

# Concentrated *Thermomics* for Early Diagnosis of Breast Cancer <sup>†</sup>

Bardia Yousefi <sup>1,2,\*</sup> , Michelle Hershman <sup>3</sup>, Henrique C. Fernandes <sup>4</sup>  and Xavier P. V. Maldague <sup>1,\*</sup><sup>1</sup> Department of Electrical and Computer Engineering, Laval University, Quebec City, QC G1V 0A6, Canada<sup>2</sup> Fischell Department of Bioengineering, University of Maryland, College Park, MD 20850, USA<sup>3</sup> Department of Radiology, University of Pennsylvania, Philadelphia, PA 19104, USA; Michelle.Hershman@pennmedicine.upenn.edu<sup>4</sup> Faculty of Computing, Federal University of Uberlandia, Uberlandia 38408-100, Brazil; henriquecoelhofernandes@gmail.com

\* Correspondence: Bardia.Yousefi.1@ulaval.ca (B.Y.); Xavier.Maldague@ulaval.ca (X.P.V.M.)

<sup>†</sup> Presented at the 16th International Workshop on Advanced Infrared Technology & Applications, 26–28 October 2021; Available online: <https://aita2021.sciforum.net/>.

**Abstract:** Thermography has been employed broadly as a corresponding diagnostic instrument in breast cancer diagnosis. Among thermographic techniques, deep neural networks show an unequivocal potential to detect heterogeneous thermal patterns related to vasodilation in breast cancer cases. Such methods are used to extract high-dimensional thermal features, known as deep *thermomics*. In this study, we applied convex non-negative matrix factorization (convex NMF) to extract three predominant bases of thermal sequences. Then, the data were fed into a sparse autoencoder model, known as SPAER, to extract low-dimensional deep thermomics, which were then used to assist the clinical breast exam (CBE) in breast cancer screening. The application of convex NMF-SPAER, combining clinical and demographic covariates, yielded a result of 79.3% (73.5%, 86.9%); the highest result belonged to NMF-SPAER at 84.9% (79.3%, 88.7%). The proposed approach preserved thermal heterogeneity and led to early detection of breast cancer. It can be used as a noninvasive tool aiding CBE.



**Citation:** Yousefi, B.; Hershman, M.; Fernandes, H.C.; Maldague, X.P.V. Concentrated *Thermomics* for Early Diagnosis of Breast Cancer. *Eng. Proc.* **2021**, *8*, 30. <https://doi.org/10.3390/engproc2021008030>

Academic Editors: Giovanni Ferrarini, Paolo Bison and Gianluca Cadelano

Published: 6 December 2021

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



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** deep learning thermography; convex non-negative matrix factorization; thermal heterogeneity; deep thermomics; breast cancer

