We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

6,900
Open access books available

186,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS

**BOOK
CITATION
INDEX**

INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Bioinformatics Solutions for Image Data Processing

Pietro Cinaglia, Luciano Caroprese,
Giuseppe Lucio Cascini, Francesco Dattola,
Pasquale Iaquinta, Miriam Iusi, Pierangelo Veltri and
Ester Zumpano

Additional information is available at the end of the chapter

Abstract

In recent years, the increasing use of medical devices has led to the generation of large amounts of data, including image data. Bioinformatics solutions provide an effective approach for image data processing in order to retrieve information of interest and to integrate several data sources for knowledge extraction; furthermore, images processing techniques support scientists and physicians in diagnosis and therapies. In addition, bioinformatics image analysis may be extended to support several scenarios, for instance, in cyber-security the biometric recognition systems are applied to unlock devices and restricted areas, as well as to access sensitive data. In medicine, computational platforms generate high amount of data from medical devices such as Computed Tomography (CT), and Magnetic Resonance Imaging (MRI); this chapter will survey on bioinformatics solutions and toolkits for medical imaging in order to suggest an overview of techniques and methods that can be applied for the imaging analysis in medicine.

Keywords: medical imaging, patterns, machine-learning, data-mining

1. Introduction

The image data processing is a relevant support for clinical diagnosis and sharing of health information, for instance, the correct interpretation of images may be crucial for early diseases detection. The use of sophisticated devices has greatly improved the acquisition of data at a very high resolution and faster rate, although the image interpretation process based on machine learning techniques is only recently taking hold. Bioinformatics tools allow to

retrieve information able to support a scientist during a diagnosis in order to detect efficiently abnormalities and to monitor their changes over time. In the last years, medical and biological images are quickly growing in terms of size and information content. The term "bioimage" concerns all images related to biological samples acquired using medical technologies such as the Computed Tomography (CT) or the Magnetic Resonance Imaging (MRI); briefly, CT is a technique based on ionizing radiation, for instance, X-rays are used to acquire the images; instead, MRI (or Nuclear-MRI) uses magnetic fields and Radio Frequencies (RF) to produce detailed pictures. An MRI Scanner measures the RF emitted by hydrogen atoms, the RF are directly related to the amount of energy previously provided to the atom and to its necessity to return at an equilibrium state after an excitation. These different approaches make that a CT scanner is better suited to detect cancer or cerebral hemorrhage, while brain tumor is more clearly visible using an MRI scanner. Bioinformatics techniques allow the developing of solutions which provide a support for images acquisition and analysis as well as the integration of biological datasets (e.g., gene data and ontologies) for patterns discovering and disorders detection. For instance, in neuroimaging an interesting branch of study concerns the correlation between brain regions and cognitive functions using several techniques such as the Functional-MRI (fMRI); the latter is based on the Blood Oxygenation Level Dependent (BOLD) signal which represents the increment of blood oxygenation in order to identify the brain region affected during activities (e.g., hands motion) and to reflect the local neuronal signaling [1]. This approach uses the BOLD signal in fMRI to study the activity of the hydrogen atoms that compose the water molecules contained in the brain [2].

BOLD signal in fMRI uses the hydrogen atoms that compose the water molecules contained in the brain to study its activity. An MRI Scanner measures the radio frequency emitted by hydrogen atoms after a first step during which this absorbs energy to the same radio frequency; this effect is related to the necessity of an atom to return to an equilibrium state after an excitation.

When a magnetic field is applied, the atoms absorb energy at a determinate radio frequency; for a clinical-standard fMRI technology at 1.5 Tesla this is approximately 64 MHz. BOLD techniques measure the changes in blood oxygenation resulting from the inhomogeneity of the magnetic field within each small volume of tissue that contains hydrogen atoms. Furthermore, the BOLD signal depends on several magnetic properties, such as Deoxy- and Oxy-hemoglobin. The first introduces an inhomogeneity into the magnetic field due to its paramagnetic properties, the latter is weakly diamagnetic. The variations in the concentrations of Deoxy- and Oxy-hemoglobin produce a decrement and an increment, respectively, in image intensity [3]. Many diseases may be studied using neuroimaging to analyze modifications and alterations in brain regions. Nevertheless, to perform detailed analysis, bioinformatics techniques and algorithms have to be developed. Recent studies focus their attention on the improvement of prediction performance in evaluation of diseases, for instance, in [4] authors recruited 120 schizophrenia patients and 120 healthy controls to compare four methods: Linear Discriminant Analysis (LDA), k-Nearest Neighbors (KNN), Gaussian process classifier (GPC) and Support Vector Machines (SVM). The results illustrate that using brain imaging is possible to study the brain areas as well as a set of specific cognitive functions and their related mental activities. Assuming that a physician needs to develop a model for

