partitioning of the frequency spectrum influences the classification results. The study involves series of experiments performed for different cepstral representations in the form of 5, 10, 15, 20 and 40 mel-frequency cepstral coefficients.**

***Keywords*—Internet of Things (IoT); smart beekeeping; pattern recognition; signal processing; mel-frequency cepstral coefficients (MFCC)**

I. INTRODUCTION

THIS paper investigates different divisions of the frequency spectrum in a problem of classification of honey bees based on audio signals represented by mel-frequency cepstral coefficients (MFCC). The proper preprocessing of audio signals is usually crucial for the success in a wide range of pattern recognition tasks as speech recognition, music genre classification, or anomaly detection of a mechanical system, just to mention a few. The mentioned mel-frequency cepstrum has been proven to be an effective feature extraction method.

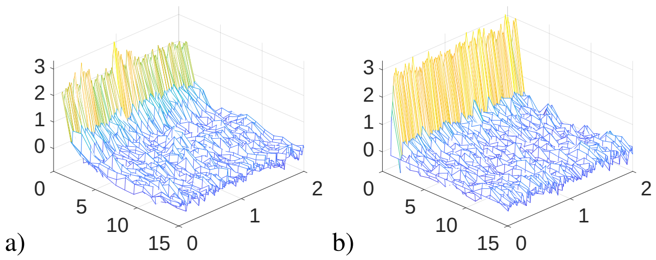

Fig. 1. Exemplary cepstrogram for audio signal from class: a) worker bee, b) drone bee, represented by 15 MFCCs.

The cepstrum of the exemplary signals from classes: worker bee and drone bee was generated by calculation of mel-frequency cepstral coefficients (MFCCs). Presentation of the frequency components on the Mel scale clearly shows that class recognition should be possible for that signal preprocessing.

U. Libal and P. Biernacki are with Department of Acoustics, Multimedia and Signal Processing, Wroclaw University of Science and Technology, Wroclaw, Poland (e-mail: {urszula.libal, pawel.biernacki}@pwr.edu.pl).

A. Internet of Things in beehive monitoring

In recent years, we observed a massive turn from manual beehive monitoring by beekeepers to applying the Internet of Things (IoT). The main goals are remote observation of the state of the bee colony, ensuring the safety and health control of the bees. Additionally, the more sophisticated methods for bee colonies allow for collecting large data sets and, thanks to that, studies on still mysterious honey bees' behaviors.

As listed in [1], [2], there are many commercial IoT systems and devices, for example:

- Beebot measures weight of a hive, as well as humidity and temperature inside.
- Easy Bee Counter counts bees leaving and returning to a hive using 48 IR sensors.
- Bee-Shop Camera Kit taking photos and video recordings, which can be stored on an SD card or sent to the beekeeper's mobile phone using the 3G/4G LTE network.
- EyeSon Hives with an image detection algorithm for analyzing the flight direction of the swarm. EyeS on Hives also uses 3G/4G LTE connectivity and allows real-time video monitoring of a hive.
- Zygi measures temperature, humidity, and weight.
- Hive-Tech detects swarming based on IR and reflectance sensors that detect real-time crowd conditions.
- HiveMind monitors the activity of bees using sound and IR sensors.

Another exciting example is a self-powered SBMaCS system [3] with temperature, humidity, weight, motion, and flame sensors. A system presented in [4] counts bees using a thermal camera. In that study, bees' flights were extracted by the 4 popular classifiers: k-nearest neighbors, neural network, random forest, and support vector machine. Bees classification with the use of an autoencoder neural network trained by MFCCs, as representations of audio samples, was proposed in [5], [6].

The AppMAIS project [7] was designed to better understand and effectively prevent colony collapse disorder, which leads to a rapid loss of adult worker bees and sudden colony mortality. The AppMAIS uses on the open-source IoT platform known as Thingsboard, which allows one to generate alerts after detection of a defined state of a colony. As pointed out in the survey in [8], the energy consumption and non-intrusiveness of the system are of essential importance.

The IoT is closely related to radio frequency identification (RFID) technology. As [9] noted, RFID tags can be applied for the continuous monitoring of individuals in a colony, which is less time consuming than watching video recordings. However, it involves mounting RFID on the backs of individual bees, which is problematic because of their size. Considering that an average colony counts from 20,000 to 80,000 bees, it would require a large number of RFID tags, and still we would receive only information about entering or leaving by a bee a specific area, defined by a location and parameters of an RFID reader.

