



# Machine Learning based on Probabilistic Models Applied to Medical Data: The Case of Prostate Cancer

*Tshikutu Bikengela Anaclet*<sup>1</sup>, *Mutapay Tshimona Remy*<sup>2</sup>, *Kafunda Katalay Pierre*<sup>3</sup>,  
*Ntumba Badibanga Simon*<sup>4</sup>, *Mbuyi Mukendi Eugène*<sup>5</sup>

<sup>1</sup> *Faculté d'Informatique, université Notre-Dame du Kasayi, Kananga, D.R.Congo*

<sup>2</sup> *Department of Computer Science, Institut Supérieur Pédagogique de la Gombe, Kinshasa, D.R.Congo*

<sup>3,4,5</sup> *Faculty of Science and Technology, University of Kinshasa, Kinshasa, D.R.Congo*

*email:*<sup>1</sup> [tshikutuanaclet2013@gmail.com](mailto:tshikutuanaclet2013@gmail.com), <sup>2</sup> [remycimona@gmail.com](mailto:remycimona@gmail.com), <sup>3</sup> [kafundakatalay@gmail.com](mailto:kafundakatalay@gmail.com), <sup>4</sup> [profntumba@gmail.com](mailto:profntumba@gmail.com), <sup>5</sup> [eugenembuyi@gmail.com](mailto:eugenembuyi@gmail.com)

## ARTICLE INFO

### Article history:

Received 29 April 2023

Revised 24 November 2023

Accepted 21 December 2023

Available online 30 December 2023

### Keywords:

Classification,  
Mixture models,  
Gaussian mixture,  
Bayesian networks,  
Bayes classifier

### IEEE style in citing this article:

T. B. Anaclet, M. T. Remy, K. K. Pierre, N. B. Simon, and M. M. Eugène, "Machine Learning based on Probabilistic Models Applied to Medical Data: The Case of Prostate Cancer," *Journal of Innovation Information Technology and Application (JINITA)*, vol. 5, no. 2, pp. 105–113, Dec. 2023.

## ABSTRACT

The growth in the amount of data in companies puts analysts in difficulties when extracting hidden knowledge from data. Several models have emerged that focus on the notion of distances while ignoring the notion of conditional probability density. This research study focuses on segmentation using mixture models and Bayesian networks for medical data mining. As enterprise data becomes large, there is a way to apply data mining methods to make sense of it using classification methods. We designed different models with different architectures and then applied these models to the medical database. The algorithms were implemented for the real data. The objective is to classify individuals according to the conditional probability density of random variables, in addition to identifying causalities between traits from tests of conditional independence and a correlation measure, both based on  $\chi^2$ . After a quick illustration of several models (decision tree, SVM, K-means, Bayes), we applied our method to data from an epidemiological study (done at the University of Kinshasa University clinics) of case-control of prostate cancer. Thus, we found after interpretation of the results followed by discussion that our model allows us to classify a new individual with an accuracy of 96%.

## 1. INTRODUCTION

In recent years, an increasing amount of data has been generated by companies, whether it is medical data, such as patient records seen in a hospital and their pathologies, banking data, such as credit card transactions, or industrial data, such as measurements from sensors on a production line, or any other type of data imaginable. In each case, the analyst seeks to extract hidden information from the data. But no matter how big the data is, it will not allow him to do this. This gave rise to the idea of classifying the data so that the most similar data are grouped in a single class. This technique is called automatic classification [1], [2]. Automatic data classification is the algorithmic categorization of objects. It consists of assigning a class or category to each object (or individual) to be classified, based on statistical data. For a long time, this mathematical technique was based on the notion of distance, i.e. two elements were closer according to the distance separating them, as presented by E Lincker, C Guinaudeau, O Pons, J Dupirein [2] and C. Mélina and L. Benoît in [3]. Then, the problem was to know which metric to choose for a good ranking. Moreover, the choice of the number of classes was very crucial and sometimes depended on one analyst to

another. Another problem was the sensitivity to atypical data, i.e. data that deviated a little from the others. The other problem was how to define the probability of an individual belonging to one class of data and not to another. To do this, a question was raised, namely: *How to classify individuals according to the conditional probability density of the random variables?*

