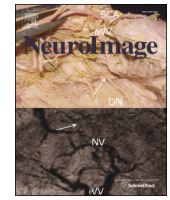
NeuroImage xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg



Data integration: Combined imaging and electrophysiology data in the cloud

Lohith G. Kini [a,*], Kathryn A. Davis [b], Joost B. Wagenaar [b]

[a] Department of Bioengineering, University of Pennsylvania, 240 Skirkanich Hall, 210 South 33rd Street, Philadelphia, PA 19104-6321, USA
[b] Department of Neurology, Hospital of the University of Pennsylvania, 3400 Spruce Street, 3 West Gates Bldg, Philadelphia PA 19104, USA

ARTICLE INFO

Article history:
Accepted 26 May 2015
Available online xxxx

ABSTRACT

There has been an increasing effort to correlate electrophysiology data with imaging in patients with refractory epilepsy over recent years. IEEG.org provides a free-access, rapidly growing archive of imaging data combined with electrophysiology data and patient metadata. It currently contains over 1200 human and animal datasets, with multiple data modalities associated with each dataset (neuroimaging, EEG, EKG, de-identified clinical and experimental data, etc.). The platform is developed around the concept that scientific data sharing requires a flexible platform that allows sharing of data from multiple file formats. IEEG.org provides high- and low-level access to the data in addition to providing an environment in which domain experts can find, visualize, and analyze data in an intuitive manner. Here, we present a summary of the current infrastructure of the platform, available datasets and goals for the near future.

Introduction

Intracranial electroencephalography (iEEG) and magnetic resonance imaging (MRI) are equally important neuroimaging modalities for studying neural activity and structure (Liu et al.; Dale and Halgren, 2001). Accurate dynamics of neural activities and interactions can be studied across large number of electrodes and with high temporal resolution using iEEG (Dale et al., 2000; Hämäläinen et al., 1993; Cohen et al., 1980; Williamson et al., 1978). Precise localization of this neuronal activity along with structural descriptions of neuronal pathways can be determined using MRI with different contrasts (Kwong et al., 1995). In addition, complementary features of neuronal activity can be studied using fMRI (Ogawa et al., 1992; Kwong et al., 1992; Belliveau et al., 1991). The explosion of technology in these modalities has spurred interest in their use in studying neuroscience and advanced translational research in neurology, particularly epilepsy.

Joint analysis of these multimodal datasets allows neuroscientists to study functional and structural relationships and has great potential to lead to vital discoveries in epilepsy otherwise not captured when studying modalities independently (So, 2000). Additional data modalities further impact the overall richness of a dataset, such as patient clinical history, genetics, electrocardiogram (EKG), etc. Integrating, analyzing, and sharing these complex datasets pose unique challenges to the data science community. Individual institutions have access to small datasets. Significant advancements in this field of study will be made using large datasets shared across multiple institutions. Unique challenges of sharing human biomedical data arise such as data format interoperability, de-identification of protected health information (PHI), and adherence to mandated government regulations.

The ability to share clinical metadata along with high-resolution data, such as iEEG and neuroimaging across multiple sites, often geographically sparsely distributed, requires novel infrastructure with a focus on data integration. A simple approach such as allowing collaborators to download the data to their local site for analysis is often not feasible, especially with the terabytes (TB) of data that comprise these datasets.

Over the past 4 years, our team of neuroscience and computer science experts has established a cloud-based resource for data sharing and collaboration, http://IEEG.org (Wagenaar et al., 2013). This platform provides data sharing and analysis capabilities to the neuroscience community, particularly in the epilepsy domain. Multiple neuroscience research centers are making their data available through the platform for collaborations where data access is controlled, and access to data is controlled by the data contributors.

