Extraction of High Level Visual Features for the Automatic Recognition of UTIs

Paolo Andreini[1], Simone Bonechi[1(✉)], Monica Bianchini[1], Andrea Baghini[1], Giovanni Bianchi[1], Francesco Guerri[1], Angelo Galano[2], Alessandro Mecocci[1], and Guendalina Vaggelli[2]

[1] Department of Information Engineering and Mathematics, University of Siena, Via Roma 56, Siena, Italy
simo_bone@alice.it
[2] Department of Medical Biotechnologies, University of Siena, Strada delle Scotte 4, Siena, Italy
http://www.diism.unisi.it,
http://www.dbm.unisi.it

Abstract. Urinary Tract Infections (UTIs) are a severe public health problem, accounting for more than eight million visits to health care providers each year. High recurrence rates and increasing antimicrobial resistance among uropathogens threaten to greatly increase the economic burden of these infections. Normally, UTIs are diagnosed by traditional methods, based on cultivation of bacteria on Petri dishes, followed by a visual evaluation by human experts. The need of achieving faster and more accurate results, in order to set a targeted and sudden therapy, motivates the design of an automatic solution in place of the standard procedure. In this paper, we propose an algorithm that combines a "bag–of–words" approach with machine learning techniques to recognize infected plates and provide the automatic classification of the bacterial species. Preliminary experimental results are promising and motivate the introduction of a visual word dictionary with respect to using low level visual features.

Keywords: Color image processing · Clustering techniques · Bag–of–words · Artificial neural networks · Support vector machines · Urinoculture screening

1 Introduction

Urinoculture represents a screening test in the case of hospitalized patients and pregnant women. For women, the lifetime risk of having a UTI is greater than 50%. Pregnant women seem no more prone to UTIs than other women. However, when the UTI occurs during pregnancy, it is more likely that the infection extends to the kidneys, giving rise to more serious pathologies. For this reason, health care providers routinely screen pregnant women for UTIs during the first 3 months of pregnancy. On the other hand, nosocomial urinary tract infections

A. Petrosino et al. (Eds.): WILF 2016, LNAI 10147, pp. 249–259, 2017.
DOI: 10.1007/978-3-319-52962-2_22

account for up to 40% of all hospital–acquired infections and, most importantly, nosocomial pathogens causing UTIs tend to have a higher antibiotic resistance than simple UTIs [1].

From the operational point on view, for the urinoculture test, the urine sample is seeded on a Petri plate that holds a culture substrate, used to artificially recreate the environment required for the bacterial growth. There exist many different culture media which allow to perform different kinds of analysis, from isolating specific types of bacteria, to promoting a wide range of microbial growth. The seeding procedure (streaking) consists on spreading the urine sample over the whole plate and can be performed both manually or automatically, with an *ad hoc* device. Then, the plate is incubated in a controlled environment for a fixed period of time (16–24 h). After the incubation phase, each plate is visually examined by a microbiologist with the aim of recognizing the possible presence of bacterial colonies and eventually their species and number, adding some more time to the medical report emission. This common situation significantly departs from the requirement to have results in quick time, to set a targeted therapy, avoiding the use of broad–spectrum antibiotics and improving the patient management.

In recent years, significant improvements in biology and medicine applications and decision support systems [2] have been obtained by using hybrid approaches, based on the combination of advanced image processing techniques [3,4], and artificial intelligence methods [5–8]. In fact, the development of automated tools for results assessment (screening systems) has attracted increasing research interest during the last decade, due to their higher repeatability, accuracy, reduced staff time, and lower costs [9]. In particular, in [10–14], different methods have been proposed to automate the uriculture screening, based on image processing and machine learning techniques. Actually, in [10], after segmentation and background subtraction, the classification of the infection type is performed using support vector machines (SVMs) and multilayer perceptron (MLPs), trained with low level visual features, such as the Cie–Lab color components, and the average colony dimension.

In this paper, we propose an algorithm that combines a "bag–of–words" approach with machine learning techniques to recognize infected plates and provide the automatic classification of the bacterial species. A dataset of 753 images has been collected in partnership with the Microbiology and Virology Laboratory of the Careggi Hospital (Florence). The images represent Petri plates that have been automatically seeded on a chromogenic substrate (Chromagar Orientation) by the Copan WASPLab specimen processor. From the dataset, a visual word dictionary based on shape and color features has been extracted. Preliminary experimental results are promising and motivate the introduction of "visual words" with respect to using low level features (such as color and texture).

