Machine learning enhanced virtual autopsy

Shane O'Sullivan[a], Andreas Holzinger[b], Kurt Zatloukal[c], Paulo Saldiva[a,d], Mohammed Imran Sajid[e], Dominic Wichmann[f]

STATE-OF-THE-ART AND OBJECTIVES

In many cases, access to biological samples is needed for molecular analyses. Biopsies or other minimally-invasive sampling techniques, in the context of virtual autopsy (virtopsy), would reduce risk and improve societal acceptance when compared with traditional autopsies.

We propose a study that compares and contrasts virtopsy with traditional autopsy, using large study groups based on four different levels of information: Macroscopic; Microscopic; Medical imaging; and Molecular phenotypes. This proposal includes the development of a machine learning enhanced virtual biobank to understand the process of diseases, and to evaluate clinical diagnosis and treatment. This virtual biobank, enhanced by machine learning and knowledge extraction approaches, will be composed of a unique collection of non-invasively generated autopsy images (e.g., X-ray, computerized tomography, magnetic resonance imaging, ultrasound) and digital pathology imaging data of corresponding biological samples.

Based on this vast resource, we will experimentally design, develop, test and evaluate machine learning algorithms that can self-learn from, as well as make predictions (based on the virtual biobank data). These ambitious studies will help us go beyond the state-of-the-art to demonstrate to what extent machine learning can enhance medical expertise, so as to understand the process of diseases and to evaluate both clinical diagnosis and treatment.

The algorithms will make data-driven predictions or decisions, by building a model from sample inputs; a variety of subject data was used. We will use the same machine learning approach developed by a research group at the Medical University of Graz.[1] They are world-leading experts in this field with long-standing experience in the application domain, health generally, pathology and metabolomics[2] in particular.

This strategy will allow achievement of results that include the use of fewer images when compared with conventional machine learning (e.g., deep learning approaches), and so helps to alleviate some major issues one would expect for this type of study set-up. In the application of virtopsy we do not have access to many cases, (i.e., sample-size is small, and not the large numbers in the thousands to millions which are usually necessary), and so consequently we need algorithms which are able to self-learn (with only a few examples), in the same way humans do. The chosen approach of interactive machine learning is an inventive way to overcome this problem. The intention is to utilize this

[a] University of Sao Paulo (USP), Faculty of Medicine, Department of Pathology. São Paulo, SP, Brazil.
[b] Medical University of Graz, Institute for Medical Informatics/Statistics, Holzinger Group. Graz, Austria.
[c] Medical University of Graz, Institute of Pathology. Graz, Austria.
[d] University of de Sao Paulo (USP), Institute of Advanced Studies. São Paulo, SP, Brazil.
[e] Wirral University Teaching Hospital, Department of Upper GI Surgery. Wirral, United Kingdom.
[f] University Hospital Hamburg Eppendorf, Department of Intensive Care. Hamburg, Germany.

method of smart injective machine learning. It is not a black box method, which demands the input of a massive amount of data. Instead, expert knowledge is fed directly into the algorithmic loop.[1] This is an ingenious method to deal with standard limitations. The necessity of large amounts of data/sample-size and cases will always be a limitation for conventional machine learning. Our technique will have a unique resource to drive the project and achieve our anticipated results.

Regarding virtual microscopy, digital pathology now provides an entirely innovative technique to analyze image data from tissues. We propose a study that conducts a systematic comparison between the morphology of tissue (as seen by digital pathology) and medical images such as magnetic resonance imaging (MRI).

In addition, pathologists will provide images using a medical care standard digital camera for macroscopic documentation and analysis. This image will be compared with corresponding medical images, as different levels of imaging information add to the originality of this study.

Finally, we propose to include nuclear mass resonance (NMR) spectroscopy in the active research of this study. This adds to the originality and scientific aspect as it involves a new level of information regarding metabolic status and morphology of tissue. Consequently, focus is not only on explaining disease biology, but it is also developing machine learning algorithms that provide image analysis – currently there is a lack of literature on large-scale programmes that compare MRI images with both NMR and microscopic images.