Currently, the IEEG-Portal contains high-quality iEEG and multimodal imaging from over 1200 subjects including 576 animal models (dog,

Abbreviations: iEEG, Intracranial electroencephalogram; EEG, Scalp electroencephalogram; EKG, Electrocardiogram; API, Application program interface; T1W, T1-weighted MRI; T2W, T2-weighted MRI; FLAIR, Fluid-attenuated inversion recovery MRI; BOLD, Blood oxygen level dependent imaging; ASL, Arterial spin labeling; HARDI, High angular resolution diffusion imaging; MRS, Magnetic resonance spectroscopy; PET, Positron emission tomography; SPECT, Single-photon emission computed tomography; MPRAGE, Magnetization-prepared rapid gradient-echo MRI; DICOM, Digital Imaging and Communications in Medicine; NIfTI, Neuroimaging Informatics Technology Initiative Data Format.

* Corresponding author at: Department of Bioengineering, 240 Skirkanich Hall, 210 South 33rd Street, Philadelphia, PA 19104-6321. Fax: +1 215 573 8393.
E-mail addresses: lkini@mail.med.upenn.edu (L.G. Kini), Kathryn.Davis@uphs.upenn.edu (K.A. Davis), joostw@seas.upenn.edu (J.B. Wagenaar).

http://dx.doi.org/10.1016/j.neuroimage.2015.05.075
1053-8119/© 2015 Published by Elsevier Inc.

Table 1
The number of patients by imaging modality per institution (from the IEEG Portal dataset).

Number of patients by imaging modality per institution	Mayo Clinic	Hospital of the University of Pennsylvania
Structural modalities		
3 T MRI–T1W	23	23
T2W	2	19
FLAIR	10	19
DWI/DTI		8
ECoG pre-implant imaging (MRI/CT)	23	14
ECoG post-implant imaging (MRI/CT)	23	21
6 months post-implant imaging (MRI/CT)		13
7 T MRI		20
HARDI imaging		20
Metabolic/functional modalities		
SPECT (ictal and interictal)	20	
18(F)-FDG PET	1	
BOLD		20
ASL		20
MRS		20

Abbreviations for imaging: positron emission tomography (18(F)-FDG PET), T1-weighted MRI (T1W), T2-weighted MRI (T2W), fluid attentuated inversion recovery sequence (FLAIR MRI), diffusion weighted imaging (DWI)/diffusion tensor imaging (DTI), X-ray computed tomography (CT), single-photon emission computed tomography (SPECT), blood oxygenation level dependent (BOLD), arterial spin labeling (ASL), high angular resolution diffusion-weighted imaging (HARDI), 1(H) magnetic resonance spectroscopy (MRS). All image sequences available as 3D NIfTI files.

mouse, rat, sheep, primate) and 733 patients with epilepsy. As this time of writing this article, there are 487 public datasets, 667 registered users, and 162 publicly accessible clinical datasets. These data were collected across multiple institutions throughout the world. Users originate from all 5 continents and represent institutions like UCLA in Los Angeles, CA, and University Hospital Motol in Prague, Czech Republic. Table 1 summarizes the data available on the IEEG-Portal.

Each human dataset can contain up to 100 electrodes and are recorded for 1–4 weeks continuously using sample rates as high as 32 kHz. Standard clinical epilepsy protocol images are also provided, including T1-weighted (T1W) isotropic axial, T2-weighted (T2W) coronal, FLAIR coronal, and diffusion-weighted image sequences. In addition, de-identified clinical metadata, such as patient medical and family history, medication history, Epilepsy Monitoring Unit reports, and scalp EEG findings are available for a subset of the patients on the portal. There are currently 46 patients with intractable epilepsy who have at least intracranial EEG, pre-operative T1W/T2W/DWI MRI, ECoG post-implant MRI or CT as well as a full clinical report. Half of these patients are from the Hospital of the University of Pennsylvania, and the other half are from the Mayo Clinic. Table 2 shows four example patients along with a sampling of the data available on the portal for these patients. All patients who have clinical reports available have lesional findings on histopathology and neuroimaging reported in the respective sections in the reports. These include lesions such as malformations of cortical development (focal cortical dysplasia or schizencephaly), vascular malformations (AVMs or cavernomas), and low-grade glial tumors. A significant minority of the patients currently on the portal has these lesional findings present on either histopathology or imaging.