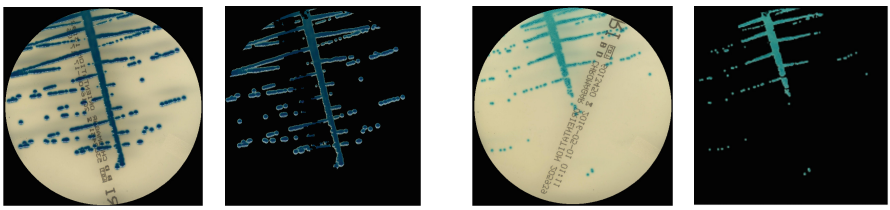
The paper is organized as follows. In the next section, we briefly describe how the image dataset has been collected, before defining the codebook generation procedure. In Sects. 3 and 4, the automatic infected plate recognition and infection classification methods are, respectively, presented, also reporting experimental results. Finally, Sect. 5 collects some conclusions.

2 Dataset Collection and Codebook Generation

The Microbiology and Virology Laboratory of the Careggi Hospital in Florence is the Tuscany reference center for microbiological and virological tests, with more than 250 urinoculture tests performed per day. In the MV–Lab, urine samples are automatically seeded on a chromogenic substrate (Chromagar Orientation) by the Copan WASPLab specimen processor. A chromogenic culture ground exploits the presence of specific enzymes, which are common in bacterial cells, to produce different colors, depending on the bacterial species; in particular, Chromagar Orientation allows to distinguish between the following pathogens:

- Escherichia Coli – produces dark pink to reddish colonies;
- Enterococcus – produces turquoise blue colonies;
- Proteus – produces a brown halo;
- Klebsiella, Enterobacater, Serratia, Citrobacter (KESC) – produces metallic blue colonies;
- Staphylococcus Aureus – produces golden opaque small colonies;
- Staphylococcus Saprophitycus – produces pink opaque small colonies;
- Candida Albicans – produces colorless colonies;
- Streptococcus Agalactiae – produces light blue colonies;
- Pseudomanas Aeuroginos – produces translucent, cream to blue colonies.

A dataset of 753 images of Petri dishes has been collected, gained via the WASPLab, which are first segmented, in order to detect and remove the culture ground.



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

Fig. 1. Results of the background removal procedure: in (a) and (c) the original images; in (b) and (d) the result of the background subtraction.

After the segmentation phase, a bag–of–words approach is then applied to generate a visual codebook, composed by color and shape descriptors. Bag–of–words is a common way to represent documents in natural language processing and information retrieval applications. In this model, a sentence or a document is represented as the collection of its words; only the word frequencies in the text are considered, disregarding grammatical rules and even the word flow. The bag–of–words approach is also widely used for image classification, when visual features are treated as visual words. This involves three main steps;

firstly, visual features are extracted from the image. Then descriptors – arrays collecting such features – are built. After this step, images are represented by a collection of vectors of the same dimension. Finally, descriptors are grouped and converted to codewords, the analogous of words in text documents. Clustering algorithms are usually employed to this aim. The set of codewords, belonging to the same cluster, produces a codebook, i.e. a "word dictionary". For the present study, different kind of features (describing color and shape) have been evaluated, to extract meaningful information from the images of Petri dishes. Based on the background removal procedure described in [10] (see Fig. 1), a set of foreground segments have been extracted from each of the 162 images that compose the training set. Each segment is then represented by a descriptor (feature vector). The selected features are described in the following.

– Color moments characterize the color distribution of a segment, and they are scale and rotation invariant. Among the possible color spaces, experimental results have shown that the use of HSV provides better performance. Three color moments are used, mean (Eq. (1)), standard deviation (Eq. (2)) and skewness (Eq. (3)); they are computed separately for each HSV channel:

$$m_i = \sum_{j=1}^{N} \frac{1}{N} v_{ij} \tag{1}$$

$$\sigma_i = \sqrt{\frac{1}{N} \sum_{j=1}^{N} (v_{ij} - m_i)^2} \tag{2}$$

$$s_i = \sqrt[3]{\frac{1}{N} \sum_{j=1}^{N} (v_{ij} - m_i)^3} \tag{3}$$

where N is the number of pixels in the segment and v_{ij} is the value of the j–*th* pixel in the i–*th* color channel.