phenotype–genotype association in Alzheimer's disease (AD), an association study of longi-tudinal phenotypic markers to AD relevant SNPs could be conducted in order to perform a task-correlated longitudinal sparse regression between MRI and Micro Array technologies, respectively, for the images and the genomic information [5]. The study of bioimaging has met a large quantitative data from heterogeneous sources and the correlation among the data is a decisive step for knowledge extraction; thus, the latter allows a scientist to study novel solutions, and bioinformatics algorithms play a primary role to match heterogeneous sources, based on different models, in order to extract the information of interest.

2. Data formats in biomedical imaging

A file format is a solution able to organize a data inside a file using a specific model to allow the storing and the handling of its information content; building a standard for these pro-cesses is necessary to extend the compatibility to the highest number of devices and tools. The increasing adoption in the clinical practice of the 3D solutions, due also to the evolu-tion of technologies in medical imaging such as the Computed Tomography (CT) and the Magnetic Resonance Imaging (MRI), produced the affirmation of the Digital Imaging and Communication in Medicine (DICOM) standard. It is supported from almost all independent manufacturers of imaging equipment that need to guarantee an easy and fast solution for data storage and exchange; furthermore, DICOM is adaptable and extensible to accommodate the requirements for an imaging technology. Briefly, this standard is able to handle the storage, viewing, and transmission of medical images, as well as data sharing in clinical research. In the DICOM format each image is stored in a separate instance (comparable to an 'object' in computer language) which includes metadata and images data. The first contains several information concerning the patient, the manufacturer and technology used to acquire the medical images (modality included), the physician, imaging parameters (e.g., slice thickness and spacing, image orientation and frames features); its size is small as all attributes are repre-sented using raw text in tags-based format. Instead, an image (or slice) concerns the numerical matrices containing a gray-scale value for each related pixel; thus, the volume rendering con-sisting of the union of all slices in the properly order. In the DICOM standard a redundancy of information is created due to the fact that metadata are repeated in a header generated for all images; this is due to the need to guarantee the independence of each slice from the others both during viewing and transmission. The flexibility of tags information contained in the header produces different information among the scanner manufacturers based on DICOM; this causes incompatibilities and critical failures in the tools that use it.

These disadvantages have led to the implementation of novel standards such as: Neuroimaging Informatics Technology Initiative (NIfTI) standard, Analyze or Minc.

The first version of NIfTI was developed as a set of extensions starting from the Analyze (release 7.5) format in order to extend the support to multidimensional data (e.g., for volumes representation) and to allow the handling of MRI data. Analyze consists of a header (a .hdr file) that contains information (e.g., dimensions and identification), and in voxel raw data (a .img file) for image representation. On the other hand, it does not support the unambiguously

establish of the image orientation, as well as the unsigned 16 bits data so its adoption is not widespread [6]. NIfTI is a minimalist format adopted in neuroimaging research consisting, as in Analyze, of a header and an uncompressed image data stored, respectively in a '.hdr' and '.img' file; furthermore, it is able to unify both elements in a single '.nii' file that have the first 348 bytes reserved for the header. In a NIfTI file, an image may be represented using up to seven dimensions of which three are for the space, one for the time and three for the diffusion gradient direction [7]; this feature allowed its growing adoption in neuroscience.

A conversion from DICOM to NIfTI is generally possible using dedicated solutions like 'dcm2nii' (http://www.nitrc.org/projects/dcm2nii/). In [8] a test-scenario based on a cloud implementation of dcm2nii is proposed; it provides the analysis of 812 volumes, stored in 991,000 DICOM files and related to 41 subjects undergoing clinical CT scanners for a neuro-logical study. Experimental results illustrate that the conversion process is extremely time-consuming; in details, authors report that it may need of a computing grid which uses up to 156 cores to process about 52 volumes/minute.

Ultimately, Minc format was developed in 1992 by the Montreal Neurological Institute (MNI); the last release (Minc2) is based on Hierarchical Data Format version 5 (HDF5) and supports large data file as well as a set of tools for DICOM/NIfTI conversion. Each MINC file concerns a multidimensional dataset with all related metadata. It is generally used only in projects developed directly by the MNI Brain Imaging Center for the releasing of library and tools [9].

