IL NUOVO CIMENTO Vol. 37 C, N. 4 Luglio-Agosto 2014
DOI 10.1393/ncc/i2014-11803-6

Communications: SIF Congress 2013

Design of a system to achieve diagnosis of brain lesions of suspected infectious origin from a database and a radiologic software(*)

D. Grosso

Institute for Maternal and Child Health IRCCS "Burlo Garofolo" - Trieste, Italy

ricevuto l'1 Febbraio 2014; approvato il 3 Maggio 2014

Summary. — This Brainsupporters project aims at creating an international database including an extensive collection of radiologic images of brain lesions and the relevant clinical, laboratory and microbiological data, and a software, to highlight and differentiate the minimal alteration in the radiologic footprint related to the local alterations of brain tissue produced by each individual pathogen that will become an objective diagnostic tool. Clinical parameters will be taken into account in defining a procedure to estimate the probability that the difference between an image and a template be due to a specific pathological agent.

PACS 87.61.-c – Magnetic resonance imaging.
PACS 87.57.-s – Medical imaging.
PACS 87.57.R- – Computer-aided diagnosis.

1. – Introduction

At present, the diagnosis of brain infections is very complex because of the low contrast and consequently hard interpretation of radiologic images.

(*) This project is carried out by:

- D. Grosso (daniele.grosso@burlo.trieste.it),
- N. Maximova (natalia.maximova@burlo.trieste.it), the scientific leader,
- D. Zanon (davide.zanon@burlo.trieste.it),
- F. Zennaro (floriana.zennaro@burlo.trieste.it)

from Institute for Maternal and Child Health IRCCS "Burlo Garofolo" Trieste, Italy and

- T. Minuzzo (tiziano.minuzzo@ve.ismar.cnr.it)

from the Italian National Research Council (CNR) - Institute for Marine Science (ISMAR).

As brain infections in immunosuppressed onco-haematologic patients undergoing chemotherapy or transplant of haemopoietic stem cells have the worst prognosis, diagnostic images should be made easier to read and more homogeneous to increase sensitivity and specificity so as to apply targeted treatments as quick as possible.

Patients who have undergone immunosuppressive therapy after transplantation have the same problems, thus the target of the study are patients of any age afflicted by oncologic or onco-hematologic diseases or that have performed hematopoietic stem cell transplant (HSCT) or solid organ transplant (SOT).

Brain infectious diseases are life threatening events not so rare in patients undergoing SOT or HSCT or doing immunosuppressive therapies.

It is today impossible to establish with absolute certainty which pathogen caused a brain infection on the basis of imaging alone and clinical and microbiological criteria are often of little help.

The Brainsupporters project has as its main objectives to create an international database collecting a wide numbers of clinical and radiological records of brain lesions of a suspected infectious origin and to develop a collection of software tools to support the diagnostic process.

The database should include an extensive collection of radiologic images of brain lesions and the relevant clinical, laboratory and microbiological data.

The software tools for image analysis should be designed, to highlight and differentiate the minimal alteration in the radiologic footprint related to the local alterations of brain tissue produced by each individual pathogen.

A starting point is to develop a system capable of estimating the probability that the difference between an image and a template be due to a pathological agent and not to biological variability. The second step is to combine information from both biological and clinical data and radiological images to reach better specificity and sensitivity, possibly high enough to try to identify the specific pathological agent.

To achieve these results we need to have a reasonable number of relatively rare cases thus many centres must be involved in the project and every centre should have interest in giving an active contribution sending images and data.

This objective may be reached by the implementation of a cooperative model between centres and giving raise to an international multicentre project.

2. – Materials and methods

2˙1. *Brainsupporters multicentre collaborative model.* – An emerging question about medical data involves their economic value (which may also be referred to as a cost).

Many centres already ask for money for every radiological image they send when they participate in a project or offer money to other centres whose collaboration is, *i.e.*, send rare cases. Here is the problem when you have to deal with limited budget, while at the same time a large database is required.

