JOM RESEARCH

Presenting results from ML analysis

Tengku Muhammad Hanis Bin Tengku Mokhtar, PhD

October 23, 2024

JOM RESEARCH

Contents

General tips

- Make an effort to use/produce a high quality data
- Be very clear regarding the goal of the machine learning analysis and the clinical goal
  - Prognosis, diagnosis, or screening purpose
- Be aware/transparent about the limitation of the data
  - Small sample
  - Certain bias in the data
  - Data imbalance
- Be aware/transparent about the limitation of the model
  - Models such as decision tree and ANN can be unstable especially in small sample data
- Choose appropriate performance metrics

Guidelines

- Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary view
- Comprehensive guide and checklist for clinicians to evaluate artificial intelligence and machine learning methodological research

JOM RESEARCH

Example of ML papers

diagnostics

MDPI

Article

Over-the-Counter Breast Cancer Classification Using Machine Learning and Patient Registration Records

Tengku Muhammad Hanis [1,*], Nur Intan Raihana Ruhaiyem [2], Wan Nor Arifin [3], Juhara Haron [4,5], Wan Faiziah Wan Abdul Rahman [5,6], Rosni Abdullah [2] and Kamarul Imran Musa [1,*]

1 Department of Community Medicine, School of Medical Sciences, Universiti Sains Malaysia, Kubang Kerian 16150, Kelantan, Malaysia
2 School of Computer Sciences, Universiti Sains Malaysia, Gelugor 11800, Penang, Malaysia
3 Biostatistics and Research Methodology Unit, School of Medical Sciences, Universiti Sains Malaysia, Kubang Kerian 16150, Kelantan, Malaysia
4 Department of Radiology, School of Medical Sciences, Universiti Sains Malaysia, Kubang Kerian 16150, Kelantan, Malaysia
5 Breast Cancer Awareness and Research Unit, Hospital Universiti Sains Malaysia, Kubang Kerian 16150, Kelantan, Malaysia
6 Department of Pathology, School of Medical Sciences, Universiti Sains Malaysia, Kubang Kerian 16150, Kelantan, Malaysia
* Correspondence: tengkuhanismokhtar@gmail.com (T.M.H.); drkamarul@usm.my (K.I.M.)

Abstract: This study aims to determine the feasibility of machine learning (ML) and patient registration record to be utilised to develop an over-the-counter (OTC) screening model for breast cancer risk estimation. Data were retrospectively collected from women who came to the Hospital Universiti Sains Malaysia, Malaysia for breast-related problems. Eight ML models were used: k-nearest neighbour (kNN), elastic-net logistic regression, multivariate adaptive regression splines, artificial neural network, partial least square, random forest, support vector machine (SVM), and extreme gradient boosting. Features utilised for the development of the screening models were limited to information in the patient registration form. The final model was evaluated in terms of performance across a mammographic density. Additionally, the feature importance of the final model was assessed using the model agnostic approach. kNN had the highest Youden J index, precision, and PR-AUC, while SVM had the highest F2 score. The kNN model was selected as the final model. The model had a balanced performance in terms of sensitivity, specificity, and PR-AUC across the mammographic density groups. The most important feature was the age at examination. In conclusion, this study showed that ML and patient registration information are feasible to be used as the OTC screening model for breast cancer.

Keywords: Asian women; breast cancer; explainable artificial intelligence; machine learning; medical consultation delays; screening model; clinical decision support systems

check for updates

Citation: Hanis, T.M.; Ruhaiyem, N.I.R.; Arifin, W.N.; Haron, J.; Wan Abdul Rahman, W.F.; Abdullah, R.; Musa, K.I. Over-the-Counter Breast Cancer Classification Using Machine Learning and Patient Registration Records. *Diagnostics* **2022**, *12*, 2826. https://doi.org/10.3390/diagnostics12112826

Academic Editor: Md Mohaimenul Islam

Received: 10 September 2022
Accepted: 15 October 2022
Published: 16 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

Breast cancer is the most common cancer among women in at least 140 countries [1]. The WHO aims to reduce global breast cancer mortality by 25% annually between 2020 and 2040, which is equivalent to 2.5 million breast cancer death worldwide [2]. Generally, breast cancer affects women above the age of 50 and the risk of having the disease increases with increased age [3–5]. The risk factors for breast cancer are mainly divided into two groups [6]. The inherent risk factors include a family history of breast cancer, age, and gender, while the extrinsic risk factors include diet and lifestyle. The risk factors differ according to the individual and population. One of the important risk factors for breast cancer is mammographic density which reflects the amount of dense and fatty tissue in the breast [7,8]. Women with denser breasts had four to six times higher chances of developing

diagnostics

MDPI

Article

Developing a Supplementary Diagnostic Tool for Breast Cancer Risk Estimation Using Ensemble Transfer Learning