To date the most widespread formats among the mentioned remain the DICOM and the NIfTI standards which are also supported by most medical device manufacturers.

3. Techniques and methods in medical imaging

Techniques and methods for image analysis are generally based on Machine Learning (ML) and Artificial Intelligence (AI); these allow to efficiently derive relevant information from heterogeneous data such as phenotypes, morphologic features, as well as patterns. In bio-informatics, a pattern may be related to diseases or alterations and it is generally used to associate an image domain with a specific alteration. Briefly, it consists of a sequence of items modeled through a mathematical function able to represent a specific structure, for instance, the patterns can be used to represent a part of an image, or to better say a part of the matrix that contains the value of each pixel/voxel. Assuming that we want to use this approach in neurosciences, a pattern could be used to discover the alterations that affect the normal func-tions of the brain; the issue could be solved using an algorithm for pattern recognition which is able to integrate information across several sources. In this context, the techniques based on Machine Learning (ML) and Data-Mining approaches are the most used for the developing of bioinformatics tools. Generally, an ML algorithm may be able to identify morphologies by basing its analysis on a training-set of known patterns (supervised learning); alternatively, it could be able to analyze directly an input dataset in order to identify novel pattern using predictions not related to training-sets previously imported but acquired during the analysis (unsupervised learning). Furthermore, a pipeline based on patterns recognition is considered

a valid approach for diagnosis and prognosis, as well as to perform studies using several medical imaging from heterogeneous sources. Assuming to conduct a study for the analysis of alterations in a specific brain's region, a pattern recognition solution may be implemented; its pipeline could consist of three main steps: the first able to acquire raw data in a structured model, the second to select the features of interest and the third to perform an algorithm that allow the analysis. The final result will be represented by a set of data with a reduced dimensionality for the space (e.g., represented using the voxel) [10] that will therefore allow to discriminate quickly the patients from the healthy controls using a classifier based, for example, on a cross-validation approach.

In medical image processing, useful approaches are based on the Neural Network (NN). An NN is conceptually inspired by the human brain system and consists of: (i) an input layer, (ii) an output layer and (iii) one or more hidden layers; the number of layers is related to the level of abstraction, as well as to the reliability of the prediction (e.g., a deep neural network hierarchically extends to over 1000 layers). For instance, the NN can be used to extract ROIs [11] using data acquired from mammograms [12]. A powerful set of techniques for learning in NN is the Deep Learning (DL); it is a growing trend for data analysis and medical imaging, becoming one of the breakthrough technologies [13]. A Deep Neural Network performs all possible mappings, importing a dataset built in a previous training step, in order to formulate predictions for the unknown cases in input. Several algorithms based on machine learning are based on DL approaches [14], such as: Convolutional Neural Network (CNN), Deep Neural Network (DNN), Deep Belief Network (DBN), Deep Botlzmann Machine (DBM) and Recurrent Neural Network (RNN); the description and key points are shown in **Table 1**.

Assuming that we want to develop an algorithm to discriminate a specific area (e.g., representing a cancer) from an entire biomedical acquisition, an algorithm for handling of 3D volume and segmentation must be implemented. For instance, UNet [15], VNet and DeepMedic [16], and its improved version DMRes [17], are able to perform a fast and precise segmentation using a CNN architecture for fast and precise segmentation.

In addition, the study of medical sciences (e.g., neurosciences) has met a large quantitative of data and the bioinformatics tools and algorithms play a primary role in reference to knowledge extractions, these allow to perform data-integration techniques obtaining an interdisciplinary view matching heterogeneous sources with different models in order to extract information of interest. In [18] the authors develop BioMediator System to provide a data integration across several biomedical domains and data types in order to allow a biologist to retrieve molecular and genomic information using a browser engine based on head constraints and global path constraints. Assuming that we want to base an application on a mediation approach for a data integration in neurosciences, a first step is related to information retrieving, for instance, the Human Database (HID, http://nbirn.net/research/function/hid.shtm) and the eXtensible Neuroimaging Archive Toolkit (XNAT, http://www.xnat.org) can be used. Subsequently, a query engine for virtual data-integration can be developed. It could implement a OGSA-DAI/DQP architecture, respectively: Open Grid Services Architecture (OGSA) Distributed Access and Integration (DAI) and OGSA Distributed Query Processing (DQP); this solution offers a streaming dataflow workflow evaluation engine, and on a distributed query evaluation engine [19].

