Faculty of Mathematics
and Computer Science

# Machine learning course (ML)

# Bayesian Learning in Medical Diagnosis

Ichim Stefan

*Department of Computer Science, Babeş-Bolyai University*
*1, M. Kogalniceanu Street, 400084, Cluj-Napoca, Romania*
*E-mail: stefan.ichim@stud.ubbcluj.ro*

**Abstract**



*Keywords:*

## 1. Introduction

Medical diagnosis has been an activity one could recall taking place from the dawn of time. Throughout the millenias, it has suffered countless reiterations and experimentations, with the final destination to reach a state where life expectancy could be pushed to its limit. In our current day and age, scientists have reached a point where medical diagnosis has become unfathomably complex. As of 2013, 26.000 different diseases have been discovered, an unbelievably high number for a qualified proffesional to memorize how each and one of them differentiate [5]. By this moment in time, one other domain which has seen a rise in popularity had been computer science. As it is known for all new paradigm shifts, it takes time and patience for them to be integrated among the sate of affairs at the time. Ideas about these two domains intertwining were concepted from late 1950s, but it was only in the early 1970s when the first 'expert systems' started seeing use [16]. Initially, these systems, also known as clinical decision support systems (CDSS's), were planned with the concept to produce the diagnosis fully themselves, without the clinician's own opinions; however, with the passing of time, modern computer algorithms have taken the role of supporting clinicians, the final decision being made from a combination of human expertise and algorithm outputs [15].

CDSS's have been divided into knowledge and non-knowledege based. The former make decisions using IF-THEN rules on a timely updated collection of data, whereas the latter use machine learning, which is also where this research's topic is included: computers learn patterns from experiences, with the drawback that their decisions have no explanations. The general machine learning approaches are represented by support-vector machines, artificial neural networks and genetic algorithms.

The aim of this paper is to create a detailed report about bayesian techniques, a branch of the non-knowledge based CDSS's, and where they situate in the space compared to other machine learning methods used in clinical diagnosis. The term bayes comes from the statistician Thomas Bayes, who managed to prove that unknown events can attain probabilistic qualities. However, it was actually Pierre-Simon Laplace that took his ideas further and concepted the bayes formula (1), on which our techniques are based off of.

Bayes' theorem fundamentally describes the probability of an event occurring based on prior knowledge of conditions that might be related to the event. The theorem expresses how a subjective degree of belief should rationally change to account for evidence. In its mathematical form, it states that the posterior probability of a hypothesis ($H$) given observed evidence ($E$) is proportional to the likelihood of the evidence given the hypothesis, multiplied by the prior probability of the hypothesis. This relationship is expressed in Equation (1), where $P(H|E)$ represents the posterior probability, $P(E|H)$ the likelihood, $P(H)$ the prior probability, and $P(E)$ the evidence probability.

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)} \tag{1}$$

In general, machine learning consists of finding the best hypothesis in a hypothesis space using certain observed data, also classified as training data. One drawback to this approach is that with each iteration, certain hypothesis can be removed entirely from the possible best hypothesis set if they appear to be incosistent with some examples. Bayesian learning methods tackle this challenge differently, in the sense that each observed training sample will increase or decrease the probability of whether our hypothesis is correct or not.

With that being said, there are also many difficulties when working with these models. To begin with, as aforementioned, bayesian models build the hypothesis incrementally, one piece of evidence at a time: needless to say, this feature can prove to be computationally demanding, linear with the number of candidate hypothesises. Besides, some assertions need to be stated before any inference takes place whatsoever, assertions which carry a considerate amount of importance, which only underline further the engineer's duty to have a clear understanding of the world of the problem and its place in that world.

The rest of the paper is organized as follows:

## 2. Related Works

The literature of non-knowledge based variants contains a wide variety of approaches tackling decision support systems in clinical diagnosis. Studies have undergone into this field through support vector machines, genetic algorithms, fuzzy logic approaches, decision trees and combinations of these previously mentioned techniques.

### *2.1. Support Vector Machines in Clinical Diagnosis*

Support Vector Machines (SVMs) have demonstrated significant success in medical diagnosis due to their ability to handle high-dimensional data and create optimal separating hyperplanes between different disease classes. Notable applications include [3] and [1] in cancer diagnosis, [10] and [17] in cardiovascular disease prediction. The main advantage of SVMs lies in their ability to handle non-linear relationships through kernel functions, though they can be computationally intensive for large datasets.

### *2.2. Artificial Neural Networks*

Neural networks have gained considerable attention in medical diagnosis, particularly with the advent of deep learning. Studies by [8], [14] and [12] have shown their effectiveness in medical image analysis and pattern recognition tasks. While neural networks can capture complex relationships in medical data, their "black box" nature poses challenges for clinical interpretation and validation.

### 2.3. Genetic Algorithms and Evolutionary Approaches

Genetic algorithms have been applied to optimize feature selection and parameter tuning in medical diagnosis systems. Research by [9] demonstrates their utility in developing adaptive diagnostic rules. These approaches excel at exploring large solution spaces but may require significant computational resources.

### 2.4. Fuzzy Logic and Decision Trees

Fuzzy logic approaches have proven valuable in handling uncertainty in medical diagnosis, while decision trees offer transparent decision-making processes. Studies combining these methods([6], [2]) have shown promising results in dealing with imprecise medical data while maintaining interpretability.