In an attempt to address some of the above problems, probabilistic models have emerged. Thus, our study will focus on the probabilistic approach [4] which has largely corrected the problem posed above. This approach is the use of the notion of conditional probability density as a Gaussian, or a mixture of Gaussians in classification. This time, it is the statistical parameters such as the mean or the variance that allow us to say that two elements are very close according to their proximity to their mean. The above methods will be applied to prostate cancer, which is the most common cancer in men over 50 in industrialized countries. As diagnostic practices and treatment options continue to evolve, small tumors can be detected, and targeted treatments can be guided to minimize the morbidity of therapy [5], [6]. The above-mentioned set of methods will be crucial for the automated processing of medical data to assist and guide the practitioner in the diagnostic decision-making and therapeutic procedure for specific treatments of prostate cancer. This research study focuses on segmentation using mixture models and Bayesian networks for medical data mining. The aim is to classify individuals according to the conditional probability density of random variables, in addition to identifying causalities between traits using conditional independence tests and a correlation measure, both based on the  $\chi^2$ , which gives us a reliability of 96%, considering how advantageous our method, unlike the algorithms developed in [1], [2], [3] which fall short as limiting their classification on elementary statistical parameters such as distance, variance, etc. Our contribution is twofold: on the scientific side, we present a probabilistic approach applied to automatic classification, presenting the virtues it brings. Secondly, a medical aspect: we have created a decision-support tool for the segmentation of Prostate Cancer data.

## 2. METHODOLOGY

In our research, we have opted for algorithms based on the mixture model [7]. Here we describe the motivation and the detailed overview of the tasks and the proposed approach. Today, it remains to be seen that prostate cancer has serious effects on men already beyond the age of 40, even leading to death. Our approach addresses methods that help to understand and extract knowledge from masses of data on prostate cancer [8], which can enable a province, or a country to understand the harms of this disease based on other factors affecting it and their probabilities on the evolution of this disease.

### 2.1. Segmentation by Mixture Model

The mixture model, by definition, is a statistical model used to parametrically estimate the distribution (density function) of random variables by modeling them as a sum of several other simple distributions as presented in [7], [8]. This definition means that it is assumed that the whole population is represented by a probability distribution which is a mixture of  $C$  probability distributions associated with the classes. The main objective is to identify the  $C$  distributions by estimating their parameters. This identification consists of assuming that the observed data are realizations of a random vector  $X$  of unknown distribution  $P$ . The objective is to reconstruct  $P$  from its realizations. The probability density  $f(x, \Theta)$  at a point is unknown. The principle of the model is to decompose this density into a sum of  $K$  components  $f_k(x, \theta_k)$  ( $k=1, \dots, c$ ) corresponding to the  $C$  classes whose parameters will be estimated  $\theta_k$  ( $k=1, \dots, c$ ) from a sample  $X$ . Mathematically, a finite mixture on a space  $X$  is a convex combination of probability laws  $f_k(x, \theta_k)$  (where the  $f_k(x, \theta_k)$  is called mixing components as explained in [7], [8], defined on  $X$  by

$$f(x, \Theta) = \sum_{k=1}^c \pi_k f_k(x, \theta_k) \quad (1)$$

Where  $\pi_k$  is the a priori probability of component  $k$ . It does check the probability conditions such that  $\sum_{k=1}^c \pi_k = 1$ , and  $0 \leq \pi_k \leq 1$ .  $\Theta$  and  $\theta_k$  are respectively the parameters of the model  $f(x, \Theta)$  and  $f_k(x, \theta_k)$

If  $X$  is a continuous distribution, i.e. we have an infinite number of components, then (1) is written

$$f(x, \Theta) = \int_1^\infty \pi_k f_k(x, \theta_k) \quad (2)$$

The probability densities  $f_k(x, \theta_k)$  can be a distribution (parameter) from the large family of statistical distributions such as (Gaussian, Student).

But in the following, we will detail the Gaussian mixture because it has some advantages compared to other distributions. In particular, during practice, the specification of the components of a mixture is not always guided by the singularity of the data to be modeled. Gaussian mixtures, which assume conditional populations distributed according to a normal distribution, are of great interest because of their flexibility, their ability to approximate a wide variety of densities, their mathematically simple use, and the generality of the normal distribution as shown by the central limit theorem.

**2.2. Gaussian Mixture Model**

To describe Gaussian mixing, we first define the density function of the Gaussian distribution. We then introduce the notion of Gaussian mixing [8].