NMR will enable us to obtain the metabolomic profile, as the metabolome provides detailed information about the chemical composition of the tissue. For NMR metabolomics, it is essentially the same principle as for MRI, but the read-out of the signal is different. This is scientifically challenging but relevant, as we can then compare MRI data with chemical composition of tissue investigated. In addition, we will compare this data with histological and macroscopic data. There has been no current literature to date demonstrating that this has been undertaken in such a coordinated approach as set out in our proposal.

Ideally, for this type of image analysis study, we intend to include many additional datasets from autopsies performed rapidly (~3 hours) after clinical death. This is important because virtopsy must be compared with traditional autopsy whilst avoiding autolysis. The proposed virtual biobank will be used to train algorithms to cope with effects due to autolysis and other post-mortem effects. This is also of huge relevance to forensic medicine. By performing autopsies in as early as 3 hours after clinical death, we are managing a more feasible situation when compared with secondary phenomena due to autolysis. As a result, we will collect better material for molecular studies, and this is cutting-edge research with the generation of new knowledge in the scientific approach.

Furthermore, accessing the affected organs and tissues by traditional autopsy provides an opportunity for molecular analyses, such as metabolomics by using NMR spectroscopy. As discussed, the innovative aspect we will introduce to this study is to compare virtopsy with traditional autopsy using four different levels of data information:

(1) Macroscopic appearance (viewed by eye in traditional autopsy using digital photography);

(2) Microscopic appearance (examined by microscope in digital pathology);

(3) Medical imaging appearance (e.g., seen by MRI, ultrasonography, etc.); and

(4) Molecular phenotypes (e.g., by NMR-based metabolomics).

In brief, the main objective is to demonstrate that it is possible to create a self-learning system, which will assist pathologists and physicians in diagnosing certain medical conditions from digital data. This information technology (IT) learning approach is a major objective of the proposal. We will create an IT-based methodology that provides an innovative technique to perform autopsies. This includes all other issues with machine learning to facilitate physicians to determine the correct diagnosis; it revolutionizes the method of performing autopsies.

CORE SCIENTIFIC APPROACH

The core scientific approach is the way in which we analyze imaging data and how we compare the MRI data with morphology or metabolic status. When we

use the power of machine learning to analyze imaging data, it is very similar in pathology versus radiology since the approaches are quite similar. However as discussed earlier, a conventional machine learning approach requires many thousands of datasets. Additionally, such a study requires exceptionally detailed annotation to achieve a particular goal; this is a valid prerequisite that one may expect to be unquestionable and leading to a great result. The limited number of datasets will constantly be an issue in any autopsy study, whether it is pathology or forensic medicine – there are simply not enough cases. This consistent limitation is overlooked by most machine learning researchers. However, by using the unique technique of interactive machine learning, we can reduce the number of datasets required at least by a factor of 10 or even more. Therefore, we increase our chances of obtaining a reasonable result very early on in the study - even with a low number of cases - and even after a short period of time, which places us in a very promising position within the specific area of artificial intelligence. We generate a solid case by applying this unique machine learning approach in this particular context.

By making the systematic comparison of tissue alterations across different organs, we generate digital slides and then compare the histologic imaging features with medical imaging features, e.g., MRI; one of the techniques we use to link virtopsy to digital pathology.

This study provides a systematic understanding that can distinguish between postmortem inflammation infarcted from autolysis, and thereby generates a new database for further imaging programs and algorithm developments. This is a "machine learning from images" study that has a novel scientific approach with a practical setup, a strong focus on what to look for, and an innovative way to analyze the data.

In most countries, there are inadequacies with regards to conducting large-scale programs for this type of study. Firstly, for some subjects, pathologists do not receive approval from the relatives, and therefore they perform fewer autopsies. Secondly, the time lapse of consent is approximately one day, resulting in performing autopsies with a marked delay after death. Thus, it is most advantageous to include additional cases that examine a situation more closely related to a living body, which allows us to distinguish between disease-related alterations and secondary alterations (caused by prolonged autolysis periods).

The virtual biobank will facilitate data fusion, data mapping, data integration and data sharing on four different levels: (1) Macroscopic; (2) Microscopic; (3) Medical imaging; and (4) NMR. The virtual biobank will consequently be the basis for a fully-fledged and powerful machine learning pipeline. We propose to include analysis of in vivo and ex vivo brain structural changes in a large sample of subjects; adding a novel parameter to the study and enabling insight into entirely original, previously unknown aspects.

