DMP title

Project Name Automated Analysis of Kidney Transplant Biopsies - DMP title

Project Identifier u0144635

Principal Investigator / Researcher Omer Granoviter

Description This project will develop a data-driven Artificial intelligence framework based on deep learning that will be fundamental to personalized diagnosis and prognosis, aiming to better the lives of patients suffering from chronic Kidney desease. More specifically, we aim to classify subtypes of rejection and predict the possibility of a transplanted kidney being rejected. The data was collected between the years 2004 and 2020 for the purpose of advancing Computational Pathology for Kidney transplant clinic usage and research.

Institution KU Leuven

1. General Information

Name applicant

Omer Granoviter

FWO Project Number & Title

Project Number: 1S61822N

Project Title: Automated Analysis of Kidney Transplant Biopsies

Affiliation

- KU Leuven

2. Data description

Will you generate/collect new data and/or make use of existing data?

- Reuse existing data

Describe in detail the origin, type and format of the data (per dataset) and its (estimated) volume. This may be easiest in a table (see example) or as a data flow and per WP or objective of the project. If you reuse existing data, specify the source of these data. Distinguish data types (the kind of content) from data formats (the technical format).

Type of data	Format	Volume	How created
Whole Slide Images (WSI)	.tiff	5 TB	Light microscopy images of Kidney biopsies.
Whole Slide Image annotations	.csv	5 MB	Annotations of Renal-Pathologists of each WSI with lesion scores and binary rejection classification.
Metadata	.csv	5 MB	For each patient in the database, metadata including age, sex, and medical conditions.
Graft functional data	.csv	5 MB	Serum creatine and proteinuria values.

3. Legal and ethical issues

Will you use personal data? If so, shortly describe the kind of personal data you will use. Add the reference to your file in KU Leuven's Register of Data Processing for Research and Public Service Purposes (PRET application). Be aware that registering the fact that you process personal data is a legal obligation.

- Yes

The project is covered by an ethical approval from the Ethical Committee of the University Hospitals Leuven The study number is S64006 entitled “An integrative clinicopathological decision support system for kidney transplant rejection”.

To quote the ethical approval: " The study is conducted in compliance with the principles of the Declaration of Helsinki (64th WMA General Assembly, Fortaleza, Brazil, October 2013), the principles of GCP and all of the applicable regulatory requirements. This protocol and related documents are submitted for approval to the EC of the University Hospitals Leuven, Herestraat 49, 3000 Leuven, Belgium. "

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? If so, add the reference to the formal approval by the relevant ethical review committee(s)

- No

To quote the ethical approval:

" All data in the databases and metadata will are pseudonymized. The Investigator and the Participating Site will treat all information and data relating to the Study disclosed to Participating Site and/or Investigator in this Study as confidential and shall not disclose such information to any third parties or use such information for any purpose other than the performance of the Study. The collection, processing, and disclosure of personal data, such as patient health and medical information is subject to compliance with applicable personal data protection and the processing of personal data (Regulation (EU) 2016/679 also referred to as the General Data Protection Regulation ("GDPR") and the Belgian Law of July 30, 2018, on the protection of natural persons with regard to the processing of personal data)."

Does your work possibly result in research data with potential for tech transfer and valorisation? Will IP restrictions be claimed for the data you created? If so, for what data and which restrictions will be asserted?

- Yes

1. The algorithmic models that are generated through this study will be published and the source code will be uploaded under relevant licensing.
2. The clinical support systems will be published and the source code will be uploaded under relevant licensing.

Do existing 3rd party agreements restrict dissemination or exploitation of the data you (re)use? If so, to what data do they relate and what restrictions are in place?

- Yes

External cohort data for the model validation process that might be shared through different research facilities will be covered under Data transfer agreements.

4. Documentation and metadata

What documentation will be provided to enable reuse of the data collected/generated in this project?

1. For the objective of classification, hyperparameter tuning is done to create the baseline performance of the model.
The experiments of the highest metrics for all the different models are saved on a local Hard-drive. For each such experiment, there is a text file (.txt) containing all the parameters used in the model. A .csv file containing all the Whole slide images used for the train, test, and validation for each cross-validation split of the experiment and a .csv file containing all the metrics.

2. A text file containing the information of the classes used for each experiment will be added to each experiment folder.

3. Each article published will be accompanied by a git repository containing the code and the default parameters used to derive the final metrics published in the paper, enabling the reproduction of the results.

Will a metadata standard be used? If so, describe in detail which standard will be

used. If no, state in detail which metadata will be created to make the data easy/easier to find and reuse.