**2.2.1. Gaussian distribution**

The Gaussian distribution, also known as the normal distribution, is the best-known of the probabilistic distributions. It has been widely used to model the distribution of continuous random variables [9]. In the case of a simple random variable X, the density function of the Gaussian distribution can be written as follows:

$$f(x, \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \tag{3}$$

Where  $\mu$  is the mean  $\sigma^2$  and is the variance. If the distribution is multidimensional, the density function of the multivariate Gaussian distribution takes the following form:

$$f(x, \mu_k, \Sigma_k) = \frac{1}{(2\pi)^{d/2} |\Sigma_k|^{1/2}} e^{-\frac{1}{2}(x-\mu_k)^t \Sigma_k^{-1} (x-\mu_k)} \tag{4}$$

with  $\Sigma_k$  the variance-covariance matrix,  $\mu_k$  the matrix of means, and d the dimension of the Space of individuals. Gaussian mixing has long been used in statistical learning. It can model any numerical data set but with arbitrary precision.

The Gaussian mixture model is a convex combination of several Gaussian components. It is particularly used in cases where the data under study cannot be modeled by a simple Gaussian. In other words, if the data structure is naturally composed of several groups, it is necessary to represent them by a Gaussian mixture model rather than a simple Gaussian distribution [9]. To formalize this, simply replace each  $f_k(x, \theta_k)$  with the density function of the Gaussian distribution. The density function of the Gaussian distribution is then written as follows:

$$f(x, \theta) = \sum_{k=1}^C \pi_k \mathcal{N}(x, \mu_k, \Sigma_k) \tag{5}$$

Where  $\mathcal{N}(x, \mu_k, \Sigma_k)$  is the k component of the Gaussian mixture.

In this context, we define the set of weights as  $\pi = \{ \pi_k \}$ , the set of means by  $\mu = \{ \mu_k \}$ , and the k-th class covariance matrices we denote  $\Sigma = \{ \Sigma_k \}$ . We also define  $\Theta = \{ \theta_k \}$  and  $\theta_k = \{ \pi_k, \mu_k, \Sigma_k \}$ .

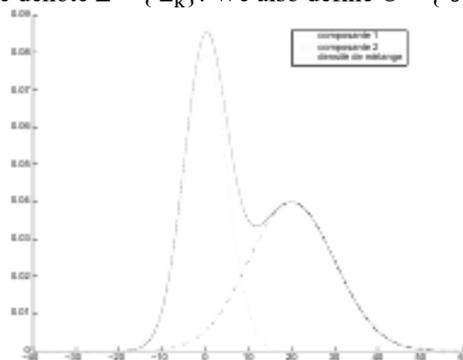


Figure 1: Density mixtures

Here is a representation of the Gaussian components and the continuous curve shows the mixing density of these two Gaussians. After a brief overview of Gaussian mixing, we now want to explain how the parameters of the Gaussian mixing model can be estimated, using the most well-known technique in this field: the EM (Estimation-Maximization) algorithm.

**2.2.2. Estimation of the parameters of mixing models**

Several estimation methods have already been developed in inferential statistics to estimate the parameters of the probability distribution of a sample, including point estimation, confidence interval estimation, etc., but the best known is the maximum likelihood method which is defined as follows:

Let the sample  $X = \{x_1, \dots, x_n\}$  with  $x_1, \dots, x_n$  independent realizations of a random vector  $X$ , the likelihood law of the data for the model of parameter  $\Theta$  is written :

$$L(\Theta) = \prod_{i=1}^n f(x_i, \Theta) \quad (6)$$

$$\Theta = \{ \theta_k \} \text{ and } \theta_k = \{ \pi_k, \mu_k, \Sigma_k \}.$$

It is easier to maximize the log-likelihood instead of the likelihood function itself. If  $\Theta$  maximises  $\ln(L(\Theta))$ , then it also maximises  $L(\Theta)$ . This is due to the monotonicity of the log-likelihood function. The log-likelihood function is written as:

$$\ln L(\Theta) = \sum_{i=1}^n \ln f(x_i, \Theta) \quad (7)$$

This method is centralized by the EM algorithm in the case of Gaussian mixtures:

### **EM algorithm**

Proposed by DA Boiko and AS Kashin in [10], [11] allows structural learning to evaluate the parameters and the score of all the networks in this neighborhood, and to choose the best one for the next iteration to achieve a good ranking.