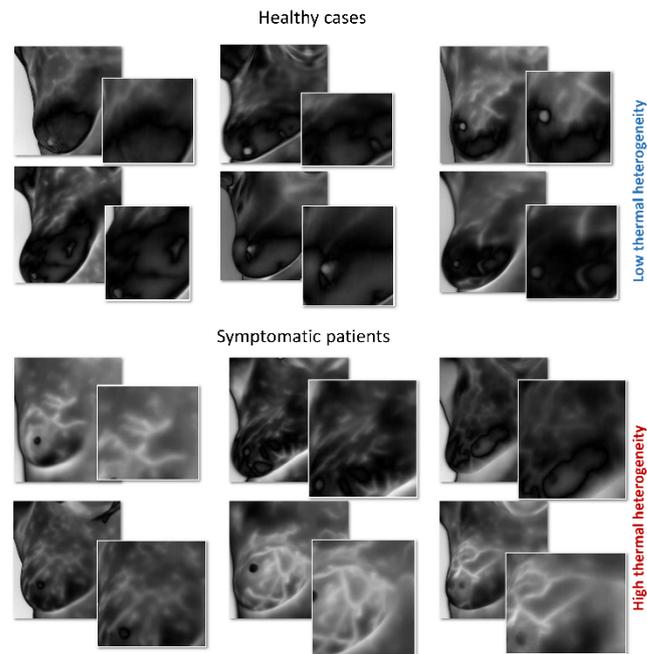
## 1. Introduction

Breast cancer shows high survival rates due to the current progresses in imaging modalities and treatment planning, but it still categorized as the most fatal cancer among women [1]. Here, a deep learning model is applied for dynamic thermography to aid the process of screening prior to mammography and in a clinical breast exam (CBE) to expedite diagnosis at an early stage. Several studies have confirmed the value of infrared thermography in detecting hypervascularity and hyperthermia in non-palpable breast cancer [2–4]. This study applies a convex matrix approximation and a deep learning model to extract the most predominant thermal patterns, called *latent space thermomics*. The model was trained and used alongside CBE for early detection of breast cancer and showed high accuracy in detecting symptomatic patients.

## 2. Materials and Methods

Thermal heterogeneity can be detected by using different matrix factorizations (MF) [5–20]. Thermal variance can be used to evaluate the thermal heterogeneity of different components. Low-rank matrix approximation, using semi- and convex NMF, provides more freedom to bases when compared to NMF, while being controlled by non-negative coefficients. This means that despite the presence of potentially negative bases, they do not overlap. This property causes these algorithms to perform the grouping of bases, driven by

their thermal variability on input thermal data. Figure 1 shows the low-rank representation of thermal images in the breast area for healthy and symptomatic participants, using convex NMF.



**Figure 1.** Twelve examples of convex NMF-driven low-rank thermal matrix approximation for six healthy cases and six symptomatic or cancer cases. Some otherwise healthy cases reported pain and changes in the breast area that are clearly projected in the thermal images.

Thermal imaging throughputs extracted from thermal images, called *thermomics* [14–17], are used extensively to deliver diagnostic solutions, such as radiomics, for early-stage breast cancers [18–20]. The efficiency of thermomics is known to have been accompanied by clinical variables. Similarly, the abundance of thermomics impedes the system’s performance by overfitting problems, which is known as the *curse of dimensionality* [2,14].

Deep learning has influenced thermomics and automated diagnosis through thermographic imaging. A sparse deep convolutional autoencoder model (SPAER) is used to extract low-dimensional features from thermographic images [14].

Three low-ranked representative bases, using convex NMF, were extracted and used as three channels of input data that were fed into the trained SPAER model. Then, SPAER extracted 16 deep thermomics from the input images, while encoding heterogeneous thermal patterns within this output vector. These 16 deep thermomics were then used to train a random forest model in making diagnostic decisions.

#### Study Cohort

We used 55 breast cancer screening participants belong to the mastology research database (DMR) to evaluate the strength of this proposed method. This cohort of patients obtained from the Hospital Universitário Antônio Pedro (HUAP) of the Federal University Fluminense [20]. Ethical Committee of the HUAP, under registration number CAEE: 01042812.0.0000.5243 supported by the Brazilian Ministry of Health approved to use this data. This data contains healthy/symptomatic patients, median age: 60 (20,120), multiple race (15 Africans (27.3%), 28 Caucasians (51%), 11 Parda (20%), and one Mulatto (1.8%)), and 18 had diabetes in their family history while 9 women went undergoing hormone replacement [20].

### 3. Results

The presented approach was tested on 55 breast cancer screening participants to investigate the capability of the model in diagnosing abnormality. The obtained results were compared to ground truth (the gold standard) to determine the system's accuracy. Three predominant low-rank matrices were calculated for 23 baseline thermal images utilizing matrix factorization. The results followed our proposed hypothesis, and thermal heterogeneity was intensified in the breast area for the symptomatic cases that had potential angiogenesis (blood vessel formation) and vasodilation, as more heterogeneity was exhibited in the ROI of symptomatic cases ("healthy with symptoms" or "cancerous" cases) than in the ROI of "healthy" cases (Figure 1). Some otherwise healthy cases were categorized as symptomatic participants, if they reported nipple changes and pain but the results of mammography and biopsy showed no signs of cancer.