In the following part of the article, we present the theoretical description of the extraction of MFCCs (in Section II), followed by the real-life signal classification results for several well-known classification algorithms (in Section III).

II. Theory

A. *Mel-frequency cepstral coefficients*

Mel-frequency cepstral coefficients (MFCCs) are very effective and widely applied, including the analysis of bee sounds and the audio signal representation method [10]–[12]. The extraction of MFCCs is performed in several steps. The steps for calculating MFCCs are presented in the diagram in Fig. 2 and are described in detail below.

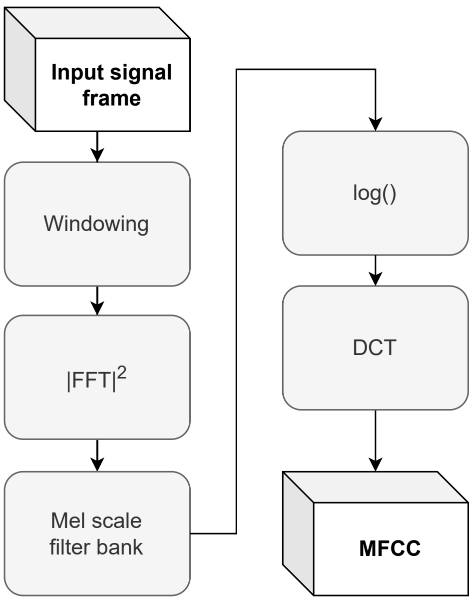

Fig. 2. Calculation of mel-frequency cepstral coefficients.

In the first step, each input signal frame is multiplied by a chosen window shape; in our case, it was the Hamming window. The goal is to prevent frequency leakage during the Fast Fourier Transform (FFT) calculation.

The energy spectrum S_m of a chosen frequency band is calculated by convolution of filter bank H_m with the energy spectral density (it is squared absolute value of power spectral density $X(f)$) of a signal frame in the following way:

$$S_m = \sum_{f=0}^{K-1} |X(f)|^2 * H_m(f), \quad (1)$$

where K is the length of power spectral density $X(f)$ of a signal frame, calculated by Fast Fourier Transform (FFT), (equal to the number of signal samples in the analyzed frame), $m = 1, \ldots, 40$ is a filter bank index.

The 40 filters $H_m(f)$ participate in the calculation of S_m by convolution with the square absolute value of the power spectral density $X(f)$ of a signal. The first 13 filters are linearly-spaced, and the last 27 are log-spaced. In Fig. 3, we present auditory filter banks $H_m(f)$, $m = 1, \ldots, 40$, for the frequency bandwidth 22,050 Hz.

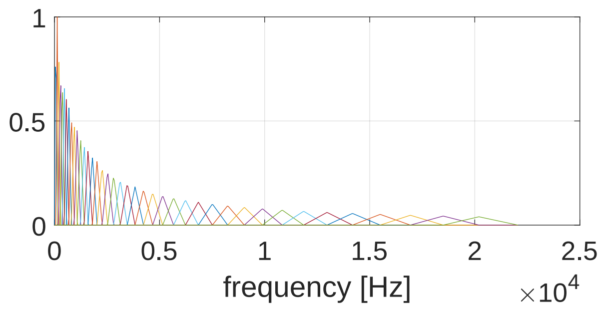

Fig. 3. Auditory filter banks H_m for bandwidth 22,050 Hz.

The mel-frequency cepstral coefficients c_i are obtained by the following transformation:

$$c_i = \sqrt{\frac{2}{M}} \sum_{m=1}^{M} log(S_m) cos\left(\frac{\pi i}{M}(m-0.5)\right), \quad (2)$$

where $M = 40$ is equal to the number of used filters and $i = 1, \ldots, n$ is the MFCC index. The formula (2) applies Discrete Cosine Transform (DCT) to logarithm of energy spectrum S_m. The base functions of the DCT are presented in Fig. 4. Each base function corresponding to the calculation of c_i, $i = 1, \ldots, n$, MFCCs. In the experiment, we extract $n = 5, 10, 15, 20$ and 40 MFCCs using the described method.