Tengku Muhammad Hanis [1,*], Nur Intan Raihana Ruhaiyem [2], Wan Nor Arifin [3], Juhara Haron [4,5], Wan Faiziah Wan Abdul Rahman [5,6], Rosni Abdullah [2] and Kamarul Imran Musa [1,*]

1 Department of Community Medicine, School of Medical Sciences, Universiti Sains Malaysia, Kubang Kerian 16150, Kelantan, Malaysia
2 School of Computer Sciences, Universiti Sains Malaysia, Gelugor 11800, Penang, Malaysia; intanraihana@usm.my (N.I.R.R.); rosni@usm.my (R.A.)
3 Biostatistics and Research Methodology Unit, School of Medical Sciences, Universiti Sains Malaysia, Kubang Kerian 16150, Kelantan, Malaysia; wnarifin@usm.my
4 Department of Radiology, School of Medical Sciences, Universiti Sains Malaysia, Kubang Kerian 16150, Kelantan, Malaysia; drjuhara@usm.my
5 Breast Cancer Awareness and Research Unit, Hospital Universiti Sains Malaysia, Kubang Kerian 16150, Kelantan, Malaysia; wfaiziah@usm.my
6 Department of Pathology, School of Medical Sciences, Universiti Sains Malaysia, Kubang Kerian 16150, Kelantan, Malaysia
* Correspondence: tengkuhanismokhtar@gmail.com (T.M.H.); drkamarul@usm.my (K.I.M.)

Abstract: Breast cancer is the most prevalent cancer worldwide. Thus, it is necessary to improve the efficiency of the medical workflow of the disease. Therefore, this study aims to develop a supplementary diagnostic tool for radiologists using ensemble transfer learning and digital mammograms. The digital mammograms and their associated information were collected from the department of radiology and pathology at Hospital Universiti Sains Malaysia. Thirteen pre-trained networks were selected and tested in this study. ResNet101V2 and ResNet152 had the highest mean PR-AUC, MobileNetV3Small and ResNet152 had the highest mean precision, ResNet101 had the highest mean F1 score, and ResNet152 and ResNet152V2 had the highest mean Youden J index. Subsequently, three ensemble models were developed using the top three pre-trained networks whose ranking was based on PR-AUC values, precision, and F1 scores. The final ensemble model, which consisted of Resnet101, Resnet152, and ResNet50V2, had a mean precision value, F1 score, and Youden J index of 0.82, 0.68, and 0.12, respectively. Additionally, the final model demonstrated balanced performance across mammographic density. In conclusion, this study demonstrates the good performance of ensemble transfer learning and digital mammograms in breast cancer risk estimation. This model can be utilised as a supplementary diagnostic tool for radiologists, thus reducing their workloads and further improving the medical workflow in the screening and diagnosis of breast cancer.

Keywords: Asian women; breast cancer; transfer learning; deep learning; artificial intelligence; diagnostic screening; mammography; radiologists

check for updates

Citation: Hanis, T.M.; Ruhaiyem, N.I.R.; Arifin, W.N.; Haron, J.; Wan Abdul Rahman, W.F.; Abdullah, R.; Musa, K.I. Developing a Supplementary Diagnostic Tool for Breast Cancer Risk Estimation Using Ensemble Transfer Learning. *Diagnostics* **2023**, *13*, 1780. https://doi.org/10.3390/diagnostics13101780

Academic Editor: Dechang Chen

Received: 28 February 2023
Revised: 14 March 2023
Accepted: 23 March 2023
Published: 18 May 2023

1. Introduction

Breast cancer is the most commonly diagnosed cancer worldwide [1]. Breast cancer is considered the leading cause of cancer-related death in the twelve regions of the world [2]. This disease accounts for one in four and one in six cancer cases and cancer deaths among women, respectively [3]. In an attempt to combat the disease, the World Health Organization (WHO) proposed a global breast cancer initiative in 2021, which will run over 20 years and consist of three key elements [4]. One of these three key elements is the promotion of the early detection of breast cancer. The early detection of this disease ensures that a patient receives timely treatment. Thus, any delay in the medical workflow of breast cancer screening and diagnosis will influence the prognosis of the disease.

Any question?

tengkuhanismokhtar@gmail.com
jom.research.malaysia@gmail.com