Our trained SPAER model [14] extracted 16 deep thermomics from the three channels' embedded convex low-ranked thermal matrix approximation, with dimensionality of  $512 \times 512 \times 3$ . The model was trained using an *Adam* optimizer for 200 epochs for a batch size of 8, with a sparsity  $\downarrow_1$  regularization value of  $10^{-5}$  in dense layers.

#### Classification Outcome

The extracted deep thermomics produced by the SPAER model were employed to train a random forest model that was used to make preliminary diagnostic decisions for both symptomatic and healthy participants. The model's accuracy was assessed by comparing the model's predictions with the ground truth, using mammography and breast biopsy. We evaluated the system's accuracy using 16 convex deep thermomics biomarkers, along with clinical and demographic covariates (i.e., age, marital status, and family history), with leave-one-out cross-validation. The accuracy yielded by the maximal multivariate model was 84.9% (79.3%, 88.7%) for NMF deep thermomics, which was challenged by NMF-SPAER+Clinical, convex NMF-SPAER+Clinical, and PCT-SPAER+Clinical, with clinical information yielding accuracies of 83.02% (79.2%, 86.8%), 79.3% (73.5%, 86.9%), and 81.1% (75.5%, 84.9%), respectively. IPCT-SPAER+Clinical, sparse PCT-SPAER + Clinical, sparse NMF-SPAER + Clinical, and convex NMF-SPAER + Clinical showed considerably similar accuracy of diagnosis, yielding median accuracies of 79.2% with different variation ranges. The lowest accuracy of these methods belonged to PCT-SPAER, which showed diagnostic accuracy of 75.5% (67.9%, 81.1%). The minimal full multivariate model belonged to sparse NMF-SPAER+Clinical, with an accuracy of 79.2% (73.6%, 83.02%), demonstrating higher restriction in the domain of the solution due to two-fold sparsity and non-negativity in the NMF and SPAER models (see Table 1).

**Table 1.** Classification results for symptomatic versus healthy cases, with leave-one-out cross-validation.

Accuracy of Multivariate Models for Breast Cancer Diagnosis		
Methods	Accuracy <sup>2</sup> (%)	t-Test <sup>3</sup>
IPCT-SPAER	79.2 (73.6, 84.9)	0.06, 0.95
PCT-SPAER	75.5 (67.9, 81.1)	2.1, 0.04
NMF-SPAER	84.9 (79.3, 88.7)	2.9, 0.004
Sparse PCT-SPAER	79.2 (73.6, 84.9)	1.2, 0.23
Sparse NMF-SPAER	79.3 (73.5, 83.02)	1.08, 0.28
Convex NMF-SPAER	75.5 (69.8, 81.1)	-
Clinical <sup>1</sup>	71.7 (66.04, 79.3)	1.4, 0.17
IPCT-SPAER + Clinical	79.3 (73.1, 81.2)	0.85, 0.39
PCT-SPAER + Clinical	81.1 (75.5, 84.9)	1.7, 0.08
NMF-SPAER + Clinical	83.02 (79.2, 86.8)	2.4, 0.01
Sparse PCT-SPAER + Clinical	79.3 (69.8, 84.9)	0.9, 0.3
Sparse NMF-SPAER + Clinical	79.2 (73.6, 83.02)	1.5, 0.15
Convex NMF-SPAER + Clinical	79.3 (73.5, 86.9)	1.39, 0.16

<sup>1</sup> Clinical and demographic covariates are family history, age, and marital status. <sup>2</sup> Classification accuracy reported by median ( $\pm$ IQR). <sup>3</sup> t-test calculated for each method versus maximal accuracy (t-statistic, p-value).