Type	Description
Convolutional Neural Network (CNN)	It is based on a neurobiological model that imports 2D inputs in order to produce a 3D output volume for the neuronal activity. Generally, it requires large dataset, as well as many layers.
Deep Neural Network (DPN)	DPN is used in non-linear computations using hidden layers (≥ 2) for classification and regression. It is a solution typically used for many areas; its general-purpose approach is very slow not being optimized for a specific context.
Deep Belief Network (DBN)	It uses undirected connections for top two layers in order to allow supervised and unsupervised training for the network. For the network initialization perform a layer-by-layer greedy learning; this approach is computationally expensive.
Deep Botlzmann Machine (DBM)	It is based on undirected links among the layers using a stochastic maximum likelihood algorithm. Top-Down feedback are supported in order to perform a bust inference.
Recurrent Neural Network (RNN)	It is a Neural Network used for stream data analysis when the output is related to previous computation; during the recursions all weight are shared for all layers. The key points are reported below: • Time dependencies modeling • Events persistence

Table 1. Deep learning approaches: Key-points and description.

4. Imaging processing solutions

Bioinformatics allows a scientist to extract knowledge from a large set of information consisting of heterogeneous resources, as well as to handle the information of interest and more generally the methods for data storage and retrieval.

For instance, to acquire information of interest from a biomedical image often requires the development of a set of instructions in accordance to the research project; in this regard, a custom algorithm is able to extract knowledge from an input exactly as required by the scientist, and generally more efficiently than a generic commercial or free solution being designed on specific requirements identified in the case study.

Assuming that a researcher wants to implement a solution able to predict structure in accordance with a set of known patterns defined in a "training" phase, a machine-learning approach is perhaps the best choice.

Briefly, the possibilities are mainly two: (i) the implementation of an algorithm "from scratch" or (ii) the importing of existing toolkits (e.g., framework and libraries) within its own algorithm. A development "from scratch" is particularly expensive and time consuming compared to the use of libraries/framework "ready to use". Therefore, it is important to evaluate the availability of existing toolkits before starting the development, as well as the efficiency of those chosen in accordance with the intended aim.

Supposing to use Python as programming-language, Scikit-Learn (http://scikit-learn.org/stable/) is a possible and useful solution for what has been said. It is a general-purpose machine learning library (written in Python) to solve complex data analysis tasks. It uses NumPy and SciPy, respectively, for: (i) storage and manipulation of multi-dimensional array, as well as (ii) for the implementation of data structure and mathematical functions [20]. Scikit-Learn contains also a large set of statistical learning algorithms. Assuming that we want to define a score to identify the variability captured by the model for each voxel then the BOLD signal should be quantified through a regression function. For instance, this approach allows a scientist to analyze fMRI data in order to encode brain activity and decode the related medical images [21].

Alternatively, a useful solution to implement algorithms that use machine-learning techniques for data analysis could be Tensorflow (http://www.tensorflow.org). It is a popular library developed by Google Inc. to implement and execute large-scale artificial neural networks based on typed and multi-dimensional array (named Tensor); a Tensor is a generalization of arrays and matrices with higher size. Its approach is based on nodes defined during a training task, in each training iteration new links are established.

Furthermore, Tensorflow's API allows the implementation of tools able to train and to test neural networks using computational graphs built in Python or C++. Assuming that a scientist wants to develop a machine-learning interface than the TensorFlow's API is a good starting point, for instance, the DICOM support may be extended in order to implement a classifier for medical imaging that uses bioimage datasets as input for tensors [22]. Therefore, this solution could therefore represent a potential tool useful for diagnostic in the clinical practice. Assuming that we want to improve the accuracy of a disease diagnosis, a deep learning solution may be implemented. For instance, the input for the model could be represented by Alzheimer's Disease Neuroimaging Initiative (ADNI) data acquired with Magnetic Resonance Imaging (MRI) or Positron Emission Tomography (PET) techniques. To evaluate the model a pre-processing step is performed in order to (i) correct the intensity inhomogeneity, (ii) normalize data into a template space, (iii) define maps for the gray matter tissue and (iv) improve the signal-to-noise ratio (e.g., using a Gaussian kernel). An interesting requirement could be represented by the possibility of capturing the highly nonlinear relationships between Input and Output; the Convolutional Neural Networks (CNNs) are a candidate solution for this aim (as previously defined: *a CNN is based on a neurobiological model that imports 2D inputs in order to produce a 3D output volume for the neuronal activity. Generally, it requires large dataset, as well as many layers*). In [23], a similar approach is developed: the authors focused on evaluating of a CNN model for image data completion. The solution is based mainly on two steps: (i) complete the missing data using CNN (a comparison with other methods based on KNN and Zero is evaluated); (ii) estimate the performance using a l2-norm regularized logistic regression classifiers on the output. In the experiments, the data are randomly divided into two sub-sets, respectively, one-third for the tests and two-thirds for the training step; in the latter, a CNN model is built in order to retrieve the nonlinear relationship between different data modalities in the network. In accordance with the authors the results show that the predicted data (obtained from PET images) can potentially be used to improve the accuracy of disease diagnosis.