– The shape of the segments is also a useful characteristic, since bacterial colonies produced by different types of infections may have different size; moreover, the shape is important to evaluate the infection severity, because it allows to distinguish between single and overlapping colonies. Two shape features have been used, the segment area ($\mathbf{A}$) and the elongation (Eq. (4)):

$$E = \sqrt{\frac{i_2}{i_1}} \tag{4}$$

where i_1 and i_2 are, respectively, the minor and the major axis of the smallest bounding ellipse of the segment.

Therefore, for each image segment in the training set, a descriptor has been computed, in the form:

$$D = [m_H, m_S, m_V, \sigma_H, \sigma_S, \sigma_V, s_H, s_S, s_V, A, E]$$

Once the descriptors were extracted, the codebook generation relies on an unsupervised clustering procedure. The set of codewords obtained in this way forms the final codebook. In particular, the k–means clustering algorithm[1] has been used to extract the codewords (each word corresponds to a cluster centroid). Let us first introduce some notations.

Let X be the dataset to be clusterized and let $x_i \in X$, for $i \in \{1, ..., N\}$, with $N =\mid X \mid$. Let $C = \{c_1, c_2, ..., c_K\}$ be the obtained K disjoint sets, with $\bar{c}_k = \frac{1}{|c_k|} \sum_{x_i \in c_k} x_i$, the centroid of the k–th cluster and $\bar{X} = \frac{1}{N} \sum_{x_i \in X} x_i$ the dataset centroid. The optimal number of clusters K has been selected by evaluating the best partition among data, according to the four different measures, described in the following.

– Silhouette (SL) [15]

$$SL(C) = \frac{1}{N} \sum_{i=1}^{N} s(x_i) \qquad s(x_i) = \frac{(b(x_i) - a(x_i, c_j))}{\max\{b(x_i), a(x_i, c_j)\}} \tag{5}$$

where, $\forall x_i \in X$, $a(x_i, c_j)$ is the average distance of the object x_i from all the points belonging to a different cluster c_j and $b(x_i)$ is the minimum average distance of x_i from all the points belonging to the other clusters.

– Calinski–Harabasz (CH) [16]

$$CH(C) = \frac{(N - \mid C \mid) \sum_{c_k \in C} \mid c_k \mid d(\bar{c}_k, \bar{X})}{(\mid C \mid -1) \sum_{c_k \in C} \sum_{x_i \in c_k} d(x_i, \bar{c}_k)} \tag{6}$$

where $d(x_i, c_j)$ is the average distance of the object x_i from all the points belonging to a different cluster c_j.

– Davies–Bouldin (DB) [17]

$$DB(C) = \frac{1}{\mid C \mid} \sum_{c_k \in C} \max_{c_l \in C \setminus c_k} \left\{ \frac{S(c_k) + S(c_l)}{d(\bar{c}_k, \bar{c}_l))} \right\} \qquad S(c_k) = \frac{1}{\mid c_k \mid} \sum_{x_i \in c_k} d(x_i, \bar{c}_k) \tag{7}$$

– Gap (G) [18]

$$G(C) = E_N^*\{\log(W_k)\} - \log(W_k) \qquad W_k = \sum_{i=1}^{k} \frac{1}{2 \mid c_i \mid} D_i \tag{8}$$

where D_i is the sum of the pairwise distance for all the points belonging to cluster c_i and the expected value $E_N^*\{\log(W_k)\}$ is determined by Monte Carlo sampling from a reference distribution.

The optimal number of clusters, computed in the range $[1, 100]$, is reported in Table 1.

[1] Other clustering methods—such as DBSCAN, OPTICS, SOM—have been tested. k–means was chosen since it offers the best trade off between simplicity and performance.

Table 1. The best value of K selected by the four different validity measures, SL, CH, DB, and G.

	SL	CH	DB	G
K^*	3	3	3	96

We have also tested different values for K within the same range (selecting $K \in \{16, 28, 45, 68\}$), in order to check the ability of the chosen metrics to produce a correct dictionary dimension. The overall codebook generation procedure is summarized in Fig. 2, whereas, in Fig. 3, an example of the obtained set of words with respect to a test image is reported.