The model with clinical variables yielded an accuracy of 71.7% (66.04%, 79.3%). Moreover, we measured the statistical significance cross-validated convex NMF-SPAER + Clinical versus other approaches using t-test (see Table 1). Convex NMF-SPAER + Clinical manifested difference to NMF-SPAER+Clinical (t-statistic = 2.9,  $p$ -value = 0.004) and NMF-SPAER+Clinical (t-statistic = 2.4,  $p$ -value = 0.01). Nevertheless, other methods exhibit statistical similarity to the cross-validated results of convex-NMF-SPAER+Clinical. This study was conducted using Python programming language and the TensorFlow and Keras Python libraries [21,22].

#### 4. Conclusions

This study challenged the thermal diagnostic system with a thermomics approach generated by a sparse deep convolutional autoencoder, known as the SPAER model, while using convex low-rank thermal matrix approximation. This approach integrated convexity and sparsity to generate low-dimensional thermomics, while simultaneously preserving thermal patterns that are vital in determining angiogenesis and vasodilation in symptomatic patients. Convex non-negative matrix factorization (convex NMF) was used to reduce 23 thermal images to three channels. Then, the SPAER model reduced the dimensionality of the input data from 786,432 to 16 imaging biomarkers. The method was compared to five state-of-the-art matrix factorizations in thermography, and tested on 55 breast cancer screening cases that underwent dynamic thermography. The best accuracy reported in this study belonged to NMF-SPAER, without demographics and clinical information preserving heterogeneous thermal patterns, which had a result of 84.9% (79.3%, 88.7%). In future work, we can analyze the effect of aggregation of other available clinical factors or imaging modalities to enhance the diagnostic power of the system.

**Institutional Review Board Statement:** The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by Ethics Committee of HAUP (under registration number CAAE: 01042812.0.0000.5243 provided by the Brazilian Ministry of Health).

#### References

1. Siegel, R.L.; Miller, K.D.; Fuchs, H.E.; Jemal, A. Cancer statistics. *Cancer J. Clin.* **2021**, *71*, 7–33. [[CrossRef](#)] [[PubMed](#)]
2. Yousefi, B.; Akbari, H.; Maldague, X.P.V. Detecting Vasodilation as Potential Diagnostic Biomarker in Breast Cancer Using Deep Learning-Driven Thermomics. *Biosensors* **2020**, *10*, 164. [[CrossRef](#)] [[PubMed](#)]
3. Gamagami, P. Indirect Signs of Breast Cancer: Angiogenesis Study. In *Atlas of Mammography*; Blackwell Science: Cambridge, MA, USA, 1996; pp. 321–326.
4. Yoshida, S.; Nakagawa, S.; Yahara, T.; Koga, T.; Deguchi, H.; Shirouzu, K. Relationship Between Microvessel Density and Thermographic Hot Areas in Breast Cancer. *Surg. Today* **2003**, *33*, 243–248. [[CrossRef](#)] [[PubMed](#)]
5. Rajic, N. Principal component thermography for flaw contrast enhancement and flaw depth characterisation in composite structures. *Compos. Struct.* **2002**, *58*, 521–528. [[CrossRef](#)]
6. Marinetti, S.; Finesso, L.; Marsilio, E. Matrix factorization methods: Application to thermal NDT/E. *NDT E Int.* **2006**, *39*, 611–616. [[CrossRef](#)]
7. Yousefi, B.; Sharifipour, H.M.; Eskandari, M.; Ibarra-Castanedo, C.; Laurendeau, D.; Watts, R.; Klein, M.; Maldague, X.P.V. Incremental Low Rank Noise Reduction for Robust Infrared Tracking of Body Temperature during Medical Imaging. *Electronics* **2019**, *8*, 1301. [[CrossRef](#)]
8. Ahmed, J.; Gao, B.; Woo, W.L. Wavelet-integrated alternating sparse dictionary matrix decomposition in thermal imaging cfrp defect detection. *IEEE Trans. Ind. Inform.* **2018**, *15*, 4033–4043. [[CrossRef](#)]
9. Yousefi, B.; Sfarra, S.; Castanedo, C.I.; Maldague, X.P. Comparative analysis on thermal non-destructive testing imagery applying Candid Covariance-Free Incremental Principal Component Thermography (CCIPCT). *Infrared Phys. Technol.* **2017**, *85*, 163–169. [[CrossRef](#)]
10. Yousefi, B.; Sfarra, S.; Sarasini, F.; Castanedo, C.I.; Maldague, X.P. Low-rank sparse principal component thermography (sparse-PCT): Comparative assessment on detection of subsurface defects. *Infrared Phys. Technol.* **2019**, *98*, 278–284. [[CrossRef](#)]
11. Wu, J.-Y.; Sfarra, S.; Yao, Y. Sparse Principal Component Thermography for Subsurface Defect Detection in Composite Products. *IEEE Trans. Ind. Inform.* **2018**, *14*, 5594–5600. [[CrossRef](#)]
12. Ding, C.H.; Li, T.; Jordan, M. Convex and Semi-Nonnegative Matrix Factorizations. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 45–55. [[CrossRef](#)] [[PubMed](#)]
13. Yousefi, B.; Sfarra, S.; Ibarra-Castanedo, C.; Avdelidis, N.P.; Maldague, X.P.V. Thermography data fusion and nonnegative matrix factorization for the evaluation of cultural heritage objects and buildings. *J. Therm. Anal. Calorim.* **2019**, *136*, 943–955. [[CrossRef](#)]