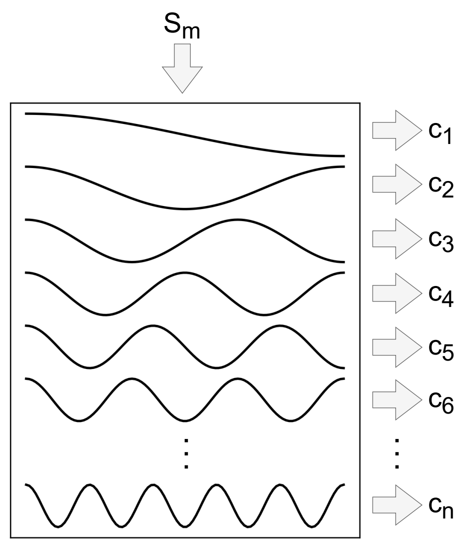

Fig. 4. Discrete cosine transform (DCT) base functions corresponding to calculation of c_i, $i = 1, \ldots, n$, MFCCs.

B. Supervised learning

To investigate the optimal representation of audio signals by MFCCs in the honey bee classification task, we applied supervised learning [15], [16]. Supervised learning means that the data set used for classifier training has to be labeled. In other words, all signals have to be tagged with one of class labels:

- worker bee,
- drone bee.

This approach allows for an easy check of classifier performance where the original (true) labels are compared with predicted (pointed out by a classifier) labels. Counting the differences in the labels for all signals from a testing set, one can estimate the predictive performance of the algorithm, usually described by classification accuracy and classification risk.

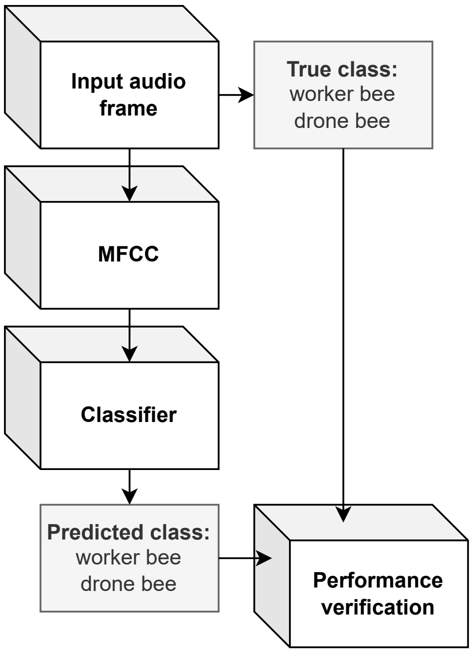

Fig. 5. Diagram of supervised learning.

In Fig. 5, a diagram is presented with particular stages of supervised learning. The first stage concerns the preparation of the data set. In the second stage, we perform signal preprocessing, in that case the calculation of MFCCs, which represent audio signals in the lower-dimensional feature space. The extraction of proper features is often one of the most crucial stages in pattern recognition. And a well-selected method can significantly improve the classification performance, minimizing the risk of misclassification. The next step involved the use of a chosen classifier. In the last step, we can verify the performance of the algorithm thanks to the comparison of true and predicted classes of signals.

Consider a sequence of independent random variables

$$(\underline{X}, Y) = \{(\underline{X}_1, Y_1), \ldots, (\underline{X}_p, Y_p)\}$$

with values in $\mathbb{R}^n \times \{0,1\}$ of a common distribution. In our case, the vectors $\underline{X}_j \in \mathbb{R}^n$ are n-dimensional vectors of mel-frequency cepstral coefficients (MFCCs), and Y_j are class labels, $Y_j \in \{0,1\}$. Classification performance can be presented by classification accuracy, which is equivalent to a percent of correctly classified signals from both classes: worker bees and drones. The complementary performance indicator, the misclassification risk, is defined as the following probability

$$R = \mathbb{P}\left(\hat{Y} \neq Y\right), \tag{3}$$

where Y is a sequence of true classes for test signals, and $\hat{Y}$ is a sequence of predicted classes for those test signals by a classifier.