It is represented as follows:

**Input:** C the number of components in the mixture, X a set of observations,

$\Theta^{(0)} = [ \pi_1^{(0)}, \dots, \pi_c^{(0)}, \mu_1^{(0)}, \dots, \mu_c^{(0)}, \Sigma_1^{(0)}, \dots, \Sigma_c^{(0)} ]$  an initial value of the model parameters.

**Output:** A (local) maximum  $\Theta^*$  of the likelihood law

1.  $S \leftarrow 0$  ;
2. Initialization of the model  $\Theta^{(0)}$ ;
3. Do
4. (Step -E: Estimation) A posteriori probability calculation :
5. For  $i= 1 \dots N$
6. For  $k= 1 \dots C$
- 7.
8. Next
9. Next
10. (Stage - M: Maximisation)
11. For  $k= 1 \dots C$
12. Estimation of the a priori probability  $\pi_k$  of the k-th class :

$$\pi_k^{(s+1)} = \frac{1}{N} \sum_{i=1}^N \gamma(z_{ik})$$

Estimation of the average  $\mu_k$  of class k

$$\mu_k^{(s+1)} = \frac{\sum_{i=1}^N \gamma(z_{ik}) \cdot x_i}{\sum_{i=1}^N \gamma(z_{ik})}$$

Estimation of the  $\Sigma_k$  covariance matrix of class k

$$\Sigma_k^{(s+1)} = \frac{\sum_{i=1}^N \gamma(z_{ik}) \cdot (x_i - \mu_k^{(s+1)}) (x_i - \mu_k^{(s+1)})^t}{\sum_{i=1}^N \gamma(z_{ik})}$$

13. Next
14.  $S \leftarrow s+1$
15. Up to convergence

The convergence of the algorithm is achieved in one of two cases: either the difference in the likelihood function between two consecutive steps is negligible, or the new estimated parameters do not change for the previous step [10], [12]. Each individual is assigned to the class to which he or she is most

likely to belong and this is determined by  $\gamma(z_{ik})$ . The EM algorithm surely converges to a local optimum likelihood. This convergence depends directly on the initialization phase of the algorithm. A bad initialization could lead to a bad model, while a good initialization will rather lead to a convergence towards a global likelihood optimum. To overcome the problems associated with the EM algorithm, a few variants have been proposed, including the SEM, and CEM algorithms [8], [11], [12], which will be developed in our next articles.

### 3. RESULTS AND INTERPRETATIONS

In our application the objective is to classify people living with prostate cancer and to determine the level of intervention and follow-up, the correlation between the different variables that are taken into account; as well as to predict the class to be assigned to a new individual. Based on the DataSet we received, we ran several machine learning algorithms developed in Python to arrive at the interpretation of the results.

#### 3.1. Data Source

The research sample was drawn from the database of the University Clinics of the University of Kinshasa. The data collection period was from January 2017 to December 2021. The first step was to collect data by selecting all patients diagnosed with prostate disease. To determine the classification of disease types, compiled based on a system of categories and grouped into a disease according to predetermined criteria known as the International Statistical Classification of Diseases and Related Health Problems, tenth revision [13], [14].

##### 3.1.1. The correlation matrix

The correlation matrix is represented as follows:

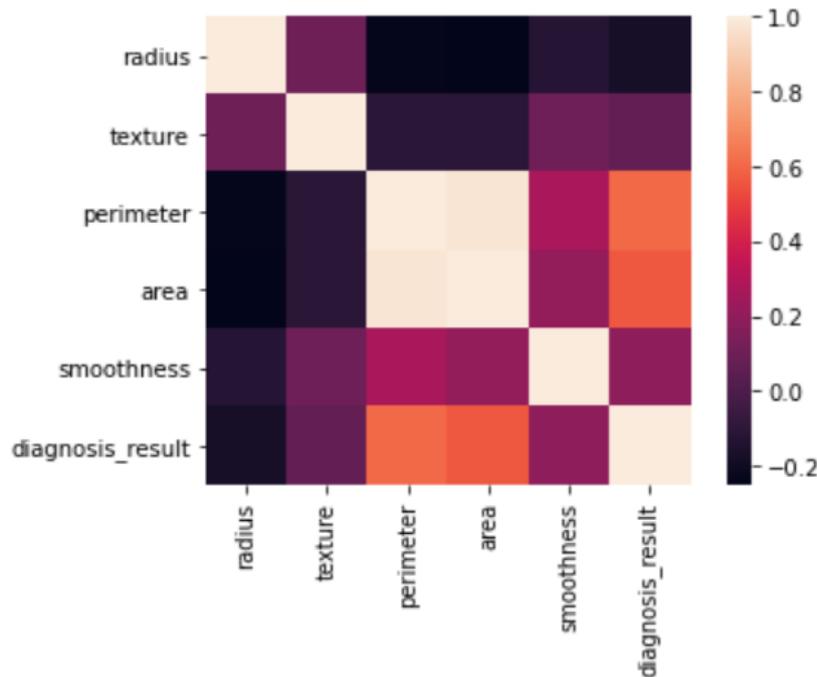


Figure 2: Correlation matrix

The main objective of this matrix is to reduce some irrelevant variables, effectively those that are more correlated with our class of interest which is the diagnosis\_result, we notice here all the variables have a low correlation with the class, hence the interest to use them in the following.

##### 3.1.2. Visualisation of training and test data

The visualization of training and test data can be represented as follows:

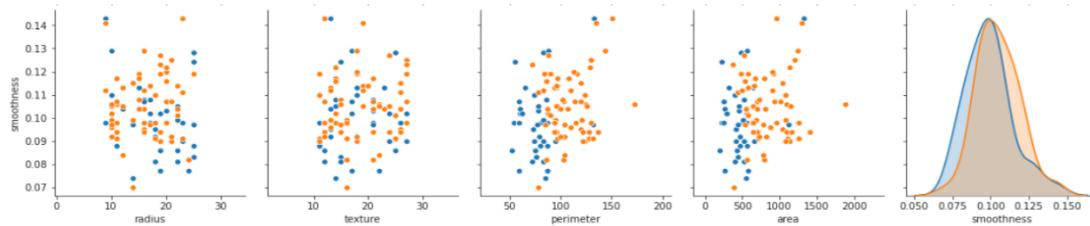


Figure 3. Training and test data

### 3.1.3. Application of the models on the data

Application of the models to the data can be summarised as follows:

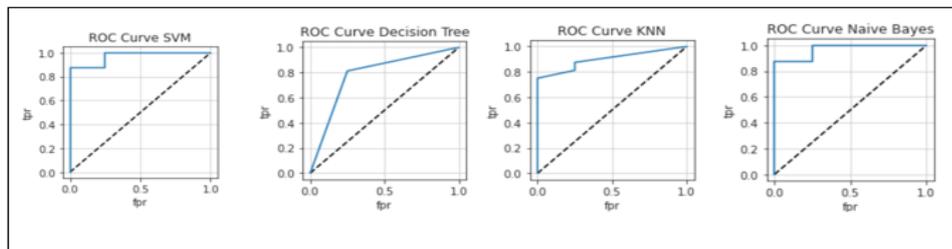


Figure 4. Application of the models to the data

Here, we used a set of well-known supervised learning algorithms [15], namely SVM, decision tree [16], KNN, and Naive Bayes [9], all used to predict a new individual with prostate cancer. We note that all these algorithms were developed on the Python framework. In a comparative study, we note an improvement in the curve for the Bayes classifier, which gives us a result with 96% better accuracy than the other models, unlike the methods developed in [2], [3], which remained at 76%.

#### Decision tree

This algorithm is, as its name suggests, a decision support tool that allows a population of individuals to be divided into homogeneous groups according to discriminating attributes to a fixed and known objective. It allows predictions to be made based on known data on the problem by reducing, level by level, the domain of solutions [15], [17], [18].

Here, each internal node of the decision tree relates to a discriminant attribute of the elements to be classified which allows to distribute these elements homogeneously between the various sons of this node to its sons represent the discriminant values of the attribute of the node. And finally, the leaves of a decision tree are its predictions concerning the data to be classified.

#### Support Vector Machine (SVM)

It is one of the most powerful supervised learning algorithms based on the concept of hyperplane and is a simple classifier called the maximum margin classifier. This algorithm generates, for a given training sample, an optimal hyperplane that maximizes the margin between data points of different classes. For a two-class data set, the closest objects in each class must be well separated from the decision boundary [17], [19], [20], [21].