4.1. Toolkits in medical imaging

More toolkits are available from several program languages and platforms in bioinformatics, and generally, these are based on Machine Learning (ML) solutions (or similar approaches). A standard which is able to produce a comparison among toolkits in ML is not defined, as each tool is often designed for determinate needs related to different problems, and its approach is therefore to be considered context-specific [24].

A list of useful toolkits is shown below and summarized in **Table 2**; only objective criteria (e.g., programming-language and features) are reported so that everyone may correlate a toolkit in reference to their needs.

- ODTbrain is a Python library that implements a back-propagation algorithm for dense diffraction tomography in 3D [25]. The three-dimensional (3D) refractive index distribution of a unique cell allows to describe its inner structure in a marker-free manner. The term dense, full-view tomographic data set denotes a set of images of a cell acquired for multiple rotational positions, densely distributed from 0 to 360 degrees. The projection tomography, based on the inversion of the Radon transform, is generally used to perform the reconstruction and its quality is greatly improved when first order scattering is taken into account. This advanced reconstruction technique is called diffraction tomography. The first implementation of diffraction tomography has been proposed in ODTbrain. The algorithm is an extension to optical projection tomography that takes into account diffraction of light due to the refractive index of the sample. In ODTbrain the reconstruction process is divided into three main steps: filtering, reconstruction and object data construction. ODTbrain is able to reconstruct 3D refractive index maps from projections of biological or artificial phase objects; the algorithm is validated performing the analysis on a simulated dataset and subsequently authors have compared results with the reconstruction qualities of Optical Diffraction Tomography (ODT) and Optical Projection Tomography (OPT).

Name	Language	Main methods
jicbioimage	Python	Microscopy data, view and explore data, generate reproducible analyses.
Scikit-Learn	Python	Classification, regression, clustering, model selection, preprocessing.
ODTbrain	Python	Back-propagation algorithm for dense diffraction tomography in 3D.
CP-CHARM	Python	Image automated classification (optimization is not required).
SCIFIO	Java	Handling of scientific images.
NIF	(framework, web-based)	Queries support for NIF.
WHIDE	(web-based)	H2SOM clustering, imaging analysis (space and collocation).
GALA	Python	Data acquisition and analysis in neurophysiology.
ACQ4	Python	Experiments automation.

Table 2. Toolkits for medical imaging.

- CP-CHARM [26] is a user-friendly image-based classification algorithm. It is inspired by WND-CHARM. The latter is a multi-purpose image classification algorithm that can be applied without optimization or modifying the starting data; features are computed on the whole image and no segmentation is required [27]. Using the CP-CHARM algorithm a user is able to extract several morphological features from an image without first being segmented; furthermore, in order to be suitable and accessible to all the biological research community, even with few expertise, CP-CHARM relies on CellProfiler an open source image analysis software for quantitative analysis of biological images [28]. The proposed method has been demonstrated to perform well on a wide range of bioimage classification problems. The proposed method has been validated firstly by showing that it could achieve performance similar to those of WND-CHARM. Then the algorithm has been used on several kinds of biological datasets, for example, data freely available from the Broad Bioimage Benchmark Collection (BBBC) [29] and tissue images from the Human Protein Atlas (HPA) [30]. CP-CHARM has been demonstrated to perform well on a wide range of bioimage classification problems.

- SCIFIO is a flexible framework for SCientific Image Format Input and Output. In other words, it is a library for reading and writing N-dimensional image data. SCIFIO is an open source plugin for the SciJava framework that offers support for handling of scientific images. SCIFIO defines a common pattern for image format construction and can be easily extended—from custom formats to new metadata schema. It is developed by the ImageJ development team at the Laboratory for Optical and Computational Instrumentation (LOCI) at the University of Wisconsin-Madison. The Open Microscopy Environment's Bio-Formats library provides the ability to convert many proprietary image formats to a common OME-TIFF format, using the OME-XML schema. This allows scientists to freely share image data without being restricted by proprietary format barriers. It is part of the SciJavasoftware stack and in use by several projects including ImageJ2, ImgLib2, and the Insight Toolkit (ITK). One of the main features of SCIFIO concerns the support of multiple domain-specific formats within a unified environment [31].