14. Yousefi, B.; Akbari, H.; Hershman, M.; Kawakita, S.; Fernandes, H.; Ibarra-Castanedo, C.; Ahadian, S.; Maldague, X. SPAER: Sparse Deep Convolutional Autoencoder Model to Extract Low Dimensional Imaging Biomarkers for Early Detection of Breast Cancer Using Dynamic Thermography. *Appl. Sci.* **2021**, *11*, 3248. [[CrossRef](#)]
15. Yousefi, B.; Castanedo, C.I.; Maldague, X.P.V. Measuring Heterogeneous Thermal Patterns in Infrared-Based Diagnostic Systems Using Sparse Low-Rank Matrix Approximation: Comparative Study. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 4501209. [[CrossRef](#)]
16. Yousefi, B.; Castanedo, C.I.; Maldague, X.P. Low-rank Convex/Sparse Thermal Matrix Approximation for Infrared-based Diagnostic System. *arXiv* **2020**, arXiv:2010.06784.
17. Yousefi, B.; Sharifipour, H.M.; Maldague, X.P.V. A Diagnostic Biomarker for Breast Cancer Screening via Hilbert Embedded Deep Low-Rank Matrix Approximation. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 4504809. [[CrossRef](#)]
18. Gillies, R.J.; Kinahan, P.E.; Hricak, H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology* **2016**, *278*, 563–577. [[CrossRef](#)] [[PubMed](#)]
19. Lambin, P.; Leijenaar, R.T.H.; Deist, T.M.; Peerlings, J.; de Jong, E.E.C.; van Timmeren, J.; Sanduleanu, S.; Larue, R.T.H.M.; Even, A.J.G.; Jochems, A.; et al. Radiomics: The bridge between medical imaging and personalized medicine. *Nat. Rev. Clin. Oncol.* **2017**, *14*, 749–762. [[CrossRef](#)] [[PubMed](#)]
20. Silva, L.F.; Saade, D.C.M.; Sequeiros, G.O.; Silva, A.C.; Paiva, A.C.; Bravo, R.S.; Conci, A. A New Database for Breast Research with Infrared Image. *J. Med. Imaging Health Inform.* **2014**, *4*, 92–100. [[CrossRef](#)]
21. Arpit, D.; Zhou, Y.; Ngo, H.; Govindaraju, V. Why regularized auto-encoders learn sparse representation? In Proceedings of the 33rd International Conference on Machine Learning, New York, NY, USA, 19–24 June 2016; pp. 136–144.
22. Google. *Python 3 Google Compute Engine Backend, T4, and P100 GPU and 27.4 Gb RAM*; Google: Mountain View, CA, USA, 2020.