To promote data exchange among the collaboration we defined the following model, inspired by NTU cooperative models from the theory of games:

- Every information in the database is related to a referrer (the oncologist, the radiologist or both).
- Every referrer must be related to a centre.
- Every referrer may upload cases into the database.

– Every uploaded case is evaluated and eventually approved by two expert oncologists and two expert radiologists, in blind. A case is flagged as approved if and only if there is a complete agreement.

– The Collaboration Weight index for every referrer is updated starting from the number of approved cases uploaded:

$$CW = k\frac{UC}{TC}\,, \tag{1}$$

where: k is a coefficient, UC is the number of approved cases he uploaded and $TC = \sum_{i=1}^{N} UC_i$ is the number of approved cases.

– Every referrer has access to (and may download):

- his own radiological images and biological or clinical data
- images and data from other referrers, with just one limit: the number N of cases he has access is limited accourding to:

$$N = f(CW), \tag{2}$$

where f is a monotonically increasing function.

– Clinical data and radiological images are processed every time a new data is approved.

– Results of the elaboration (templates, atlases and statistic results from biological data) are accessible to every centre.

With this model we are giving the opportunity to access rare cases not every centre may encounter because of their relative low number of patients. On the other side, if every case has a value, while increasing the number of cases you have access you are increasing your clinical cases database (and your knowledge which may result in a more accurate diagnosis).

2˙2. *Architecture, data workflow*. – The true referrer for every centre is the oncologist. Every centre has ben sent an invitation letter. If an oncologist is interested in joining the collaboration he asks for registration in the collaboration site (`www.brainsupporters.eu`). He is given an account and he may access the collaboration database and submit his cases. He may create primary report for his patients and insert follow up data or radiological images (anonymized and compressed). Our protocol suggest he may submit T1, T2 and flair NMR images and eventually CT or fMRI images.

From the literature [1, 2] we learn that more than 50% of leukemia patients have infections from *Aspergillus* and more than 25% from *Candida*. About 90% of the patients were positive for *Toxoplasma gondii* before they underwent transplantation and about 15% may have *Toxoplasma* infection, as defined by fever and a positive polymerase chain reaction (PCR) found for *Toxoplasma gondii* in blood.

Now we know what we have to look for; if we search for lesions from *Aspergillus*, *toxoplasmosis* and *candidiasis* we should cover the great majority of possible disease causes.

We have subjects who belong to one of the following categories: I0 unknown, I1 proven IFD, I2 ... and, at this stage, we wish our software tools to classify them in one of the following categories:

– I0 unknown,

– I1 negative (no evidence of IFD),

– I2 possible IFD (probability of IFD $p \geq p_0$),

– I3 proven IFD (probability of IFD $p \geq p_1$).

It is required a criteria for proven IFD (invasive fungal disease) except for endemic mycosis. We adopted EORTC/MSG Consensus Group revised definitions [3].

2˙3. *Database for research and pseudoanonimyzation.* – Medical image processing and analysis requires testing and validation with real clinical data. In our case both clinical and radiological data are to be stored for research purposes. It must be taken into account that a common database of a multicentre project may involve data exchanges among different centres while preserving, at the same time, the privacy of the patients.

The kind of database required for this project must admit follow-up insertion thus complete data anonymization is not a practicable issue. An operator may be allowed to access any of its patient "clinical folder" and insert new "follow-up" data.

A practical solution to this problem is pseudoanonymization [4] where any patient is given a numeric ID and eventually an alias as a key to retrieve its data.

2˙4. *Database engine specifications.* – First of all it is necessary to build a container to host clinical data and the images of the patients. This database should be accessible and easy to use. It should allow data access even from tablets and possibly smartphones, even without installing supplementary software. It should allow at least a user/group permission scheme designed in a way that many implementation of permission schemes should be possible. The database engine should be open and allows easy interfacing with statistical analysis tools like R([1]). The database engine should be based on diffused, well tested, scalable, and stable tools.