### 2.5. Hybrid Approaches

Recent research has focused on combining multiple techniques to leverage their respective strengths. For instance, [11] integrated neural networks with fuzzy logic to balance accuracy with interpretability. Similarly, [7] combined SVMs with genetic algorithms for optimal feature selection in disease diagnosis.

### 2.6. Bayesian Methods in Context

While the aforementioned approaches have their merits, Bayesian methods offer distinct advantages in medical diagnosis. Unlike deterministic approaches, they provide natural handling of uncertainty through probability distributions, the ability to incorporate prior medical knowledge, incremental learning capabilities, and probabilistic outputs that align with clinical decision-making.

However, the literature reveals several challenges in implementing Bayesian approaches, including computational complexity in handling large hypothesis spaces, difficulty in specifying appropriate prior distributions, and the need to balance model complexity with interpretability.

This review of existing approaches sets the stage for our detailed exploration of Bayesian techniques in subsequent sections, where we will examine how these methods address the limitations of other approaches while introducing their own unique challenges and solutions.

## 3. Other sections to be added

[4] [13]

## 4. Discussion

## 5. Conclusions and future work

## References


[1] Alam, J., Alam, S., Hossan, A., 2018. Multi-stage lung cancer detection and prediction using multi-class svm classifier , 1–4doi:10.1109/IC4ME2.2018.8465593.

[2] Barach, P., Levashenko, V., Zaitseva, E., 2019. Fuzzy decision trees in medical decision making support systems 8, 37–42. doi:10.1177/2327857919081009.

[3] Chen, H.L., Yang, B., Liu, J., Liu, D.Y., 2011. A support vector machine classifier with rough set-based feature selection for breast cancer diagnosis. Expert Systems with Applications 38, 9014–9022. URL: https://www.sciencedirect.com/science/article/pii/S0957417411001400, doi:https://doi.org/10.1016/j.eswa.2011.01.120.

[4] Dafonte, C., Garabato, D., Álvarez, M.A., Manteiga, M., 2018. Distributed fast self-organized maps for massive spectrophotometric data analysis. Sensors 18.

[5] Espe, S., 2018. Malacards: The human disease database. Journal of the Medical Library Association : JMLA 106, 140–141. doi:10.5195/jmla.2018.253. published: 2018/01/02.

[6] Fan, C.Y., Chang, P.C., Lin, J.J., Hsieh, J., 2011. A hybrid model combining case-based reasoning and fuzzy decision tree for medical data classification. Applied Soft Computing 11, 632–644. URL: https://www.sciencedirect.com/science/article/pii/S1568494609002774, doi:https://doi.org/10.1016/j.asoc.2009.12.023.

[7] Huerta, E., Duval, B., Hao, J.K., 2006. A hybrid ga/svm approach for gene selection and classification of microarray data. EvoWorkshops 2006, LNCS 3907 3907, 34–44. doi:10.1007/11732242_4.

[8] Karabulut, E.M., Ibrikçi, T., 2012. Effective diagnosis of coronary artery disease using the rotation forest ensemble method. Journal of Medical Systems 36, 3011–3018. doi:10.1007/s10916-011-9778-y.

[9] Rani, P., Kumar, R., Ahmed, N.M.O.S., et al., 2021. A decision support system for heart disease prediction based upon machine learning. Journal of Reliable Intelligent Environments 7, 263–275. doi:10.1007/s40860-021-00133-6.

[10] Sali, R., Shavandi, H., Sadeghi, M., 2016. A clinical decision support system based on support vector machine and binary particle swarm optimisation for cardiovascular disease diagnosis. International Journal of Data Mining and Bioinformatics 15, 312. doi:10.1504/ijdmb.2016.078150.

[11] Samuel, O.W., Asogbon, G.M., Sangaiah, A.K., Fang, P., Li, G., 2017. An integrated decision support system based on ann and fuzzy_ahp for heart failure risk prediction. Expert Systems with Applications 68, 163–172. URL: https://www.sciencedirect.com/science/article/pii/S0957417416305516, doi:https://doi.org/10.1016/j.eswa.2016.10.020.

[12] Tate, A., Underwood, J., Acosta, D., Julià-Sapé, M., Majós, C., Moreno-Torres, A., Howe, F., van der Graaf, M., Lefournier, V., Murphy, M., Loosemore, A., Ladroue, C., et al., 2006. Development of a decision support system for diagnosis and grading of brain tumours using in vivo magnetic resonance single voxel spectra. NMR in Biomedicine 19, 411–434.

[13] Tolstikhin, I.O., Bousquet, O., Gelly, S., Schölkopf, B., 2018. Wasserstein auto-encoders, in: 6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings, pp. 1–20.

[14] Uğuz, H., 2012. A biomedical system based on artificial neural network and principal component analysis for diagnosis of the heart valve diseases. Journal of Medical Systems 36, 61–72. doi:10.1007/s10916-010-9446-7.

[15] Wikipedia, a. Clinical decision support system. https://en.wikipedia.org/wiki/Clinical_decision_support_system.

[16] Wikipedia, b. Computer-aided diagnosis. https://en.wikipedia.org/wiki/Computer-aided_diagnosis.

[17] Çomak, E., Arslan, A., İbrahim Türkoğlu, 2007. A decision support system based on support vector machines for diagnosis of the heart valve diseases. Computers in Biology and Medicine 37, 21–27. URL: https://www.sciencedirect.com/science/article/pii/S0010482505001484, doi:https://doi.org/10.1016/j.compbiomed.2005.11.002.