# MEDICAL APPLICATIONS of Laser Molecular Imaging and Machine Learning

# MEDICAL APPLICATIONS of Laser Molecular Imaging and Machine Learning

**Yury V. Kistenev**
**Alexey V. Borisov**
**Denis A. Vrazhnov**

**SPIE PRESS**
Bellingham, Washington USA

Library of Congress Control Number: 2021938748

Published by
SPIE
P.O. Box 10
Bellingham, Washington 98227-0010 USA
Phone: +1 360.676.3290
Fax: +1 360.647.1445
Email: books@spie.org
Web: http://spie.org



Cover image: Great Bergens, courtesy of Shutterstock

Printed in the United States of America.
Last updated 14 July 2021.
For updates to this book, visit http://spie.org and type “PM333” in the search field.

SPIE.

# Table of Contents

# Preface

This book examines various biophotonics applications associated with modern machine learning techniques and laser molecular imaging and spectroscopy. Most of the existing books focus on either a specific instrumental method, such as terahertz and IR spectroscopy or Raman scattering, or a limited number of mathematical tools for raw data analysis. We describe a thorough review of molecular imaging technologies and current machine learning approaches to perform data analysis of gaseous, liquid samples of biological origin and biotissues. Much of the material highlights applications of machine learning to develop non-invasive medical diagnostics tools.

Here, we present the basics of machine learning methods, which consider the specificity of laser molecular imaging and spectroscopy medical data features, such as the high dimensionality of raw data and a low number of samples leading to a lack of representation. Modern trends such as deep learning are not applied broadly in similar tasks because of the small volume of available samples. There are two main reasons for this. The first is the high variability of biological systems, which makes biophysical relations difficult to discover. The second is ethical restrictions on studies with living beings. These reasons require new methods to deal with high-dimensional but low-numbered data (contrary to big data, which operates with low-dimensional yet outnumbered data).

Speaking of the development of both biophotonics hardware and software, we try to make future forecasts based on current trends in these fields. We also discuss available hardware platforms: home and self, medical screening, and specialized devices for end-level diagnosis. General trends include personalized medicine and bringing high-tech diagnostics from hospitals directly to individuals.

This book focuses on the most suitable approaches for medical screening and monitoring. Some ideas can be used in personal diagnosis tool design and production. Machine learning pipeline algorithms can be useful for high-accuracy multi-modal diagnosis.

This book is intended for specialists in the fields of biomedical optics, laser spectroscopy, bioengineering, and medical engineering. To provide practical help to readers who plan to use the machine learning methods in their research, the Supplemental Materials include sample datasets and the Python modules for the most useful algorithms described in the book. The link to the Supplemental Materials website is

https://github.com/biophotonics-lab-tsu/monograph

For convenience, we use Roman superscript numbers to link key terms and specific methods with the chapter that defines them. Also, the first mention of a method is italicized.

## Acknowledgments

The authors would like to thank their colleagues from the Laboratory of Biophotonics of Tomsk State University, especially Anatasia Knyazkova and Olga Zakharova, for their help with the technical work in this book.

The materials have been collected under partial support of the Russian Foundation for Basic Research (Grant No. 7-00-00186), RFBR, and Administration of Tomsk Region (grant No. 18-42-703012), FCPIR, contract No. 14.578.21.0082.

This work was supported by the Government of the Russian Federation (Agreement No. 075-15-2021-615 of 04 June 2021) to support scientific research projects implemented under the supervision of leading scientists at Russian institutions (Russian institutions of higher education).