We used a mysql server installation with a web interface based on DaDaBIK([2]) (enterprise release 5.1.2) which is an open source database interface realized with PHP([3]). It is open source but requires a fee (about 190$ for the enterprise version). Dadabik is a solid project and is well supported. During the preliminary test phase we noticed that there was a bug preventing the correct visualization of details in a form. With the help of their support we discovered it was necessary to replace

```
if ($primary_key_field_field != "" && $field_element['type_field'] ===
'select_single'){
```

at line 4904 with

([1]) A multiplatform free software environment for statistical computing available online at the following address: `www.r-project.org`.
([2]) `http://www.dadabik.org/`.
([3]) `http://www.php.net/`.

TABLE I. – *Database tables.*

N.	Table	Note
01	onebase_ht_web_requeste_registration	cf_id
02	onebase_at_organizations	id_organization
03	onebase_at_organizations_users	id_organization_user
04	onebase_at_users_phonenumbers	id_user_phonenumber
05	onebase_at_patients	id_patient
06	onebase_ht_patients_primary	id_primary
07	onebase_ht_patients_followup	id_followup
08	onebase_ht_examinations	id_examination
09	onebase_ct_diagnostic_culture_pathogen	id_diagnostic_culture_pathogen
10	onebase_ct_diagnostic_imaging	id_diagnostic_imaging
11	onebase_ct_diagnostic_liquid	id_diagnostic_liquid
12	onebase_ct_diagnostic_molecular_liquid	id_diagnostic_molecular_liquid
13	onebase_ct_diagnostic_molecular_methods	id_diagnostics_molecular_methods
14	onebase_ct_evaluation	id_evaluation
15	onebase_ct_hscs_source	id_hscs_source
16	onebase_ct_immunoglobulins_levels	id_immunoglobulins_levels
17	onebase_ct_lymphocytes	id_lymphocytes
18	onebase_ct_lymphocytesCD4	id_lymphocytesCD4
19	onebase_ct_major_immunosuppressive_therapy	id_major_immunosuppressive_therapy
20	onebase_ct_monoclonal_antibody	id_monoclonal_antibody
21	onebase_ct_neutrophils	id_neutrophils
22	onebase_ct_serum_ferritin	id_serum_ferritin
23	onebase_ct_treatment_types	id_treatment_type
24	onebase_ct_type_of_donor	id_type_of_donor

```
if ($primary_key_field_field != "" &&
$fields_labels_ar[$i]['type_field'] === 'select_single'){
```

in include/business_logic.php to eliminate the bug.

We installed it on an apache-based web server, with other tools to convert and analyze the uploaded images and to perform statistical analysis on biological and clinical data.

2˙5. *The database architecture.* – Table I contains a list of all the tables([4]) included in the database([5]).

Table 01 contains web requests of registration automatically generated from the relative web site form. Table 02 contains a list of organization, table 03 contain organizations-users links and table 04 is a directory of people joining the collaboration.

Most important tables are the following ones:

– 05 which contains pseudoanonymized patient data

– 06 is filled during the patient registration

([4]) A table is an organized set of data elements.
([5]) A collection of tables togheter with their relationships.

Fig. 1. – S_0

Fig. 2. – $N = N(S_0)$

Fig. 3. – MinMax2(N)

– 07 dedicated to contains biological and clinical data acquired during the followup

– 08 contains examination data (from images metadata)

Tables from 09 to 24 are reserved for value selections. Other tables are reserved for the database catalog, the user interface and the implementation of permission rules.

2˙6. *Some algorithms we are testing.* – We are looking for small regions with low contrast and possibly spread all over the brain thus we need to enhance image readability and obtain a first ROI collection as a first step toward automatic segmentation.

We are now testing three different approaches; the first one is to generate an image with values obtained from the standard deviation on a window([6]). The objective here is to have a measure of the variability of the signal at a given scale. This is based on the assumption that the signal inside a lesion should be more uniform because of physiological processes.