A team with in-depth experience in machine learning and knowledge extraction will support and help in building the machine learning pipeline and novel knowledge extraction tools. In particular, the highly needed visualization tools that enable the experts to find, e.g., anomalies, similarities, dissimilarities in arbitrarily high dimensional data sets. These would be otherwise inaccessible to the human end user, if following an interactive machine learning approach with the doctor-in-the-algorithmic-loop.[3,4]

Our unique technique is that we propose to construct machine learning algorithms that self-learn from and make predictions based on the study group data stored in our virtual biobank. Imaging biomarkers depend on access to biomaterial, and these biological features are readily available at many medical universities. Some academic hospitals remain very open to both traditional and imaging autopsies for clinical and academic studies.[5,6] The results from the proposed virtopsies are highly reproducible and most developed countries can facilitate and provide virtopsies.

In most cases, a study of this kind would usually require 1-2 years of data collection. However, some research groups possess a record of previously existing virtopsy data, which can enable them to perform machine learning image analysis from the onset. This data allows researchers to develop algorithms which they then validate, adapt, use and test in the autopsy cases which are to be collected during the random period of their study.

Initially, they can train the algorithms on existing datasets, which are available from previous virtopsy studies, improving the likelihood of achieving expected results. However, it is critical to include many datasets from autopsies performed rapidly (~3 hours) after

clinical death to avoid the risk that ultimately the algorithms will fail to recognize autolysis, which is the most common phenomenon in these cases. The time-course advantage is invaluable for the validation of this study. Using an "early cohort" in the mix can address specific questions that cannot be explained by using a "delayed cohort". Therefore, the "early cohort" allows to witness alterations in the dead body as early as possible, as well as to follow how this has changed over time and recognize how features are influenced by autolysis. Subsequently, the outcome is having a superimposition of the pre-existing disease – be it an inflammation, infarction or tumor. This is then superimposed by the effects of autolysis. The intention is to avoid making studies on autolysis in pathology since the primary aim is to discover the underlying disease.

EXPECTED RESULTS AND HYPOTHESIS

The focus and strengths of this study are on novel integrative and interactive machine learning approaches, which can complement and extend current deep learning approaches. We also envision helping to solve problems that would occur when comparing radiology data with pathology data. With this unique opportunity, we can compare and contrast the digital pathology with the radiology to proceed further in viewing these images; the aspiration of many radiologists. We have an advantage in the ability to collect tissue from the deceased, as opposed to the case of living subjects. Additionally, we can access the lungs, the heart and the brain (if needed); which is a unique opportunity that drives the imaging field in general, not only in the context of virtopsy.

The ability to compare radiology data with histology and digital pathology allows us to see how far the principles of machine learning image analysis can be applied - with different technologies and different imaging types. We create a biobanking library of digital images together with material data, which makes it a compelling resource. These are major factors of the core scientific approach that increase the likelihood of success.

This virtopsy proposal aims to use autopsies while incorporating new technologies - such as medical imaging - to improve how infectious diseases are examined. This research will help to understand the processes of diseases and evaluate both clinical diagnosis and treatment. It will enhance understanding in respect of therapy resistance by cancer patients with minimal tumor burden, and even the cause of death in such patients. Essential baseline data outcomes will quantify and compare tumor mass in cases involving diverse patients (according to their metabolic conditions). We hypothesize that virtopsy is dramatically more advantageous at mapping cardiovascular alterations, hemorrhages, and misuse of medical devices.[7] As a result of its strong application opportunities and high impact, this study is scientifically significant and innovative – particularly the substantial savings in time and finance, and delivery of rapid reliable data and diagnosis:

- Virtopsy provides new possibilities to clarify the pathogenesis of specific diseases, particularly those which affect organs or clinical conditions that impede in vivo tissue sampling. This is predominantly relevant to postmortem sampling of selected areas of the brain, that shed light on the biology of a variety of mental disorders such as dementia, Alzheimer's disease, schizophrenia and depression;