III. Experiment

A. Honey bee data set

We performed a series of numerical experiments on our data set, available online [13], consisting of 10,000 audio samples recorded for flying worker bees and 1700 audio samples recorded for flying drones. The share of drones, male bees, in the honey bee swarm is around 15% of all bees during late spring and early summer. The remaining 85% are more numerous worker bees, all females of smaller size compared to drones. In a swarm, there is usually one queen, who's main task is laying eggs. The imbalanced data set mimics the naturally occurring shares of workers and drones. The audio samples have 1 second length. The bee sounds were recorded in close proximity to an entrance to a beehive by a directional microphone. All samples were tagged with one of the two class labels: worker bee or drone bee.

All audio samples were pre-processed before the classification stage to extract vectors of lower dimension consisting of n MFCCs obtained for a fixed number n of base functions of the discrete cosine transform (DCT).

B. Classification results

The data set was divided into training and testing sets in proportion 70% to 30%. The calculated MFCCs, representing sound samples, took part in supervised learning and testing of the following classification algorithms:

- support vector machines (SVM) [14],
- linear discriminant analysis (LDA) [15], [16],
- random forest (RF) [17], [18],
- k-nearest neighbors (KNN) [20].

The support vector machine classifier aims to determine a hyperplane separating, with a maximum margin, coefficient vectors representing signals from two classes. Linear discriminant analysis is another classification method that calculates, similarly to SVM, a linear hyperplane by minimizing a discriminant function. The LDA assumes that probability density functions in classes are n-dimensional normal distributions with equal covariance matrices in classes.

The random forest is created by a set of decision trees whose majority vote points to the predicted class. At each division of a tree, a different subset of features is selected. In our experiments, we always randomly chose 4 features from n. Depending on a case, the feature set (the set of MFCCs) counts

$n = 5, 10, 15, 20$ or 40. There is also another popular approach with random choice of around $\sqrt{n}$ features for each division in a tree, but we consciously decided to compare exactly the same version of random forest classifier in all the cases, hence a constant number of 4 random features.

The k-nearest neighbors (KNN) classifier assigns a class label that occurs more times among an odd number k of nearest neighbors.

The classification was performed for 5, 10, 15, 20 and 40 MFCCs. The results are presented in Table I.

TABLE I
CLASSIFICATION ACCURACY [%].

No. of MFCC	SVM	LDA	RF	KNN
5	92.62	92.70	99.88	99.68
10	94.21	93.76	99.82	99.68
15	94.55	94.33	99.85	99.68
20	94.92	94.67	99.08	99.68
40	95.32	95.44	93.33	99.68

The empirical misclassification risk obtained for the tested classifier: linear support vector machines (SVM), linear discriminant analysis (LDA), random forest (RF) of 10 trees with 4 features, and k-nearest neighbors (KNN) with $k = 105$ is shown in Fig. 6.

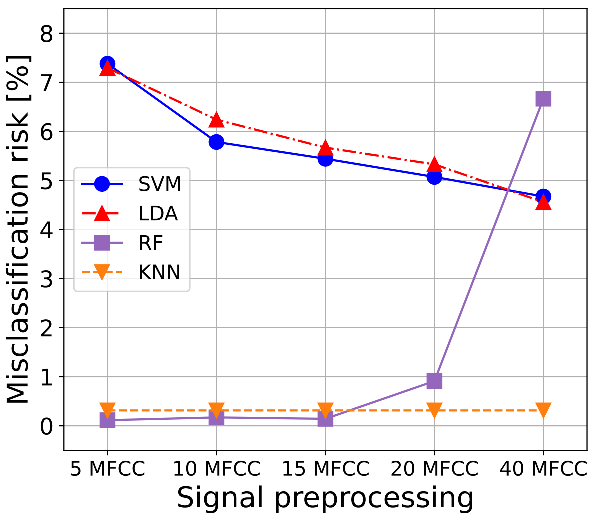

Fig. 6. Misclassification risk for classifiers trained on mel-frequency cepstral coefficients.