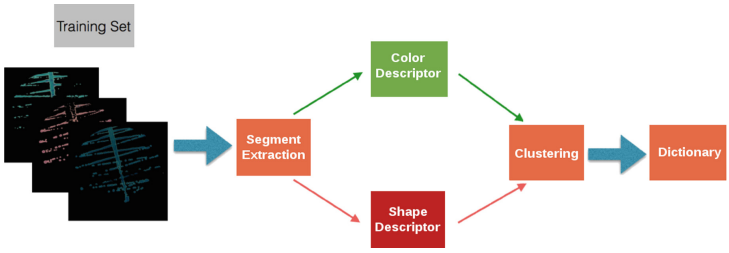
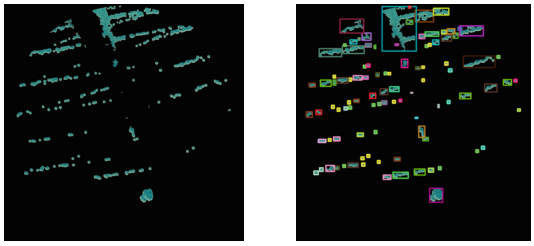


Fig. 2. Codebook generation schema.



(a) (b)

Fig. 3. In (a), the original image without the background and, in (b), words found within the image.

3 Infected Plate Recognition

Automatic recognition of infected plates aims at distinguishing between positive and negative tests, where a positive plate is characterized by the presence of bacterial colonies, whereas on negative plates no bacteria are grown. Nevertheless, sometimes, plates can be considered as negative even if bacterial colonies are

actually present on the culture medium, especially in the case of atypical infections that, in specific circumstances, can be considered as a contamination of the urine sample. The rate between infected and not infected samples is imbalanced, with a higher prevalence of the latter. Actually, about 70% of the samples are negative[2]. Hence, automatically recognizing infected plates with a high accuracy is fundamental, since it can significantly reduce the biologist workload.

Table 2. Dataset composition (infected plate classification).

Dataset	Positive	Negative	Total
Training set	162	162	324
Test set	84	345	429

In order to detect infected plates, using the codebook generation procedure previously described, six codebooks, with a different number of words (3, 16, 28, 45, 68, and 96, respectively), have been extracted from images belonging to the dataset (Table 2). Then, two different classifiers, namely, MLPs and SVMs, have been trained[3], obtaining the results summarized in Table 3.

Table 3. Accuracy gained by SVMs/MLPs based on the different dictionaries.

K	3	16	28	45	68	96
MLP parameters	3–10–2	16–50–2	28–30–2	45–25–2	68–70–2	96–200–2
MLP accuracy	94.63%	94.41%	96.73%	96.2%	96%	95.1%
MLP TP Rate	0.92	0.79	0.91	0.88	0.8	0.82
SVM parameters (Kernel/C/Gamma)	Poly2 1/0.3	RBF 1/0.1	Poly2 1/0.2	Poly2 1/0.3	RBF 1/0.1	RBF 15/0.4
SVM accuracy	95.57%	93.94%	94.87%	95.33%	95.8%	96.5%
SVM TP rate	0.92	1	0.88	0.88	0.91	0.9

As we can observe from the results in Table 3, the performance of the two classification models are very similar, with the best accuracy obtained by MLPs using the codebook with 28 words. However, it is worth noting that, in medical applications, not only the accuracy of the system is important, but also its capacity of avoiding false negatives since, actually, a false negative could lead to ignore the infection and to expose the patient to possible risks. Therefore,

[2] For this reason, the training set dimension has been reduced to balance the number of positive and negative patterns (see Table 2).

[3] The MLP structures are described in Table 3. Both hidden and output neurons are sigmoidal. Two neurons constitute the ouput layer in order to improve the network *flexibility* in modeling complicated relationships. All the architectural parameters (for MLPs and SVMs) were chosen via a trial–and–error procedure and crossvalidation.

Table 4. Accuracy (a) and confusion matrix (b) obtained by an RBF kernel SVM, with $C = 1$ and $\gamma = 0.1$, on the test set composed by 429 images.

(a) Accuracy

	Number of Images	**Percentage**
Incorrectly Classified	26	6.06 %
Correctly Classified	**403**	**93.94 %**

(b) Confusion Matrix

Negative	**Positive**
319	26
0	84

observing the True Positive (TP) rate, the SVM trained using the codebook with 16 words must be preferred with respect to other alternatives. Table 4 collects detailed results obtained in this case.