- The Neuroscience Information Framework (NIF) is a framework that promotes the integrating access to Web-Based neuroscience resources [32]. It is supported from the Institutes and Centers forming the NIH Blueprint for Neuroscience Research. The framework is based on an Open Source design and offers dynamic and web-accessible resources focused on neuroscience that are described using an integrated terminology, also it is able to support concept-based queries as well as the integration of neuroscience information with complementary areas of biomedicine.

- Web-based Hyperbolic Image Data Explorer (WHIDE) combines more features related to principles of machine learning and scientific-information in order to analyze the aspects (space and collocation) of Toponome Imaging System (TIS) images. WHIDE uses Hierarchical Hyperbolic SOM (H2SOM) clustering to resolve non-linear features and dynamic interactive manipulation of the colors, as well as to organize the clusters in a hierarchical structure. Authors tested the tool for TIS analysis but is not excluded that it is applicable to other MBI data [33].

- Graph-based Active Learning of Agglomeration (GALA) is an algorithm, implemented in python language, for image segmentation. GALA belongs to a class of segmentation algorithms called agglomerative algorithms, in which segments are formed by merging smaller segments. It works by repeatedly consulting a gold standard segmentation (prepared by human annotators) as it agglomerates sub-segments according to its current best guess. More specifically, GALA accumulates a training dataset used to fit a classifier that guides the subsequent agglomeration decisions. GALA includes several scientific Python libraries: numpy, scipy and others, to perform segmentation analysis; also, it implements a solution based on machine-learning approach [34].

- ACQ4 is a modular software for data acquisition and analysis in neurophysiology research. It is available for download at http://www.acq4.org. ACQ4 integrates the task of acquiring, managing and analyzing experimental data. It is developed as general-purpose tools with the main aim to combine traditional electrophysiology, photostimulation and imaging for experiments automation. The system is highly modular and therefore it is quite simple to add new functionalities.

In [35] authors present several use-cases for ACQ4 to illustrate its functionalities reported as a set of experiments that are possible using this tool; below are briefly listed: (i) Multiphoton calcium imaging during whisker deflection; (ii) Laser scanning photostimulation; (iii) In vitro patch clamp with drug perfusion and (iv) In vivo recording during an operant conditioning task. ACQ4 uses free and open-source tools such as Python, NumPy/SciPy for numerical computation, PyQt for the user interface and PyQtGraph for scientific graphics. ACQ4.

- jicbioimage is a tool implemented in python language for automated and reproducible bioimage analysis; jicbioimage has been used on over 15 internal projects at various stages of the publication pipeline [36]. Using jicbioimaget an user is able to (i) read bioimage data in several format, for this aim the features of Python-BioFormat are imported by authors; (ii) transform and segment images using methods based on numpy, scipy, and scikit-image and (iii) examine the versions for an experiment: briefly, the versions for an initial image are stored by jicbioimage during all transformations to help a scientist to understands the steps performed by the tool and its sub-processes.

- eIMES 3D (standing for Evolution Imaging System 3D for Mobile) is a system that supports image reconstruction with dedicated features for the mobile environments [37, 38]. Using eIMES 3D, a cancer network data infrastructure can be defined and implemented in order to integrate information regarding rare and complex diseases. Furthermore, it provides a hardware infrastructure to connect multiple devices, as well as to create workstations (WorkSpaces) with independent and asynchronous features for information retrieving in order to acquire 3D images from a central database. Briefly, eIMES 3D allows:

  - full control and management of data and imaging by means of artificial intelligence algorithms;

  - advanced stereoscopic 3D visualization by using theWebGL innovative technology;

○ sharing of medical data;

○ distribution of 3D imaging on different output devices;

○ query the system through a search of the various case studies.

The AI algorithms allow to provide a new layer of information by applying set of rules fixed by international protocols, or by expert of the domain. These conceptually use known information and a set of production rules to derive new information or to change beliefs as a result of new knowledge. The logical formalization of AI algorithm is defined using a set of logical rules in order to extract new layers deriving knowledge (deductive process). Furthermore, the AI algorithms can apply an abductive process: using novel observation, that modifies the existing protocol, the information can be updated with the new specifications.