The experiment results for the KNN classifier have a constant accuracy value equal to 99.68%. The very small misclassification risk $R_{kNN} = 0.32\%$ for KNN is clearly connected to the optimally minimal value of Bayesian risk R^*, stated by the theorem given by Devroye, Györfi and Lugosi in [21]:

Theorem 1: For all odd k and all distributions,

$$R_{kNN} \leq R^* \left(1 + \frac{1}{\sqrt{ke}}\right). \tag{4}$$

Asymptotically, while the number k of nearest neighbors goes to infinity, we have:

$$lim_{k\to\infty} R_{kNN} \leq R^*. \tag{5}$$

Due to the fact that the Bayesian risk R^* (also called the Bayes probability of error [19]) is optimal, it means that R^* is the lowest possible value of misclassification risk among all classifier, including KNN, what can be denoted as:

$$R^* \leq R_{kNN}. \tag{6}$$

The direct conclusion of the theorem is that while $k \to \infty$, the risk for k-nearest neighbors classifier is asymptotically equal to the Bayesian risk R^*, which for the full probabilistic information about distributions in classes is minimal for all existing classifiers. Since the theoretical distributions are unknown for real-life problems, the Bayesian risk R^* gives information about separability of the classes. The close to zero risk values for KNN, for a large number $k = 105$ of neighbors, can be interpreted as very close to the theoretical optimal Bayesian risk R^*, and show that, indeed, the representations of signals by MFCCs discriminate the worker bee and drone classes.

The high performance of the random forest (RF) and k-nearest neighbors (KNN) classifiers can be problematic in real-life applications, due to the need for extensive calculations for the large data set. The linear classification algorithms: support vector machines (SVM) and linear discriminant analysis (LDA) performed very similar for different numbers of MFCCs. The linear classifiers have an advantage in practice, as the already trained classifier can be explicitly applied without additional access to the data set. The trained classifier (SVM and LDA) defines the decision areas divided into the two classes by a linear hyperplane, calculated during the training stage.

IV. CONCLUSION

All sets of MFCCs were calculated for the same frequency band of 44.1 kHz. In each case, the frequency band was divided into 40 subbands and transformed by a cosine transform, which led to low-dimensional representations in the form of 5, 10, 15, 20 and 40 MFCCs. Each cosine base function corresponds to the particular coefficient on the mel frequency scale. The results of classification of honey bee sounds show clearly that the number of MFCCs has an impact on the performance of the classifiers.

From a practical point of view, we would recommend the application of one of the linear classifiers: support vector machines (SVM) or linear discriminant analysis (LDA). The increase of the misclassification risk from higher-dimensional representation for 40 MFCCs to lower-dimensional for 10 MFCCs is equal around 1.11% for SVM and 1.68% for LDA. It means the drop in the classification performance is very low, and in real-life application, it would be beneficial to base the classification system on only 10 MFCCs, without a significant change in the classifier performance. It is worth mentioning that the LDA also had the shortest computation time.

Future work should focus on the creation of a smartphone application that allows many beekeepers in a simple way to have access and the possibility to monitor their beehives in the presented way. The proposed solution should be incorporated into an IoT system to monitor the condition of a honey bee colony. The recording of audio signals and their processing would be performed on a device located near the beehive. At times defined by a beekeeper, it would report expected events using the cellular phone network. Moreover, the beekeeper would be able to 'call' the system at any time to get an update on the hive status on demand. The proposed solution could be extended by daily reports with additional data such as weather conditions, temperature in a hive, etc.

REFERENCES

[1] S. Kontogiannis, "An Internet of Things-Based Low-Power Integrated Beekeeping Safety and Conditions Monitoring System", *Inventions* 4(3):52 (2019). [Online]. Available: https://doi.org/10.3390/inventions4030052

[2] G. Voudiotis, S. Kontogiannis and C. Pikridas, "Proposed Smart Monitoring System for the Detection of Bee Swarming", *Inventions* 6(4):87 (2021). [Online]. Available: https://doi.org/10.3390/inventions6040087

[3] E. Ntawuzumunsi, S. Kumaran and L. Sibomana, "Self-Powered Smart Beehive Monitoring and Control System (SBMaCS)", *Sensors (Basel)* 21, 3522 (2021). [Online]. Available: https://doi.org/10.3390/s21103522

[4] S.M. Williams, S. Bariselli, C. Palego, R. Holland nad P. Cross, "A comparison of machine-learning assisted optical and thermal camera systems for beehive activity counting", *Smart Agricultural Technology* 2, 100038 (2022). [Online]. Available: https://doi.org/10.1016/j.atech.2022.100038