**Yury V. Kistenev**
**Alexey V. Borisov**
**Denis A. Vrazhnov**
July 2021

# List of Abbreviations and Acronyms

| | |
|---|---|
| AI | Artificial intelligence |
| ANN | Artificial neural network |
| AR–PAM | Acoustic-resolution photoacoustic microscopy |
| ATR | Attenuated total reflection |
| AUC | Area under the curve |
| BAS | Breath air sample |
| BCC | Basal cell carcinoma |
| BF | Body fluid |
| BLP | Bootstrapped Latin partition |
| CARS | Coherent anti-Stokes Raman scattering |
| CCA | Canonical correlation analysis |
| CI | Confidence interval |
| CNN | Convolutional neural network |
| COPD | Chronic obstructive pulmonary disease |
| CP | Conducting polymer |
| CRDS | Cavity ring-down spectroscopy |
| CTC | Circulating tumor cell |
| CV | Cross-validation |
| CW | Continuous wave |
| DAS | Direct absorption spectroscopy |
| DBSCAN | Density-based spatial clustering of applications with noise |
| DCIS | Ductal carcinoma *in situ* |
| DL | Deep learning |
| DNA | Deoxyribonucleic acid |
| DNN | Deep neural network |
| ELM | Extreme learning machine |
| EV | Explained variance |
| FAD | Flavin adenine dinucleotide |
| Fisher-KPP | Fisher-Kolmogorov-Petrovsky-Piskunov |
| FLDV | Fiber laser Doppler vibrometer |

| | |
|---|---|
| FLIM | Fluorescence-lifetime imaging microscopy |
| FMS | Frequency modulation spectroscopy |
| FN | False negative |
| FOS | First-order statistics |
| FP | False positive |
| FPOU | Fokker-Planck equation describing the random Ornstein-Uhlenbeck process |
| FRET | Fluorescence resonance energy transfer |
| FS | Feature selection |
| FTIR | Fourier-transform infrared spectroscopy |
| GC–MS | Gas chromatography–mass spectrometry |
| GLCM | Gray-level co–occurrence matrix |
| HOG | Histogram of oriented gradients |
| HSI | Hyperspectral imaging |
| IC | Invasive carcinoma |
| ICA | Independent component analysis |
| IR | Infrared |
| KFCV | *K*-fold cross-validation |
| LC | Lung cancer |
| LDA | Linear discriminant analysis |
| LIF | Laser-induced fluorescence |
| LOD | Limit of detection |
| LOO | Leave-one-out |
| LOOCV | Leave-one-out cross-validation |
| LPAS | Laser photoacoustic spectroscopy |
| MCL | Markov cluster algorithm |
| MDP | Method of differential absorption |
| ME | Masking effect |
| ML | Machine learning |
| MNG | Multi-nodular goiter |
| MOS | Metal-oxide semiconductor |
| MPAS | Multipass absorption spectroscopy |
| MPM | Malignant pleural mesothelioma |
| MPM | Multiphoton microscopy |
| MRI | Magnetic resonance imaging |
| MS | Mass spectrometry |
| NDIR | Nondispersive infrared spectroscopy |
| NMF | Nonnegative matrix factorization |
| NN | Neural network |
| OAD | Optical acoustic detector |
| OLP | Oral lichen planus |
| OR–PAM | Optical-resolution photoacoustic microscopy |
| OSCC | Oral squamous cell carcinoma |

| | |
|---|---|
| PA | Photoacoustic imaging |
| PACT | Photoacoustic computed tomography |
| PAFC | Photoacoustic flow cytometry |
| PAS | Photoacoustic spectroscopy |
| PAT | Photoacoustic tomography |
| PC | Principal component |
| PCA | Principal component analysis |
| PCCA | Partial canonical correlation analysis |
| PCS | Physio-chemical spectrogram |
| PET | Positron emission tomography |
| PLS | Partial least squares |
| PLS-DA | Partial least squares discriminant analysis |
| $Q$ factor | Quality factor |
| QCL | Quantum cascade laser |
| QEPAS | Quartz-enhanced photoacoustic spectroscopy |
| RBF | Radial base function |
| RET | Resonance energy transfer |
| RNA | Ribonucleic acid |
| ROC | Receiver operator curve |
| ROI | Region of interest |
| RS | Raman scattering |
| RSCV | Random subsampling cross-validation |
| SE | Swamping effect |
| SERS | Surface-enhanced Raman scattering |
| SHG | Second-harmonic generation |
| SIFT–MS | Selected ion flow tube mass spectrometry |
| SIMCA | Soft independent modeling of class analogy |
| SNV | Standard normal variate |
| SOS | Second-order statistics |
| SPECT | Single-photon-emission computed tomography |
| SVM | Support vector machine |
| TDLAS | Tuned diode laser absorption spectroscopy |
| TDS | Time-domain spectroscopy |
| TEL | Total edge length |
| THz | Terahertz |
| TN | True negative |
| TP | True positive |
| TPAF | Two-photon autofluorescence |
| VOC | Volatile organic compound |
| WFT | Windowed Fourier transform |
| WMS | Wavelength modulation spectroscopy |