4 Infection Species Recognition

Automatic infection classification aims at recognizing the infection strain(s) present on the infected plates. There is a huge number of different bacterial strains, which can possibly be present in an urine sample and, as discussed in Sect. 2, using the Chromagar Orientation medium, nine different bacterial types can be distinguished. The most common infection is E. Coli, with about 60–70% of occurrence, whereas all other species are much less frequent. As a consequence, our image dataset is too small to represent, with statistical significance, all the nine classes. Actually, for example, only four images of Pseudomanas Aeuroginos and Proteus, and two of Staphylococcus Saprophitycus are present. Therefore, the number of classes to be recognized has been reduced, grouping together some underrepresented classes with similar (color) properties (see Table 5).

Table 5. Dataset composition (infection classification).

Infection classes	Training set	Test set	Total
E. Coli	94	48	142
Enterococcus Spp	26	14	40
KESC	22	12	34
Other (Proteus, S. Aureus, Pseudomonas, Candida)	20	10	30

As it can be observed in Table 5, the dataset is imbalanced; as expected, E. Coli is highly prevalent, whereas the other classes contain a small number of samples. From the classification point of view, this is an undesirable situation and, to address this problem, the training set has been pre–processed and artificially balanced[4]. Then, the codebook has been generated using the positive

[4] The Weka Class Balancer function has been used to balance the data. This function reweights the instances in the data so that each class has the same total weight. The total sum of weights across all instances will be maintained. Only the weights in the first batch of data received by this filter are changed.

Table 6. Accuracy gained by SVMs/MLPs based on the different dictionaries.

K	3	16	28	45	68	96
MLP parameters	3–10–4	16–25–4	28–20–4	45–30–4	68–60–4	96–100–4
MLP accuracy	23.8%	88%	88%	90.47%	92.85%	85.71%
SVM parameters (Kernel/C/Gamma)	RBF 10/0.2	Poly2 1/0.6	Poly2 1/0.1	Poly2 112/0.1	Poly2 1/0.3	RBF 1/0.2
SVM accuracy	54.7%	73.8%	80.95%	77.38%	82.15%	78.57%

sample images. Actually, six codebooks with a different number of words (3, 16, 28, 45, 68, and 96, respectively) have been extracted from the images of the dataset (Table 5). Two different classification architectures, MLPs and SVMs, have been trained, producing the results reported in Table 6.

In this case, the best accuracy has been obtained using the codebook with 68 words. It is worth noting that, differently from the infected plate detection, a greater number of words seems to be necessary to conveniently characterize the infection type. This may reflect the more complex nature of this problem (see Table 7).

Table 7. System overall accuracy and confusion matrix obtained by an SVM with a polynomial kernel of degree 3, $C = 1$, $\gamma = 0.1$, and by an MLP architecture, with 68–60–4 units.

(a) Accuracy

	Number of Images	**Percentage**
Total Number of Images	84	
Incorrectly Classified	15 / 6	17.85 / 7.14 %
CorrectlyClassified	**59 / 78**	**82.15 / 92.85 %**

(b) Confusion Matrix

E.Coli	**Enterococcus**	**KESC**	**Other**
45 / 47	0 / 0	1 / 0	2 / 1
1 / 1	10 / 13	0 / 0	3 / 0
0 / 0	1 / 1	10 / 11	1 / 0
5 / 0	1 / 3	0 / 0	4 / 7

The accuracy for each single class gained here is very similar to the results obtained in [10] even if, unfortunately, a precise comparison is not possible due to the differences in the training dataset and in the classification approach (different number of classes in the two problems).

Finally, briefly considering the computational complexity of the proposed approach, we can notice that the image segmentation module (which is out of the scope of this article) represents the most demanding task, taking 6 to 10 s for each image, whereas the whole procedure constituted by the two phases of word

frequency histogram extraction and image classification took 3 s, on average, for each image[5]. All the experiment were carried out using an Intel i5 CPU.

5 Conclusions

Urinary tract infections can be caused by many different microbes, including fungi, viruses, and bacteria. Bacteria are actually the most common cause of UTIs. The body is usually able to rapidly removing bacteria that enter the urinary tract before they cause symptoms. However, sometimes bacteria overcome the body natural defenses and, in fact, roughly 150 million of infections occur annually worldwide. In this paper, an automatic method capable of detecting the presence of UTIs and to establish their type, was described. The system shows a good accuracy in distinguishing positive and negative samples, and also a very good sensitivity. Moreover, the proposed procedure is able to recognize different infection types with a high accuracy. Unfortunately, some classes are underrepresented, and this lead to group different infection types together (based on similar colors) in order to obtain a meaningful representation. A larger dataset (hopefully available soon) could avoid this issue and allow to distinguish among a higher number of different infections, even with very similar colors.