2˙6.1. MinMax. If a lesion is of the order of magnitude of just a few voxels it may be helpful to analyze min and max values and try to uniform the signal around every spot signal, possibly while filtering noise and preserving edges at the same time.

This could be done with a filter acting on a parallelepiped-like interval:

$$S(P^*, r) := \{P(x_1 \dots x_n) | r - x_{i=1\dots n} \leq x_{i=1\dots n} \leq r + x_{i=1\dots n}\}. \tag{3}$$

The filter should cause a minimal variation of the gradient module and direction. The following one is a good candidate:

$$\min(I, r) := \min(I(P \in S(P, r))) \to \min_2(I(P \in S(P, r))), \tag{4}$$

$$\max(I, r) := \min(I(P \in S(P, r))) \to \max_2(I(P \in S(P, r))). \tag{5}$$

Every point with value $v = v_{\min}$ in S ($\min \equiv \min_1$ the absolute min value on interval $S(P^*, r)$) is replaced with the second order min value $\min_2$ ($v := \min_2$) and the same happens for $\max_1$ which is replaced with $\max_2$.

Figure 3 shows the effect of a filter with $r = 1$ acting on fig. 1 after it has been corrupted by noise (fig. 2).

This filters are used to remove salt and pepper noise but they may be useful in our case too.

([6]) A parallelepiped-like shape including $N = (2r + 1)^3$ voxels.

Fig. 4. – S_0

Fig. 5. – $EDI(I)$

Fig. 6. – $G(I, EDI(I))$

2˙6.2. EDI images. Accourding to Shannon theory it is possible to define the entropy of a system (in bits):

$$H(X) = -\sum_{i=1}^{N} p_i \log_2(p_i), \tag{6}$$

where p_i is the probability of state i and N is the total number of possible states of a system. This equation is related to Gibbs entropy $S = -K_b \sum p_i \ln p_i$ of a thermodinamic system.

Starting from eq. (6) it is related to define the information (in bits) which is necessary to transmit on a channel to change the entropy of a system:

$$\Delta I = -(H(X_1) - H(X_0)), \tag{7}$$

X_0 and X_1 are the configurations of the system before and after the transmission and $H(X_0)$ and $H(X_1)$ are the entropies of the system before and after the transmission.

Equation (7) is also a measure of the entropy reduction due to the loss in uncertainty about the state of the target system.

Starting from this equation it is possible to generate an image $EDI(I)$ having in every point P_i a value obtained computing the average information transferred by the symbol in the same position P_i in image I.

Such an image may be referred to as an Entropy (spatial) Distribution Image and is a feature space image generated from the original one (with an uninvertible transformation).

Figure 5 is obtained by computing an EDI transform on fig. 4 while fig. 6 has been obtained combining figs. 4 and 5 in hsv space with a given hue h, value v proportional to values in S_0 and signal s proportional to values in $EDI(I)$.

The photo in fig. 4 has been taken with a CCD camera by using a flash whose light reflected on the wall results visible in both $EDI(I)$ and $G(I, EDI(I))$.

Figure 8 has been obtained from the RMI in fig. 7 while fig. 10 has been obtained from the CT image in fig. 9. In both cases many anatomical tracts remain visible even in the transformed images while values in the transformed images have the same units.

This approach may put in evidence local fluctuation of the spatial distribution of the information in an image I when the configurations of bit related to the values we are searching for are relatively rare, even if not too different from the average (which may be the cause of small lesions with low contrast with a uniform background).