- No

5. Data storage and backup during the FWO project

Where will the data be stored?

1. The database as exposed to the researcher is anonymized. The only patient information the researcher is exposed to is a patient identifier number (unique per subject), and the biopsy number, type of staining, lesion scores, and metadata. No information can lead the researcher to the un-anonymous information of the subjects.

2. Data is stored in the KUL drive (https://drives.kuleuven.be/) and the previously defined anonymized form. The data stored in the drives is secure under the regulations of the university.

3. For collaborative work, outside researchers will be granted access to the online database following a data transfer agreement as stated in the study protocol S64006 (An integrative clinicopathological decision support system for kidney transplant rejection). To cite the study protocol: "Only pseudonymized data or scans, without link to the patient providing the data / images will be used. No physical patient material will be shared with these third parties".

4. For usage of distributed deep learning training on Condor the pseudonymized data is stored on secure Biomed servers. Following the end of the experiments, the data is deleted.

5. For local training and analysis the data is stored on a local computer SSD drive at KU Leuven. This data can only be accessed by the researcher and admin credentials.

How is backup of the data provided?

The data is stored on the KUL drive (https://drives.kuleuven.be/) with automatic daily back-up procedures.

Is there currently sufficient storage & backup capacity during the project? If yes, specify concisely. If no or insufficient storage or backup capacities are available then explain how this will be taken care of.

- Yes

What are the expected costs for data storage and back up during the project? How will these costs be covered?

Don't know, are there any costs?

Data security: how will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

The data is entirely pseudonymized. There is no danger of personal data leakage from the Whole slide image database, meta data and classification data. Additionally, the data is stored in a university-secured environment.

6. Data preservation after the FWO project

Which data will be retained for the expected 5 year period after the end of the project? In case only a selection of the data can/will be preserved, clearly state the reasons for this (legal or contractual restrictions, physical preservation issues, ...).

1. All the Biopsy slides.
2. All the metadata and labels will be preserved.
3. The code of the project will be preserved including steps to reproduce the results. Reproduction of results is easy and of low cost.
4. Specific state-of-the-art model results will be saved in a KU Leuven archive.

Where will the data be archived (= stored for the longer term)?

1. The data and meta-data will be stored on the KUL Drive (with automatic back-up procedures) for at least 10 years, conform the KU Leuven RDM policy.
2. The code including steps to reproduce results will be saved in designated git repositories.

What are the expected costs for data preservation during the retention period of 5 years? How will the costs be covered?

No additional costs are expected for the preservation of the data.

7. Data sharing and reuse

Are there any factors restricting or preventing the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)?

- Yes. Specify:

1. The previously described data might be shared with designated research facilities and collaborations with multiple international kidney transplant centers with similar expertise (e.g., University of Cologne, Prof. Jan Becker; Universities Amsterdam and Leiden, Prof. Jesper Kers; University of Lyon, Prof. Olivier Thaunat; University of Vienna, Dr. Nicolas Kosakowski and Dr. Zeljko Kikic; Necker Hospital Paris, Prof. Rabant).
2. The data sharing is constrained under Data transfer agreements that are signed will designated collaborators.

Which data will be made available after the end of the project?

The dataset will not be made publicly available as it is an IP of UZ/KU Leuven. It will be available following the signing of Data transfer agreements with specific parties.

Where/how will the data be made available for reuse?

- In a restricted access repository

1. The source code will be released to a git repository (e.g. GitHub, GitLab).
2. Data will be available upon request following the signing of a data-sharing agreement.

When will the data be made available?

- Upon publication of the research results

1. The data will be shared with specific parties following data-sharing agreements.
2. Specific project data such as source code and analysis scripts will be published upon publication.

Who will be able to access the data and under what conditions?

1. Access to the dataset will be granted upon request and following the signing of data transfer agreements.

What are the expected costs for data sharing? How will the costs be covered?

Data will be shared digitally. The costs of data sharing if any will exist will be covered by designated grants designed for the specific collaboration, or by the parties requesting the data.

8. Responsibilities

Who will be responsible for data documentation & metadata?

The data documentation and metadata are the responsibility of the Department of Pathology of the University Hospital Leuven.

Who will be responsible for data storage & back up during the project?

The data documentation and metadata are the responsibility of the Department of Pathology of the University Hospital Leuven.

Who will be responsible for ensuring data preservation and reuse ?

The responsibility for ensuring data preservation and reuse is of the Department of Pathology of the University Hospital Leuven.

Who bears the end responsibility for updating & implementing this DMP?

The PI bears the end responsibility of updating & implementing this DMP.