Health Informatics Vision: From Data via Information to Knowledge
J. Mantas et al. (Eds.)
IOS Press, 2019

doi:10.3233/SHTI190081

A Platform for Collection and Analysis of Image Data on Stroke

Daria DOLOTOVA[a,1], Victoria DONITOVA[b], Ivan ARHIPOV[a], Faat SHARIFULLIN[c], Tatiana ZAGRIAZKINA[d], Boris KOBRINSKII[b,d] and Andrey GAVRILOV[a,e]
[a] *Gammamed-Soft, Ltd, Moscow, Russia*
[b] *Federal Research Center "Computer Science and Control" of RAS, Moscow, Russia*
[c] *N.V.Sklifosovskiy Scientific Research Institute of Emergency Care, Moscow, Russia*
[d] *Pirogov Russian National Research Medical University, Moscow, Russia*
[e] *Lomonosov Moscow State University, Moscow, Russia*

Abstract. Identifying imaging biomarkers (IBs) of stroke remains a priority in neurodiagnostics. There is a number of different methods for image analysis and learning rules applicable in this field, but all of them require large arrays of DICOM images and clinical data. In order to amass such dataset, we have designed a platform for systematic collection of clinical data and medical images in different modalities. The platform provides easy-to-use tools to create formalized radiology reports, contour and tag the regions of interest (ROIs) on the DICOM images, and extract radiomics data. Subsequent analysis of the obtained data will allow identifying the most relevant IBs that predict clinical outcome and possible complications. The results of the analysis will be used to develop predictive algorithms for stroke diagnostics.

Keywords. Stroke, imaging biomarkers, big data, machine learning, decision support system

1. Introduction

One of the most relevant directions in modern medicine is the identification of imaging biomarkers (IBs), or characteristics of physiological and pathological processes that can be assessed through analysis of medical images [1]. Research and adoption of IBs to clinical practice could expedite the analysis of medical images, help in the search for predictors of acute conditions, improve the diagnostics process, and, as a result, the quality of healthcare. Process of IBs identification includes following stages [2]: 1) data selection; 2) segmentation of regions of interest (ROIs); 3) feature extraction; 4) statistical analysis and modeling.

Identifying IBs would be particularly helpful in diagnostics of stroke, one of the leading causes of mortality and morbidity. The first priority in stroke diagnostics is evaluation of the stroke type in order to determine the appropriate treatment. The main difficulty in the research of the stroke IBs is the absence of representative datasets containing both images and clinical data. There are some datasets for image analysis in MRI or CT perfusion that are openly available to the research community [3,4]. A lot of

[1] Corresponding Author, Daria Dolotova; E-mail: dariadolotova@gmail.com.

scientists proposed mathematical models based on these datasets, international registries, or their local databases [5]. But the first-choice modality for primary stroke diagnostics in Russia is nonenhanced CT. In view of a limited number of such datasets, the aim of this project was to develop a platform that allows gathering of comprehensive information on stroke patients, both clinical data and DICOM images.

2. The Platform Design

The developed platform contains a package of client-server applications for Microsoft Windows: medical documentation module, DICOM images module, module for extraction of image characteristics. Data is stored in a Microsoft SQL database, and DICOM images are stored separately in a networked storage. DICOM images module supports importing images from portable devices as well as connecting to the healthcare facilities' PACS with the anonymization feature for the stored data. Connection to the facilities' archives is realized through DICOM Q/R and DICOM-Retrieve protocols. The platform also includes functionality for integration with electronic health record system to optimize the process of data entering.

2.1. Medical Documentation Module: Gathering Clinical Data

For every patient the following clinical data was entered into the platform:

- Clinical diagnosis (in accordance with the classification adopted in Russia, one of the following groups was assigned: transient ischemic attack, ischemic stroke, haemorrhagic stroke)
- Risk factors (arterial hypertension, diabetes mellitus, cigarette smoking, obesity, alcohol consumption, and others)
- Clinical case data (outcome, length of hospital confinement, treatment method, scores for different neurological scales: NIHSS, mRS, GCS)

Formalized radiology reports of: cerebral CT without contrast, cerebral CT angiography and cerebral CT with contrast were also included. Special screen forms with formal fields were developed to simplify the entering of clinical information into the system (Figure 1).

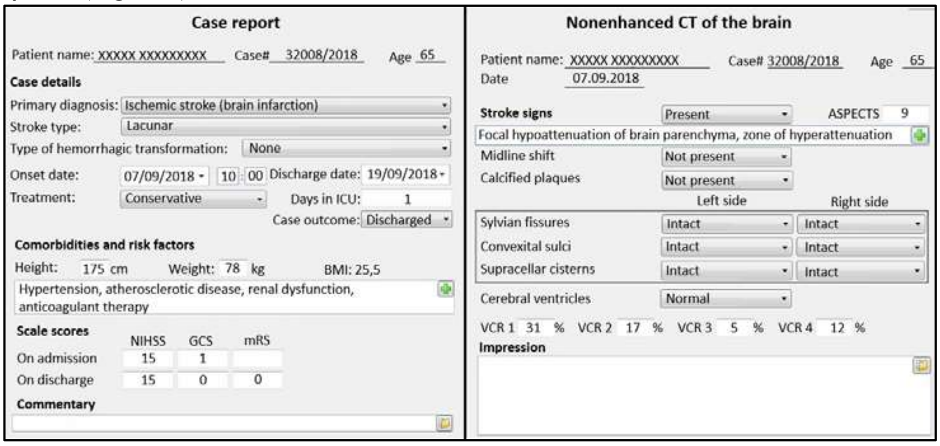

Figure 1. The user interface for entering clinical data and radiology report (nonenhanced brain CT).