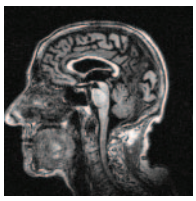

Fig. 7. – I_1

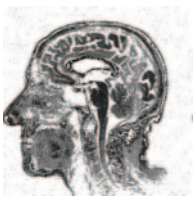

Fig. 8. – I_{1EDI}

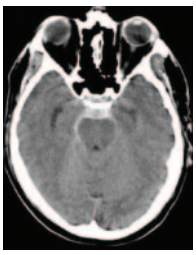

Fig. 9. – I_2

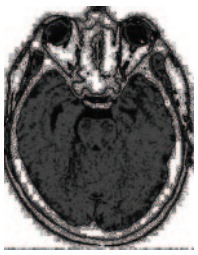

Fig. 10. – I_{2EDI}

2˙6.3. LEF images. The previous technique may be refined by computing the mean entropy on a parallelepiped-like ROI of a given size

$$s(Q) = -\frac{1}{k} \sum_{P \in W(Q,r)} f_Q \log_2(f_Q) \tag{8}$$

with $k = (2r+1)^n$ and $W(Q,r) = \{Q | Q_1 - r \leq Q_1 \leq Q_1 + r, \ldots, Q_n - r \leq Q_n \leq Q_n + r\}$.

Figure 12 has been obtained from starting from the image in fig. 11 with $r = 2$ which is of the order of magnitudo necessary to let the wall texture become visible. Details of different size in the image in fig. 12 were lost but if we recombine the images, like in fig. 13 we obtain a readable image with a more visible texture than in the original image.

The image in fig. 13 has been generated combining the ones in figs. 11 and 12 in hsv space with a given hue h (different than the one used before). The value v is again proportional to values in S_0 and the signal s is proportional to values in S_{0LEF}.

The aim of this filter is to enhance readability of small lesions and hopefully put them in evidence.

Images in figs. 15 and 16 have been obtained from the RMI in fig. 14 with a LEF filter of radius $r = 1$ and $r = 3$ thus they show details of different order of magnitude.

2˙6.4. CORED. To find a localized difference D between two images I and J of the same patient, taken during different follow-up or between the image I of a patient and a template J it is necessary to coregister the images to minimize the term M due to misregistration

$$J = I + D + M. \tag{9}$$

Fig. 11. – S_0

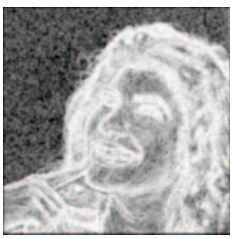

Fig. 12. – S_{0LEF}

Fig. 13. – $F(S_0, S_{0LEF})$

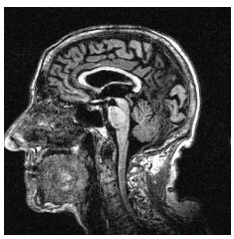

Fig. 14. – I_1

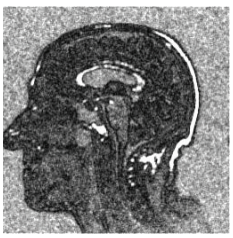

Fig. 15. – $I_{LEF(1)}$

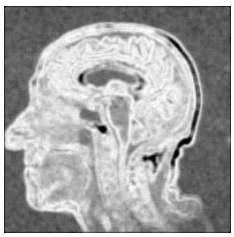

Fig. 16. – $I_{LEF(2)}$

To solve the coregistration problem we are interested in finding the optimal transform $T_{p_1 \dots p_n}(I) = J$ depending on parameters $p_1 \dots p_n$ which satisfies the following equation:

$$\lim_{p_1 p_n \to p_1^* \dots p_n^*} \|M(p_1 \dots p_n)\| \to 0. \tag{10}$$

We require this transformation to be invertible and smooth.

We are now exploring different solutions mainly based on EDI transform and Mutual Information and diffent approaches to the optimization problem.

3. – Results and discussion

Brain infections in immunosuppressed patients have the worst prognosis. It should be useful to have a radiological diagnostic tool capable of increasing sensitivity and eventually of identifing at least the three more frequent pathogens.