5. Conclusion

The fast growth of medical imaging may be seen monitoring the increase of available solutions and the interest in its research field, as well as in clinical practice. The increasing adoption in the clinical practice of the 3D solutions, due also to the evolution of technologies in medical imaging, such as the Computed Tomography and the Magnetic Resonance Imaging, produced a large amount of data. Knowledge extraction from medical images is still a complex task; this chapter recalls several images techniques and approaches used in bioinformatics, also describing some useful toolkits for the development of custom solutions.

Author details

Pietro Cinaglia[1], Luciano Caroprese[2], Giuseppe Lucio Cascini[1], Francesco Dattola[3], Pasquale Iaquinta[3], Miriam Iusi[3], Pierangelo Veltri[1] and Ester Zumpano[2]*

*Address all correspondence to: e.zumpano@deis.unical.it

1 Department of Surgical and Medical Sciences, Magna Graecia University of Catanzaro, Italy

2 Department of Informatics, Modeling, Electronics and System Engineering, University of Calabria, Italy

3 e way Enterprise Business Solutions Cosenza, Italy

References

[1] Matthews P, Jezzard P. Functional magnetic resonance imaging. Journal of Neurology, Neurosurgery, and Psychiatry. 2004;**75**(1):6-12

[2]   Tang Y, Meng L, Wan CM, et al. Identifying the presence of Parkinson's disease using lowfrequency fluctuations in BOLD signals. Neuroscience Letters. 2017;**645**:1-6

[3]   Heeger DJ, Ress D. What does fMRI tell us about neuronal activity? Nature Reviews Neuroscience. 2002;**3**(2):142-151

[4]   Li Z, Wang Y, Quan W, Wu T, Lv B. Evaluation of different classification methods for the diagnosis of schizophrenia based on functional near-infrared spectroscopy. Journal of Neuroscience Methods. 2015;**241**:101-110

[5]   Wang H, Nie F, Huang H, et al. From phenotype to genotype: An association study of longitudinal phenotypic markers to Alzheimer's disease relevant SNPs. Bioinformatics. 2012;**28**(18):i619-i625

[6]   Larobina M, Murino L. Medical image file formats. Journal of Digital Imaging. 2014;**27**(2): 200-206

[7]   Li X, Morgan PS, Ashburner J, Smith J, Rorden C. The first step for neuroimaging data analysis: DICOM to NIfTI conversion. Journal of Neuroscience Methods. 2016;**264**:47-56

[8]   Bao S, Plassard AJ, Landman BA, Gokhale A. Cloud engineering principles and technology enablers for medical image processing-as-a-Service. Proceeding of IEEE International Conference on Cloud Engineering. 2017;**2017**:127-137

[9]   Vincent RD, Neelin P, Khalili-mahani N, et al. MINC 2.0: A flexible format for multi-modal Images. Frontiers in Neuroinformatics. 2016;**10**:35

[10] Haller S, Lovblad KO, Giannakopoulos P, Van de ville D. Multivariate pattern recognition for diagnosis and prognosis in clinical neuroimaging: State of the art, current challenges and future trends. Brain Topography. 2014;**27**(3):329-337

[11] Sonka M, Hlavac V, Boyle R. Image Processing, Analysis and Machine Vision. 1995. 10.1007/978-1-4899-3216-7

[12] Sahiner B, Chan HP, Petrick N, et al. Classification of mass and normal breast tissue: A convolution neural network classifier with spatial domain and texture images. IEEE Transactions on Medical Imaging. 1996;**15**(5):598-610

[13] MIT Technology Review. 2013 [Online]. Available https://www.technologyreview. com/s/513696/deep-learning

[14] Ravi D, Wong C, Deligianni F, et al. Deep Learning for Health Informatics. IEEE Journal of Biomedical and Health Informatics. 2017;**21**(1):4-21

[15] Ourselin S, Joskowicz L, Sabuncu MR, et al. Medical Image Computing and Computer-Assisted Intervention – MICCAI. 2016, 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings. Springer; 2016

[16] Kamnitsas K, Ledig C, Newcombe VFJ, Simpson JP, Kane AD, Menon DK, Rueckert D, Glocker B. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Medical Image Analysis. 2017;**36**(1):61-78

[17] Kamnitsas K, Ferrante E, Parisot S, Ledig C, Nori AV, Criminisi A, Rueckert D, Glocker B. DeepMedic for brain tumor segmentation. In: Crimi A, Menze B, Maier O, Reyes M, Winzeck S, Handels H, editors. BrainLes 2016. LNCS, vol. 10154, pp. 138-149