Analyzing large-scale EEG and neuroimaging data requires substantial computational resources. Leveraging cloud resources provides a scalable solution to benchmark experiments, share gold standard datasets, and advance towards more integrative collaborative research in the neuroscience community.

IEEG.org, like other databases such as the Human Connectome Project (Van Essen et al., 2012), the European EEG database (http://epilepsy-database.eu) (Klatt et al., 2012; Ihle et al., 2012), and LONI IDA (Dinov et al., 2010) are critical to the standardization of neuroimaging data analyses, avoiding bias, and allowing for significant research advances (Ihle et al., 2012). Benchmarking experiments (i.e. testing algorithms on novel data) requires a central body to curate “gold standard” training data and withhold testing data. IEEG.org has been developed to allow users to share data and use its resources to validate and benchmark new algorithms.

Architecture of IEEG.org

The IEEG-Portal is developed using the Google Web Toolkit in Java and JavaScript. It is hosted on Amazon's EC2 service and the data are stored on the Amazon S3 service with reduced redundancy. Fig. 1 shows an abstract schematic of the various components of the IEEG.org

Table 2
Sample patient profiles from the IEEG portal.

Patient ID	I002_P002	Mayo Clinic Study 005	I002_P005	I002_P006
Institution	Hospital of the University of Pennsylvania	Mayo Clinic	Hospital of the University of Pennsylvania	Hospital of the University of Pennsylvania
Epilepsy type	Localization-related; Left frontal onset, left cortical dysplasia	Bi-temporal onset	Localization-related; Left temporal onset (secondary to hemorrhagic HSV encephalitis)	Localization-related; Symptomatic meningitis
Engel outcome	1	1	4	1
Seizure type	Simple partial and generalized tonic–clonic of left temporal onset	Focal complex partial endpoint of seizure not evolving to secondary generalized	Complex partial seizures of left temporal onset and status epilepticus	Complex partial seizures of right temporal onset with secondarily generalized tonic-clonic seizure
No. of seizures	1		0	5
Age during first seizure	4	21	30	11
Age at admission	20	26	35	32
Sample past anti-epileptic medications	Carbamazepine, topiramate, levetiracetam, clobazam, valproic acid	–	Phenytoin, valproic acid, levetiracetam	Carbamazepine, zonisamide
Imaging available	T1 MPRAGE, T2 FLAIR, post-implant T1 MPRAGE and FLAIR, post-implant CT, 6 months post-op T1, T2, FLAIR, DWI	MRI T1 pre-implant, SPECT ictal/interictal, MRI T1 post-implant, CT post-implant	Pre-implant T1 MPRAGE and T2, Post-implant T1 MPRAGE and FLAIR, Post-implant head CT	T1 MPRAGE, T2 FLAIR, T2 susceptibility-weighted, DTI, pre-implant T1 MPRAGE and T2 FLAIR, Post-implant T1 MPRAGE and FLAIR, post-implant CT, 6 months post-op T1, T2, FLAIR

Four patients and their profiles on the IEEG Portal. Dates of hospital admission for epilepsy monitoring have been standardized to start on Jan-01-2000. All other dates have been adjusted accordingly. Note that some data are not available and marked by –. All patients who have lesional epilepsy will have lesion findings noted under MRI report summaries in the respective patients' clinical reports. All lesional patients will have appropriate pathology and imaging findings reported, whether lesions are malformations of cortical development, vascular, or low-grade glial tumors.

Please cite this article as: Kini, L.G., et al., Data integration: Combined imaging and electrophysiology data in the cloud, NeuroImage (2015), http://dx.doi.org/10.1016/j.neuroimage.2015.05.075