To achieve this goal we planned the following steps:

1. design and implement a database hosting biological data and radiological images;
2. develop some software tools to enhance contrast, segment and eventually measure lesion parameters and combine clinical and radiological data with an objective in mind: to obtain the probability of presence of a specific pathological agent;
3. develop a software to perform elastic image registration to achieve better image comparison against a template and combine clinical and radiological data to obtain a better estimation of the probability of presence of a specific pathological agent.

We wrote down a proposal and submitted it to EBMT (Europian Society for Bone and Marrow Transplation) and it has been appreciated. The proposal has been reviewed and approved by our Ethic Committee and it has become a project.

To start collecting a reasonable number of radiologic images of infectious brain lesions in pediatric patients with confirmed or suspect infectious etiology it is necessary to start an international collaboration because these patologies are relatively rare.

We realized the scientific database to host clinical and biological data and images from any centre. The database uses pseudoanonymization to preserve the patient's privacy.

An internet site (`www.brainsupporters.eu`) has been created, and is freely accessible to oncohaematologic centres worldwide.

Both database and web server have been installed with the support of the National Research Council (CNR ISMAR in Venice) and the database is online.

Now we are ready to start collecting images and clinical data and we are inserting data from our centre for testing purpose. Collected images may be accessed through a gallery with advanced research and image comparison functions still under development.

We are now installing a local mirror at Burlo Garofolo of the database server and testing all the software tools required for image elaboration before they enter in production.

We have adopted coherent criteria for diagnosis and have selected a multicenter team for the definition of diagnostics criteria for brain lesions of infectuous origin (in collaboration with the Catholic University of Sacred Heart, Institute of Microbiology, Rome, Italy). We will use these protocols to try to develop diagnostic and radiological follow-up protocols, currently not available for this kind of complicances and these target patients.

Characteristics of the typical pathogen findings —hardly detected by radiologists— will be detected by a software, which will use algorithms we are actually testing with images from our centre.

The collected images will be subject to a blind evaluation (with no microbiological diagnosis) specialists in blind; only proven diagnosis will be taken into account for testing new algorithms

Images will then be classified in different groups, to evaluate if and how the "radiologic footprint" of lesions due to the same pathogen changes according to the host (patient).

Our project has been exhibited at the London 2013 EBMT annual meeting, at the ESOT Vienna 2013 congress, at the 99th congress of SIF in Trieste and at the EMMIT 2013 Euro-Mediterranean Medical Informatics and Telemedicine, and finally in Barcelona at the EMBT infectious diseases working party, with success.

* * *

This work was produced, supported and perpetrated by the Brainsupporters collaboration, under the auspices of the Burlo Garofolo, Trieste.

REFERENCES

[1] Leventakos Konstantinos, Lewis Russell E. and Kontoyiannis Dimitrios P., *Clin. Infect. Dis.*, **50** (2010) 405 doi:10.1086/649879.

[2] Martino R., Maertens J., Bretagne S., Rovira M., Deconinck E., Ullmann A. J., Held T., Cordonnier C. for the European Group for Blood and Marrow Transplantation Infectious Diseases Working Party, *Clin. Infect. Dis.*, **31** (2000) 1188.

[3] De Pauwa Ben, Walsha Thomas J., Donnellya J. Peter, Stevens David A., Edwards John E., Calandra Thierry, Pappas Peter G., Maertens Johan, Lortholary Olivier, Kauffman Carol A., Denning David W., Patterson Thomas F., Maschmeyer Georg, Bille Jacques, Dismukes William E., Herbrecht Raoul, Hope William W., Kibbler Christopher C., Jan Kullberg Bart, Marr Kieren A., Muoz Patricia, Odds Frank C., Perfect John R., Restrepo Angela, Markus, Ruhnke, Segal Brahm H., Sobel Jack D., Sorrell Tania C., Viscoli Claudio, Wingard John R., Zaoutis Theoklis and Bennett John E., *Clin. Infect. Dis.*, **46** (2008) 1813 doi:10.1086/588660.

[4] Noumeir Rita, Lemay Alain and Lina Jean-Marc, *J. Digital Imaging*, **20** (2007) 284.