[18] Wang K, Tarczy-Hornoch P, Shaker R, Mork P, Brinkley JF. BioMediator data integration: beyond genomics to neuroscience data. In: American Medical Informatics Association Annual Symposium Proceedings; 2005. pp. 779-783

[19] Wang L, Alpert KI, Calhoun VD, Cobia DJ, Keator DB, King MD, Kogan A, Landis D, Tallis M, Turner MD, Potkin SG, Turner JA, Ambite JL. SchizConnect: Mediating neuro-imaging databases on schizophrenia and related disorders for large-scale integration. Neuroimage. Jan 2016;**124**:1155-1167

[20] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research. 2011;**12**:2825-2830

[21] Abraham A, Pedregosa F, Eickenberg M, et al. Machine learning for neuroimaging with scikit-learn. Frontiers in Neuroinformatics. 2014;**8**:14

[22] Zhang YC, Kagen AC. Machine Learning Interface for Medical Image Analysis. Journal of Digital Imaging. 2017;**30**(5):615-621

[23] Li R, Zhang W, Suk HI, Wang L, Li J, Shen D, Ji S. Deep learning based imaging data completion for improved brain disease diagnosis. Medical Image Computing and Computer-Assisted Intervention. 2014;**17**(Pt 3):305-312

[24] Erickson BJ, Korfiatis P, Akkus Z, Kline T, Philbrick K. Toolkits and Libraries for Deep Learning. Journal of Digital Imaging. 2017;**30**(4):400-405

[25] Muller P, Schurmann M and Guck J. ODTbrain: a Python library for full-view, dense diffraction tomography. BMC Bioinformatics. 2015;**16**:367:1-367:9

[26] Uhlmann V, Singh S, Carpenter AE. CP-CHARM: segmentation-free image classification made accessible. BMC Bioinformatics. 2016;**17**:51

[27] Orlov N, Shamir L, Macura T, Johnston J, Eckley DM, Goldberg IG. WND-CHARM: Multi-purpose image classification using compound image transforms. Pattern Recognition Letters. 2008;**29**(11):1684-1693

[28] Dao D, Fraser AN, Hung J, Ljosa V, Singh S, Carpenter AE. Cell Profiler Analyst: interactive data exploration, analysis and classification of large biological image sets. Bioinformatics. 2016

[29] Ljosa V, Sokolnicki KL, Carpenter AE. Annotated high-throughput microscopy image sets for validation. Nature Methods. 2012;**9**(7):637

[30] Uhlen M, Oksvold P, Fagerberg L, Lundberg E, Jonasson K, Forsberg M, et al. Towards a knowledge-based human protein atlas. Nature Biotechnology. 2010;**28**(12):1248-1250

[31] Hiner MC, Rueden CT, Eliceiri KW. SCIFIO: An extensible framework to support scientific image formats. BMC Bioinformatics. 2016;**17**(1):521

[32] Gardner D, Akil H, Ascoli GA, et al. The neuroscience information framework: a data and knowledge environment for neuroscience. Neuroinformatics. 2008;**6**(3):149-160

[33] Kölling J, Langenkämper D, Abouna S, Khan M, Nattkemper TW. WHIDE–a web tool for visual data mining colocation patterns in multivariate bioimages. Bioinformatics. 2012;**28**(8):1143-1150

[34] Nunez-iglesias J, Kennedy R, Plaza SM, Chakraborty A, Katz WT. Graph-based active learning of agglomeration (GALA): A Python library to segment 2D and 3D neuroimages. Frontiers in Neuroinformatics. 2014;**8**:34

[35] Campagnola L, Kratz MB, Manis PB. ACQ4: An open-source software platform for data acquisition and analysis in neurophysiology research. Frontiers in Neuroinformatics. 2014;**8**:3

[36] Olsson TS, Hartley M. jicbioimage: A tool for automated and reproducible bioimage analysis. PeerJ. 2016;**4**:e2674

[37] Iaquinta P, Iusi M, Caroprese L, Turano S, Palazzo S, Dattola F, Pellegrino I, Veltri P, Zumpano E. eIMES 3D: An innovative medical images analysis tool to support diagnostic and surgical intervention. Procedia Computer Science. 2017;**110**:459-464

[38] Turano S, Biamonte R, Zumpano E, Veltri P, Confessore G, Dattola F, Pellegrino I, Iaquinta P, Iusi M, De Stefano R, Romeo F, Palazzo S. ReCaTuR, Rare Cancer Network Calabria, Implementing a software system based on showing of 3D stereoscopic imaging data. Annals of Oncology. 2016;**27**(suppl 4):iv100-iv100