- In the recent episodes of Ebola, portable ultrasound imaging may have been applied in remote regions of Africa. In addition, this application may have led to a faster identification of Zika virus as being the cause of microcephaly during the recent outbreak in Brazil. It may have reduced the number of Ebola and microcephaly cases worldwide;

- In circumstances where corpses are suspected to have high-risk pathogens or infectious diseases, imaging autopsies will protect medical staff. The innovative methods eliminate the risks of contamination associated with the use of knives, needles, and scissors in conventional procedures;

- Virtopsy also obviates negative outcomes for DNA sequencing and pathogen detection, particularly in biopsies required for immune testing;

- This project addresses many unanswered questions in digital pathology. These relate to the lack of international literature on large-scale programs. Such programmes that compare MRI data with molecular phenotypes in an incisive, coordinated way that is expounded in this study;

- Image-guided biopsies have become an important factor in reference to the ethics of autopsies. They will detect the cause of death in the majority of cases; in doing so they improve quality control and diminish the economic burden of conventional autopsies;[7]

- The results from this project are highly reproducible in other countries. Medical imaging autopsy is a valid alternative method, and it will increase autopsy

acceptance rates,[7] particularly among cultures that oppose autopsies on ethical and/or religious grounds.

In summary, we plan to combine all the aforementioned diverse steps of autopsy and data analysis. The anticipated result will provide a new method of analyzing digital data from virtopsy and digital pathology; identifying improvements in diagnostic techniques, and assisting physicians with finding the correct diagnosis. We are confident that the entirety of our multi-faceted approach will provide the desired results and advance the state-of-the-art.

ACKNOWLEDGEMENTS

REFERENCES

1. Holzinger A, Plass M, Holzinger K, Crişan GC, Pintea CM, Palade V. Towards interactive Machine Learning (iML): applying ant colony algorithms to solve the traveling salesman problem with the human-in-the-loop approach. In: Buccafurri F, Holzinger A, Kieseberg P, Tjoa A, Weippl E, editors. Availability, reliability, and security in information systems. Cham: Springer; 2016. p. 81-95. (Lecture Notes in Computer Science; 9817). CD-ARES 2016. http://dx.doi.org/10.1007/978-3-319-45507-56.

2. Cacciatore S, Hu X, Viertler C, et al. Effects of intra- and post-operative ischemia on the metabolic profile of clinical liver tissue specimens monitored by NMR. J Proteome Res. 2013;12(12):5723-9. PMid:24124761. http://dx.doi.org/10.1021/pr400702d.

3. Holzinger A. Interactive machine learning for health informatics: when do we need the human-in-the-loop? Brain Informatics. 2016;3(2):119-31. PMid:27747607. http://dx.doi.org/10.1007/s40708-016-0042-6.

4. Holzinger A. Introduction to Machine Learning & Knowledge Extraction (MAKE). Mach Learn Knowl Extr 2017;1(1):1. http://dx.doi.org/10.3390/make1010001.

5. Wichmann D, Obbelode F, Vogel H, et al. Virtual autopsy as an alternative to traditional medical autopsy in the intensive care unit. Ann Intern Med. 2012;156(2):123-130. PMid:22250143. http://dx.doi.org/10.7326/0003-4819-156-2-201201170-00008.

6. Wichmann D, Heinemann A, Weinberg C, et al. Virtual autopsy with multiphase postmortem computed tomographic angiography versus traditional medical autopsy to investigate unexpected deaths of hospitalized patients: a cohort study. Ann Intern Med. 2014;160(8):534-41. PMid:24733194. http://dx.doi.org/10.7326/M13-2211.

7. Saldiva PHN. Minimally invasive autopsies: a promise to revive the procedure. Autops Case Rep. 2014;4(3):1-3. PMid:28573111. http://dx.doi.org/10.4322/acr.2014.021.

Correspondence
Shane O'Sullivan
Department of Pathology - Hospital das Clínicas - Faculdade de Medicina - Universidade de São Paulo (USP)
Av. Enéas Carvalho de Aguiar, 155 – São Paulo/SP – Brazil
CEP: 05403-000
Phone: +55 (11) 3061-7173
doctorshaneosullivan@gmail.com