2.2. DICOM images module: Segmentation and Tagging

A tool for contouring and tagging the ROIs was created based on a free hand technology that allows users to easily contour pathological formations and tag them accordingly. In order to enhance the efficiency of contouring an algorithm that fills in the contour was developed. The algorithm works by interpolation of contours made manually by the radiologists on separate slices of the CT. As a result of the finished contour a 3D ROI reconstruction is created (Figure 2), and the range of mathematical parameters is calculated.

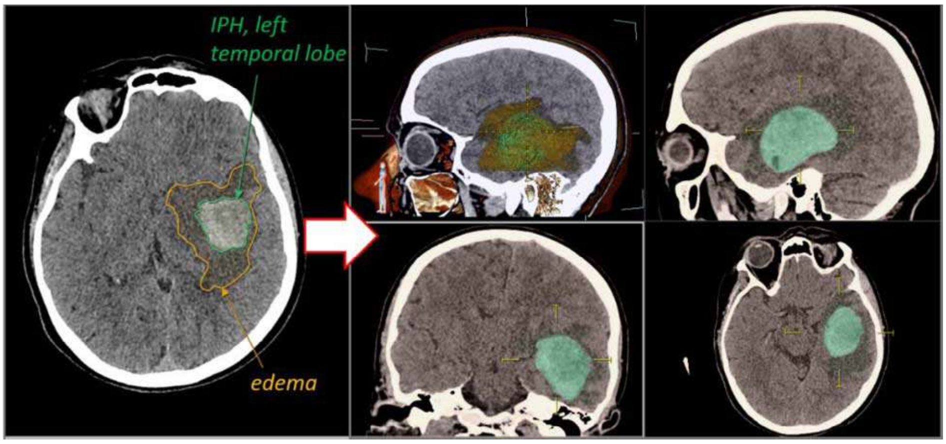

Figure 2. An example of image with segmented and tagged areas (ROI) of intra-cerebral haemorrhage and edema.

A thesaurus of tags for the ROIs was developed with the participation of the radiologists. The localization of a lesion, vascular zone in case of watershed lesions and other characteristics of a lesion were registered by tagging. Types of lesions included ischemic and hemorrhagic stroke (with intraparenchymal, intraventricular, and subarachnoid subtypes). There were also additional tags for areas of edema, lacune, aneurysm and others. Operational comfort in the process of tagging was ensured by the means of trigram search.

2.3. Module for Feature Extraction

The module for extraction of image characteristics was built by means of an open source library PyRadiomics [6]. Features describing ROIs intensity, morphology, spatial and histogram distribution can be calculated and exported to analytic packages in .csv format.

3. The First Implementation In Healthcare Facility

The platform was installed in N.V. Sklifosovskiy Scientific Research Institute of Emergency Care. So far, we have collected data on a hundred stroke patients with 170 CT reports with corresponding DICOM images that were uploaded from the Institute's PACS. Three radiologists reviewed, contoured and tagged the ROIs on the CT images. In total, there are about 20000 contours, and for all them different features will be calculated in the follow-up process.

At the beginning of the implementation numerous peculiarities related to data entry and clinical terminology were revealed. Additional algorithms tracking outliers, missing values, and tags incompatibility were developed to minimize the amount of data entry errors.

This work will be continued by adding more cases, including control cases of patients that have known risk factors for stroke, but have not developed any clinical signs of the stroke yet. The resulting dataset will be anonymized and published for open use.

4. Conclusion

During this research a platform for collection and analysis of stroke images was developed. Currently its functionality covers almost all stages of workflow of IBs identification except for the stage of feature analysis and modeling. In comparison with concurrent projects the developed platform has several advantages. Most of existing software programs do not support integration of clinical data in conjunction with imaging data and calculated parameters. Many of such programs have a narrow field of application. Some are designed only for contouring of ROIs, some for calculation of ROI parameters, thus prompting the researchers to combine multiple software products that can lead to compatibility issues. There is also an option to assess the parameters over time, if there are several CT protocols for the same case.

The project's long-term goal is to identify predictive IBs for stroke in order to assess and calculate individual stroke risk levels. A health management system that estimates individual risk of stroke for patients and produces a preventive measure plan to reduce risk levels is now in testing, and the results of presented work will be introduced into this system [7].

Acknowledgements

Part of this work was financially supported by the Russian Foundation for Basic Research, the research project №18-29-26007.

References

[1] J. Park H and Kim, Radiomics as a Quantitative Imaging Biomarker: Practical Considerations and the Current Standpoint in Neuro-oncologic Studies, *Nuclear Medicine and Molecular Imaging* **52** (2018), 99-108.

[2] L.E. Court, X. Fave, D. Mackin, J. Lee, J. Yang and L. Zhang, Computational resources for radiomics, *Translational Cancer Research* **5** (2016) 340-348.

[3] Isles Challenge 2018, Ischemic Stroke Lesion Segmentation, (2018). http://www.isles-challenge.org (accessed May 14, 2019).

[4] Kaggle. https://www.kaggle.com (accessed May 14, 2019).

[5] R. Feng, M. Badgeley, J. Mocco and E.K. Oermann, Deep learning guided stroke management: a review of clinical applications, Journal of neurointerventional surgery **10** (2018), 358-362.

[6] Pyradiomics, Welcome to pyradiomics documentation! (2016). https://pyradiomics.readthedocs.io (accessed May 14, 2019).

[7] O.G. Grigoriev, B.A. Kobrinskii, G.S. Osipov, A.I. Molodchenkov and I.V. Smirnov, Health Management System Knowledge Base for Formation and Support of a Preventive Measures Plan, *Procedia Computer Science* **145** (2018), 238-241.