

Literature Review of Machine Learning Feature Selection Methods with emphasis on Medical Application on Breast Cancer

Thony Price Niklas Lindqvist

March 2018

Abstract

This document will provide a review of the past work connected to the topic of our Degree Project in Computer Science, Feature Selection Methods for Classification of Breast Cancer. It highlight different efforts of previous research that will inspire our project and be cited while making claims concerning the topic.

1 Introduction

The purpose of this litterature review is to build a foundation of knowledge for our upcoming project. The more we know about the topic beforehand the easier it will be to choose a relevant research question, refer to relevant material in out report as well as navigate the present research in the field. This document will also easily convert to a chapter about previous research.

The structure of the document:

- **Title:** Litterature title and reference.
- **Notes:** Key points from text.
- **Usage:** Potential use to us.

2 Litterature

Title: Machine learning for medical diagnosis: history, state of the art and perspective [6].

Notes: An overview of the development of intelligent data analysis in medicine from a machine learning perspective: a historical view, a state-of-the-art view, and a view on some future trends in this subfield of applied artificial intelligence.

Usage: The paper goes far back and covers the outlook for Machine Learning in the 60's to state of the art. This overview of how this dicipline evolved is good to present in our project too.

Title: Improve Computer-Aided Diagnosis With Machine Learning Techniques Using Undiagnosed Samples [8].

Notes:

Usage:

Title: Global trends in breast cancer incidence and mortality 1973–1997 [2].

Notes:

Usage:

Title: Beyond randomized controlled trials - Organized mammographic screening substantially reduces breast carcinoma mortality [13].

Notes:

Usage:

Title: Discovering Mammography-based Machine Learning Classifiers for Breast Cancer Diagnosis [12].

Notes:

Usage:

Title: Classifier ensemble construction with rotation forest to improve medical diagnosis performance of machine learning algorithms [10].

Notes:

Usage:

Title: Support vector machines combined with feature selection for breast cancer diagnosis [1].

Notes:

Usage:

Title: A comparative study on the effect of feature selection on classification accuracy [7].

Notes:

Usage:

Title: MUC-4 Evaluation metrics [11].

Notes:

Usage:

Title: UCI Machine Learning Repository [4].

Notes:

Usage:

Title: Radiographer involvement in mammography image interpretation: A survey of United Kingdom practice [3].

Notes:

Usage:

Title: Breast Cancer in Low- and Middle-Income Countries: Why We Need Pathology Capability to Solve This Challenge [9].

Notes:

Usage:

Title: An introduction to variable and feature selection [5].

Notes:

Usage:

2.1 Conclusion

Here is a conclusion!

References

- [1] AKAY, M. F. Support vector machines combined with feature selection for breast cancer diagnosis. *Expert Systems with Applications* 36, 2, Part 2 (2009), 3240 – 3247.

- [2] ALTHUIS, M. D., DOZIER, J. M., ANDERSON, W. F., DEVESA, S. S., AND BRINTON, L. A. Global trends in breast cancer incidence and mortality 1973–1997. *International Journal of Epidemiology* 34, 2 (2005), 405–412.
- [3] CULPAN, A. Radiographer involvement in mammography image interpretation: A survey of united kingdom practice. *Radiography* 22, 4 (2016), 306 – 312.
- [4] DHEERU, D., AND KARRA TANISKIDOU, E. UCI machine learning repository, 2017.
- [5] GUYON, I., AND ELISSEEFF, A. An introduction to variable and feature selection. *An Introduction to Variable and Feature Selection* 3 (2003), 1157–1182.
- [6] IGOR, K. Machine learning for medical diagnosis: history, state of the art and perspective. *Artificial Intelligence in Medicine* 23 (2001), 89–109.
- [7] KARABULUT, E. M., ÖZEL, S. A., AND İBRIKÇİ, T. A comparative study on the effect of feature selection on classification accuracy. *Procedia Technology* 1 (2012), 323 – 327.
- [8] LI, M., AND ZHOU, Z. H. Improve computer-aided diagnosis with machine learning techniques using undiagnosed samples. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans* 37 (Nov 2007), 1088–1098.
- [9] MARTEI, Y. M., PACE, L. E., BROCK, J. E., AND SHULMAN, L. N. Breast cancer in low- and middle-income countries: Why we need pathology capability to solve this challenge. *Clinics in Laboratory Medicine* 38, 1 (2018), 161 – 173.
- [10] OZCIFT, A., AND GULTEN, A. Classifier ensemble construction with rotation forest to improve medical diagnosis performance of machine learning algorithms. *Computer Methods and Programs in Biomedicine* 104, 3 (2011), 443 – 451.
- [11] PH.D., N. C. Muc-4 evaluation metrics. In *Fourth message understanding conference (MUC-4)* (McLean, Virginia, June 1992).
- [12] RAMOS-POLLÁN, R., GUEVARA-LÓPEZ, M. A., SUÁREZ-ORTEGA, C., DÍAZ-HERRERO, G., FRANCO-VALIENTE, J. M., RUBIO-DEL SOLAR, M., GONZÁLEZ-DE POSADA, N., VAZ, M. A. P., LOUREIRO, J., AND RAMOS, I. Discovering mammography-based machine learning classifiers for breast cancer diagnosis. *Journal of Medical Systems* 36 (Aug 2012), 2259–2269.
- [13] TABÁR, L., VITAK, B., H CHEN, H., F YEN, M., DUFFY, S., AND SMITH, R. Beyond randomized controlled trials - organized mammographic screening substantially reduces breast carcinoma mortality. 1724–31.