[5] U. Libal and P. Biernacki, "Detecting drones at an entrance to a beehive based on audio signals and autoencoder neural networks", in Proc. IEEE Signal Processing Symposium (SPSympo), Karpacz, Poland, 26-28 September 2023; pp. 99–104. [Online]. Available: https://doi.org/10.23919/SPSympo57300.2023.10302687

[6] U. Libal and P. Biernacki, "MFCC Selection by LASSO for Honey Bee Classification", *Appl. Sci.* 14(2):913 (2024). [Online]. Available: https://doi.org/10.3390/app14020913

[7] A.S. Hamza, R. Tashakkori, B. Underwood, W. O'Brien, C. Campell, "BeeLive: The IoT platform of Beemon monitoring and alerting system for beehives", *Smart Agricultural Technology* 6, 100331 (2023). [Online]. Available: https://doi.org/10.1016/j.atech.2023.100331

[8] H. Hadjur, D. Ammar and L. Lefèvre, "Toward an intelligent and efficient beehive: a survey of precision beekeeping systems and services", *Comput. Electron. Agric.* 192, 106604 (2022). [Online]. Available: https://doi.org/10.1016%2Fj.compag.2021.106604

[9] P. Nunes-Silva, M. Hrncir, J.T.F. Guimarães, et al. "Applications of RFID technology on the study of bees", *Insect. Soc.* 66, 15–24 (2019). [Online]. Available: https://doi.org/10.1007/s00040-018-0660-5

[10] Abdul, Z. K., Al-Talabani, A. K. Mel Frequency Cepstral Coefficient and its Applications: A Review", *IEEE Access* 10: 122136–122158 (2022). [Online]. Available: https://doi.org/10.1109/ACCESS.2022.3223444

[11] B.S. Soares, J.S. Luz, V.F. de Macêdo, R.R.V. e Silva, F.H.D. de Araújo, D.M.V. Magalhães, "MFCC-based descriptor for bee queen presence detection", *Expert Syst. Appl.* 201, 117104 (2022). [Online]. Available: https://doi.org/10.1016/j.eswa.2022.117104

[12] A. Terenzi, N. Ortolani, I. Nolasco, E. Benetos, S. Cecchi, "Comparison of Feature Extraction Methods for Sound-Based Classification of Honey Bee Activity", *IEEE/ACM Trans. Audio Speech Lang. Process.* 30: 112–122 (2022). [Online]. Available: https://doi.org/10.1109/TASLP.2021.3133194

[13] P. Biernacki, "Dataset for Honey Bee Audio Detection" [Dataset], Zenodo. 2023. [Online]. Available: https://zenodo.org/doi/10.5281/zenodo.10359685 (accessed on 13 December 2023).

[14] C. Cortes and V. Vapnik, "Support-vector networks", *Mach. Learn.* 20(3): 273–297 (1995). [Online]. Available: https://doi.org/10.1007/BF00994018

[15] R.O. Duda, P.E. Hart and D.H. Stork, "Pattern Classification", Wiley Interscience (2000).

[16] C.M. Bishop, "Pattern Recognition and Machine Learning", Information Science and Statistics, Springer-Verlag New York (2006).

[17] T.K. Ho, "Random Decision Forests", in Proc. the 3rd Int. Conf. on Document Analysis and Recognition (ICDAR), Montreal, QC, 14-16 August 1995, pp. 278–282. [Online]. Available: https://doi.org/10.1109/ICDAR.1995.598994

[18] T.K. Ho, "The Random Subspace Method for Constructing Decision Forests", *IEEE Trans. Pattern Anal. Mach. Intell.* 20(8): 832–844 (1998). [Online]. Available: https://doi.org/10.1109/34.709601

[19] L. Devroye, "On the asymptotic probability of error in nonparametric discrimination", *Ann. Stat.* 9: 1320–1327 (1981). [Online]. Available: https://doi.org/10.1214/aos/1176345648

[20] T. Cover and P. Hart, "Nearest neighbor pattern classification", *IEEE Trans. Inf. Theory* 13: 21–27 (1967). [Online]. Available: https://doi.org/10.1109/TIT.1967.1053964

[21] L. Devroye, L. Györfi and G. Lugosi, "A Probabilistic Theory of Pattern Recognition", Stochastic Modelling and Applied Probability, Springer-Verlag, New York (1996).