#### K-nearest neighbor (KNN)

It is one of the most widely used supervised learning algorithms for its simplicity, much more so in classification problems than in regression. It has weaknesses in that it uses all the data for training while making predictions [17], [19], [22].

#### Naive Bayes (NB)

It is a probabilistic classifier, based on Bayes' theorem. In simple terms, a Naive Bayes classifier assumes that the presence of a particular feature in a class is not related to the presence of any other feature. To this, each feature makes an independent and equal contribution to the result [17], [20] [23], [24] [25].

**3.1.4. Segmentation with the Gaussian Mixture Model [26]**

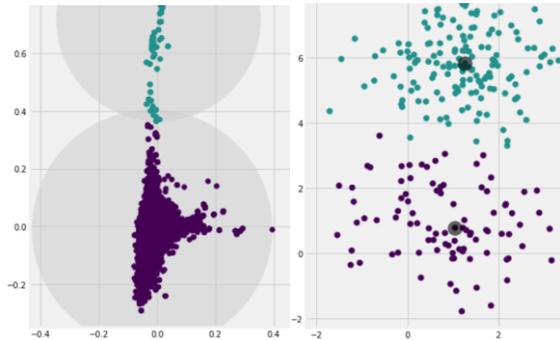


Figure 5. Segmentation with the Gaussian Mixture Model

Our data are initially classified by our model into two classes: positive, and negative. This leads us to a co-relation matrix (figure 6), the main objective of which is to reduce certain irrelevant variables, effectively those that are more correlated with our class of interest, which is the diagnosis\_result. We note here that all the variables have a low correlation with the class, hence the interest in using them in the following.

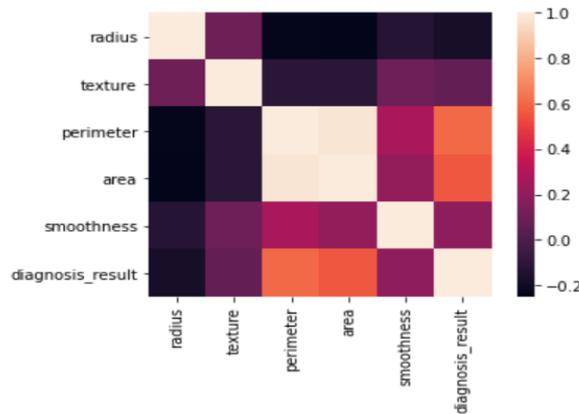


Figure 6. Bayesian networks and co-relation matrix

**3.1.5. Validation of the combined model by the confusion matrix**

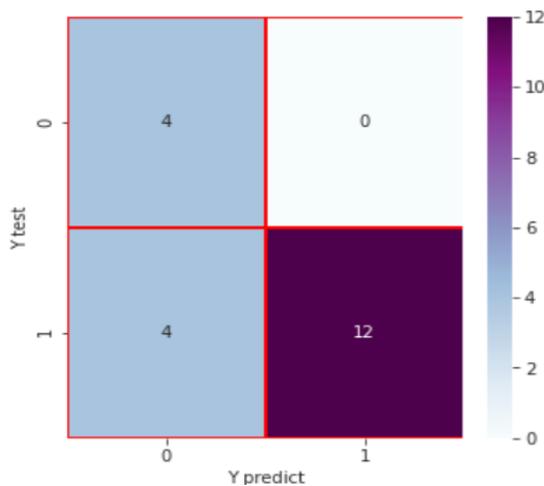


Figure 7: Confusion matrix

We can say here that our model performs well based on the results of our confusion matrix, on 4 images of patients with prostate cancer the model correctly identified them and on 16 scanned images of patients without pain, the model correctly identified 12 and got 4 wrong.