Aknowledgements. The authors would like to thank Prof. Rossolini and the whole staff of the MV–Lab of the Careggi Hospital for their willingness to provide real data, and for their invaluable experience in interpreting their microbiological meaning.

References

1. National Institute of Diabetes and Digestive and Kidney Diseases, Urinary Tract Infections in Adults. https://www.niddk.nih.gov/health-information/health-topics/urologic-disease/urinary-tract-infections-in-adults/Pages/facts.aspx
2. Berlin, A., Sorani, M., Sim, I.: A taxonomic description of computer-based clinical decision support systems. J. Biomed. Inform. **39**, 656–667 (2006). Elsevier
3. Deserno, T.M.: Biomedical Image Processing. Springer-Verlag, New York (2011)
4. Belazzi, R., Diomidous, M., Sarkar, I.N., Takabayashi, K., Ziegler, A., McCray, A.T., Sim, I.: Data analysis, data mining: current issues in biomedical informatics. Methods Inf. Med. **50**(6), 536–544 (2011). Schattauer Publishers
5. Agah, A.: Artificial Intelligence in Healthcare. CRC Press, Boca Raton (2014)
6. Heckerling, P.S., Canaris, G.J., Flach, S.D., Tape, T.G., Wigton, R.S., Gerber, B.S.: Predictors of urinary tract infection based on artificial neural networks and genetic algorithms. Int. J. Med. Inform. **76**(4), 289–296 (2007)
7. Bianchini, M., Maggini, M., Jain, L.C.: Handbook on Neural Information Processing. Intelligent Systems Reference Library, vol. 49. Springer-Verlag, Heidelberg (2013)
8. Bandinelli, N., Bianchini, M., Scarselli, F.: Learning long-term dependencies using layered graph neural networks. In: Proceedings of IJCNN-WCCI 2012, pp. 1–8 (2012)

[5] Instead, the codebook generation required about 15 min, using a training set of pre–segmented images.

9. Bourbeau, P.P., Ledeboer, N.A.: Automation in clinical microbiology. J. Clin. Microbiol. **51**(6), 1658–1665 (2013)
10. Andreini, P., Bonechi, S., Bianchini, M., Garzelli, A., Mecocci, A.: Automatic image classification for the urinoculture screaning. Comput. Biol. Med. **70**, 12–22 (2016). Elsevier
11. Andreini, P., Bonechi, S., Bianchini, M., Garzelli, A., Mecocci, A.: ABLE: an automated bacterial load estimator for the urinoculture screening. In: ICPRAM, pp. 573–580. Springer (2016)
12. Andreini, P., Bonechi, S., Bianchini, M., Mecocci, A., Massa, V.: Automatic image analysis and classification for urinary bacteria infection screening. In: Murino, V., Puppo, E. (eds.) ICIAP 2015. LNCS, vol. 9279, pp. 635–646. Springer, Heidelberg (2015). doi:10.1007/978-3-319-23231-7_57
13. Andreini, P., Bonechi, S., Bianchini, M., Mecocci, A., Massa, V.: Automatic image classification for the urinoculture screening. In: Neves-Silva, R., Jain, L.C., Howlett, R.J. (eds.) Intelligent Decision Technologies. SIST, vol. 39, pp. 31–42. Springer, Heidelberg (2015). doi:10.1007/978-3-319-19857-6_4
14. Ferrari, A., Signoroni, A.: Multistage classification for bacterial colonies recognition on solid agar images. In: Proceeding of IEEE IST 2014, pp. 101–106 (2014)
15. Rousseeuw, P.J.: Silhouettes: a graphical aid to the interpretation and the validation of cluster analysis. J. Comput. Appl. Math. **20**, 53–65 (1987)
16. Calinski, T., Harabasz, J.: A dendrite method for cluster analysis. Commun. Stat. **3**(1), 1–27 (1974)
17. Davies, D.L., Bouldin, D.W.: A cluster separation measure. IEEE Trans. Pattern Anal. Mach. Intell. **1**(2), 224–227 (1979)
18. Tibshirani, R., Walther, G., Hastie, T.: Estimating the number of clusters in a data set via the gap statistic. J. Royal Stat. Soc. Ser. B **63**(2), 411–423 (2001)