ISSN 1870-4069

# Clustering Models for the Identification of Coexisting Bacteria in Groups of Patients with the Polymicrobial Syndrome Bacterial Vaginosis (BV)

Henry Jesús Hernández Gómez, Juana Canul Reich

Juarez Autonomous University of Tabasco,
Academic Division of Information Sciences and Technologies,
Mexico

`henryhernandezgomez@hotmail.com, juana.canul@ujat.mx`

## 1 Motivation

Bacterial vaginosis is the most frequent infection in women of child-bearing age [2]. The pathogenesis is unclear, but involves overgrowth due to multiple bacterial pathogens and a decrease in the usual vaginal flora with a predominance of lactobacillus [7, 5]. Its complications, such as endometritis, salpingitis, oophoritis [4], preterm premature rupture of membranes (PPROM), and chorioamnionitis [6]. Particularly, it is well known that a BV-positive case is a consequence of a disequilibrium state of bacteria; it is not a unique bacterium but a coexistence of bacteria leading to a BV-positive condition. However, these bacteria may differ from patient to patient.

Thus, the present research arises from the need to support physicians in understanding the unclear pathogenesis of BV by identifying clusters of patients that presumably share coexisting bacteria in a grouped form, and from there, to analyze the contexts of bacterial coexistence. So it is our interest to tackle this problem using a clustering approach.

The dataset used for this research was generated by the laboratory of research in infectious and metabolic diseases of the Juarez Autonomous University of Tabasco. It was obtained as part of research in molecular epidemiology in BV during the years 2016 to 2018 in an urban population in southeastern Mexico [8]. The dataset consists of 201 patient records with 19 attributes, which was explored through AHC methods without performing any pre-processing process.

This study contributes to the effort to provide information that helps to learn more about the coexisting bacteria in the groups of patients diagnosed as BV-positive. Furthermore, the benefit of identifying the groups translates into the selection of specific treatments according to the bacteria coexisting in each group. It also becomes a support tool to obtain a priori knowledge of the contexts that may occur in clinical cases.

## 2 Previous Works in the Area

Nowadays, there exist few research studies that have conducted an analysis of the BV disease using ML methods.

Song et al. [9] contributed to integrate Superpixel methods with Deep Learning methods based on convolutional neural network (CNN) for the automatic assisted diagnosis of BV. The classification results yield an accuracy of 99%, the sensitivity of 100% and specificity of 98.04%.

Baker et al. [1] built a classification model by breaking down the groups of microbes based on their correlation. Likewise, it reduced the number of factors, increasing the interpretability of the classification models. The classifications were made using Genetic Programming, Random Forest, and Logistic Regression, the precision of the models was evaluated using ROC curves. The accuracy obtained from the models was between 90% and 95% when they were classified using the dataset with the Nugent score.

Cruciani et al. [3] designed a new phylogenetic microarray-based tool (VaginArray) that includes 17 probe sets specific for the most representative bacterial groups of the human vaginal ecosystem. The VaginArray was applied to evaluate the efficacy of rifaximin vaginal tablets for the treatment of BV. The results showed the ability of rifaximin to reduce the growth of various BV-related bacteria (Atopobium vaginae, Prevotella, Megasphaera, Mobiluncus, and Sneathia spp.)

## 3 Research Objectives

To build a bacterial community clustering model of the polymicrobial syndrome bacterial vaginosis (BV) that allows identifying characteristics of similarity and dissimilarity of the clusters, as well as bacterial coexistence contexts with biological significance.

## 4 Methodology

In order to achieve the stated research objective, it was necessary to construct a study scheme that shows the clustering approaches considered and the phases to be performed that will lead to visual clustering models that allow the analysis of bacterial coexistence contexts, as shown in Figure 1.

## 5 State of the Research

Currently, the research shows the progress of more than 75% of the activities carried out. The activities involved in the construction and validation of results through metrics were completed for two of the three investigated approaches.

Furthermore, agglomerative hierarchical clustering and partitional clustering have been constructed. Finally, each model was converted to a data visualization