### 3.1.6. Validation of the model with the Rock Curve

After training with the models on the data, we found the results to be 96% satisfactory. This is shown in the figure below by the rocky curve. By seeing this figure, we can say that our models have learned well and that they can classify a new individual with good accuracy (prostate cancer or not).

```
Out[99]: <matplotlib.legend.Legend at 0xb3c8f13ac8>
```

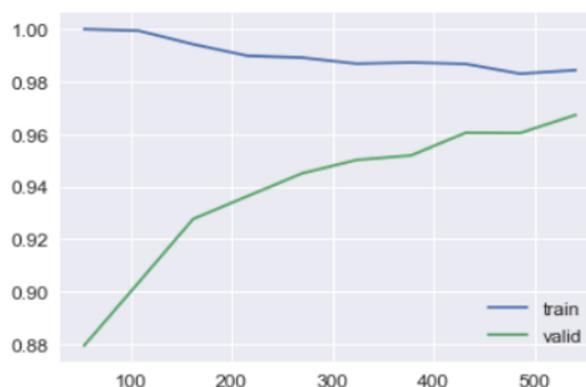


Figure 8. Rock curve

## 4. CONCLUSION

The growing volume of medical data is putting analysts in a difficult position when it comes to extracting hidden knowledge from data. Several models have emerged, focusing on the notion of distances while ignoring the notion of conditional probability density. Our work has focused on segmentation using mixture models and Bayesian networks for medical data mining. We designed different models with different architectures and then applied these models to the medical database. The algorithms were implemented for real data. The aim was to classify individuals according to the conditional probability density of random variables, in addition to identifying causalities between traits using tests of conditional independence and a correlation measure, both based on  $\chi^2$ . After a brief illustration of several models (decision tree, SVM, K-means, Bayes), we applied our method to data from an epidemiological case-control study of prostate cancer (carried out at the University of Kinshasa's university clinics). After interpretation and discussion, we found that our model enabled us to classify a new individual with 96% accuracy, unlike non-probabilistic models which classify with 78% accuracy. These results show that the methods we have developed are sufficiently accurate, fast, and robust to be used in a clinical context. These tools prove their ability to offer a gain in turnaround time and reproducibility of prostate diagnostic and therapeutic decisions. To close, we noted the slowness of the EM algorithm in case of joint estimation of certain parameter subscript parameters  $\mu_k, \Sigma_k, v_k$ . Thus, in our future research we are thinking of separating the estimation of means and covariance matrices from the estimation of degrees of freedom using the ECM algorithm to speed up the result.

## References

- [1] H. Serhal, N. Abdallah, J.-M. Marion, P. Chauvet, and M. Oueidat, "Overview on prediction, detection, and classification of atrial fibrillation using wavelets and AI on ECG," *Computers in Biology and Medicine*, vol. 142, no. 105168, p. 105168, 2022.
- [2] E. Lincker, C. Guinaudeau, O. Pons, J. Dupire, H. Céline, V. Mousseau and H. Caroline, "Automatic classification of unbalanced and noisy data: application to textbook exercises," in *20th International Conference on Content-based Multimedia Indexing (CBMI 2023)*, 2023.
- [3] C. Mélina and L. Benoît, "Artificial intelligence in nutrition research: perspectives on current and future applications," *Applied Physiology, Nutrition, and Metabolism*, vol. 47, no. 1, pp. 1-8, 2022.
- [4] O. Guiliang, H. Yulin and Z. H. Joshua, "A compressed naive hidden Bayesian classifier," in *International Joint Conference on Neural Networks (IJCNN)*, 2021.

- 
- [5] K. Fizazi, J. M. Piulats, M. N. Reaume, P. Ostler, R. McDermott and al, "Rucaparib or physician's choice in metastatic prostate cancer," *New England Journal of Medicine*, vol. 388, no. 8, pp. 719-732, 2023.
- [6] J. T. Wei, D. Barocas, S. Carlsson, F. Coakley, S. Eggener, R. Etzioni and al, "Early detection of prostate cancer: AUA/SUO guideline part I: prostate cancer screening," *The Journal of Urology*, vol. 210, no. 1, pp. 46-53, 2023.
- [7] C. Biernacki, G. Celeux et G. L. F. Govaert, «Model-based cluster and discriminant analysis with the MIXMOD software,» *Computational Statistics & Data Analysis*, vol. 51, n° %12, p. 589, 2006.
- [8] F. Gerard, B. Hugonnier and S. Varin, "Bayesian networks and causal discovery: what lessons for the synthetic indicator of the quality of education systems in OECD countries," *International Journal of Research in Social Sciences*, vol. 13, no. 1, p. 76, 2023.
- [9] W. Yuan, B. Eckart, K. Kim, V. Jampani, D. Fox and J. Kautz, "Deepgmr: Learning latent gaussian mixture models for registration," in *Computer Vision-ECCV 2020: 16th European Conference, Glasgow, UK, August 23-28, 2020*.
- [10] D. A. Boiko, A. S. Kashin, V. R. Sorokin, Y. V. Agaev, R. G. Zaytsev and V. P. Ananikov, "Analysing ionic liquid systems using real-time electron microscopy and a computational framework combining deep learning and classic computer vision techniques," *Journal of Molecular Liquids*, vol. 376, no. 121, p. 407, 2023.
- [11] D. Daneshvar and A. Behrood, "Estimation of the dynamic modulus of asphalt concretes using random forests algorithm," *International Journal of Pavement Engineering*, vol. 23, no. 3, pp. 57-58, 2022.
- [12] M. McIsaac and J. R. Cook, "Statistical methods for incomplete data: Some results on model misspecification," *Statistical Methods in Medical Research*, vol. 26, no. 1, p. 253, 2017.
- [13] W. H. Organization, *International Statistical Classification of Diseases and related health problems: Alphabetical index*, World Health Organization, 2004.
- [14] S. K. BOUNEBACHE, C. QUANTIN, É. BENZENINE and al., "Bibliographic Review of Database Linking Methods: Applications and Perspectives in the Case of Public Health Data," *Journal de la société française de statistique*, vol. 159, no. 3, pp. 81-82, 2018.
- [15] P. C. Sen, M. Hajra and M. Ghosh, "Supervised classification algorithms in machine learning: A survey and review," in *Emerging Technology in Modelling and Graphics: Proceedings of IEM Graph 2018, 2020*.
- [16] C. S. Lee and P. Y. S. Cheang, "Predictive analytics in business analytics: decision tree," *Advances in Decision Sciences*, vol. 26, no. 1, pp. 24-25, 2022.
- [17] C. Noel and J. Schiltz, *Finite mixture models for an underlying BETA distribution with application to COVID-19 data*, Joint work with Jang SCHILTZ, University of Luxembourg, 2022.
- [18] p. Clin, F. Grogard, D. Andrivon, L. Mailleret and M. F. Mamelin, "The proportion of resistant hosts in mixtures should be biased towards the resistance with the lowest breaking cost," *PLOS Computational Biology*, vol. 19, no. 5, p. 234, 2023.
- [19] S. S. Prasetyowati, "Performance Analysis of the Hybrid Voting Method on the Classification of the Number of Cases of Dengue Fever," *International Journal on Information and Communication Technology (IJoICT)*, vol. 8, no. 1, pp. 12-14, 2022.
- [20] R. Hermansyah and R. Sarno, "Sentiment analysis about product and service evaluation of pt telekomunikasi indonesia tbk from tweets using textblob, naive bayes & K-NN Method," in *International Seminar on Application for Technology of Information and Communication (iSemantic)*, 2020.
- [21] M. Sheykhmousa, M. Mahdianpari, H. Ghanbari and al., "Support vector machine versus random forest for remote sensing image classification: A meta-analysis and systematic," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, no. 1, pp. 6315-6321, 2020.
- [22] K. Taunk, S. De, S. Verma and A. Swetapadma, "A brief review of nearest neighbor algorithm for learning and classification," in *international conference on intelligent computing and control systems (ICCS)*, 2019.
- [23] S. Lauritzen, "Propagation of Probabilistics, Means and Variances in Mixed Graphical Association Models," *Journal of the American Statistical Association*, vol. 87, no. 1, p. 1106, 2016.
- [24] W. Gerych, T. Hartvigsen, L. Buquicchio, E. Agu and E. A. Rundensteiner, "Recurrent bayesian classifier chains for exact multi-label," *Advances in Neural Information Processing Systems*, vol. 34, no. 1, pp. 1590-1591, 2021.
- [25] G. B. Marcot and M. A. Hanea, "What is an optimal value of k in k-fold cross-validation in discrete Bayesian network analysis?," *Computational Statistics*, vol. 36, no. 3, pp. 2013-2015, 2021.
- [26] Z. Luo and Y. Deng, "A matrix method of basic belief assignment's negation in Dempster-Shafer theory," *IEEE Transactions on Fuzzy Systems*, vol. 28, no. 9, pp. 2272-2273, 12 3 2019.