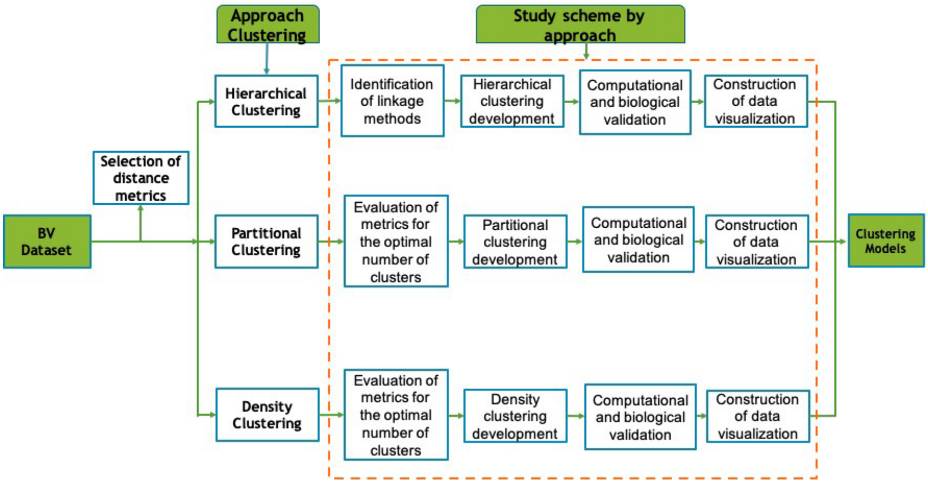

**Fig. 1.** Study scheme.

that allows analysis of the contexts of bacterial coexistence between groups of patients. The density clustering algorithms so far are about to start the process of constructing the data visualizations.

As the research is pioneering the use of clustering methods to create visual models for further analysis of bacterial coexistence contexts, each model is subjected to biological validation by an expert in the field who evaluates the biological significance of the groups.

## 6 Preliminary Results

The overall results that have been achieved so far in BV research through clustering models are:

Two experiments were designed and evaluated in the agglomerative hierarchical clustering (AHC) methods through five linkage methods and a distance measure. The first experiment aimed to identify clusters with the presence and absence of VB. The result obtained was a segmentation of the dataset into two groups, one with elements with positive classes and a second group with elements of negative and indeterminate classes.

The clustering partitioning (PC) methods were evaluated through four methods from the same experiments perspective as the AHCs. The findings of the PCs assessed in the first experiment show a similarity of clustering to the AHCs. However, in the second experiment, the algorithms show a dissimilar clustering with the common VB-positive diagnostic. From the findings across the two clustering approaches evaluated, we can infer that the best performance achieved for the second experiment is from the AHCs.

We have created data visualization tools for further analysis of bacterial coexistence. The identified clusters show a context of bacterial coexistence of two anaerobic bacteria with a prevalence of more than 30%. The findings of the dissimilar clusters with a common diagnosis, there is a coexistence of three bacteria with a prevalence of more than 30%.

We have also experimented with other measures of similarity to improve the results of experiment two of the PC.

## References

1. Beck, D., Foster, J.A.: Machine learning techniques accurately classify microbial communities by bacterial vaginosis characteristics. PloS one 9(2), e87830 (2014)
2. Coleman, J.S., Gaydos, C.A.: Molecular diagnosis of bacterial vaginosis: an update. Journal of clinical microbiology 56(9), e00342–18 (2018)
3. Cruciani, F., Biagi, E., Severgnini, M., Consolandi, C., Calanni, F., Donders, G., Brigidi, P., Vitali, B.: Development of a microarray-based tool to characterize vaginal bacterial fluctuations and application to a novel antibiotic treatment for bacterial vaginosis. Antimicrobial agents and chemotherapy 59(5), 2825–2834 (2015)
4. for Disease Control, C., Prevention, et al.: Sexually transmitted diseases treatment guidelines. Morbid Mortal 42, 57–59 (1993)
5. Muzny, C.A., Taylor, C.M., Swords, W.E., Tamhane, A., Chattopadhyay, D., Cerca, N., Schwebke, J.R.: An updated conceptual model on the pathogenesis of bacterial vaginosis. The Journal of infectious diseases 220(9), 1399–1405 (2019)
6. Prajarto, H.W., Pramono, B.A.: The association of cervical length, bacterial vaginosis, urinary tract infection and premature rupture of membranes to the imminent preterm labour. Diponegoro International Medical Journal (DIMJ) 1(2), 10–6 (2020)
7. Reid, G.: Is bacterial vaginosis a disease? Applied microbiology and biotechnology 102(2), 553–558 (2018)
8. Sanchez Garcia, E.K., Contreras Paredes, A., Martinez Abundis, E., Garcia Chan, D., Lizano, M., de la cruz Hernandez, E.: Molecular epidemiology of bacterial vaginosis and its association with genital microorganisms in asymptomatic women. Journal of medical microbiology 68(9), 1373–1382 (2019)
9. Song, Y., Ni, D., Zeng, Z., He, L., Chen, S., Lei, B., Wang, T.: Automatic vaginal bacteria segmentation and classification based on superpixel and deep learning. Journal of Medical Imaging and Health Informatics 4(5), 781–786 (2014)