PKMine; AI for extracting input parameters from different sources to support PBK modelling

Marc Teunis[1]*, Marie Corradi[1], Ivo Djidrovski[2], Susana Proenca[3], Huan Yang[3], Thomas Luechtefeld[4], Thomas Hartung[5,6] Nynke Kramer[7]

1. University of Applied Sciences, Utrecht; 2. Utrecht University, Utrecht; 3. esqLABS GmbH, Saterland;
4. Insillica.co, Bathesda; 5. John's Hopkins University, Baltimore;
6. Center for Alternatives to Animal Testing, Baltimore;
7. Wageningen University & Research, Wageningen

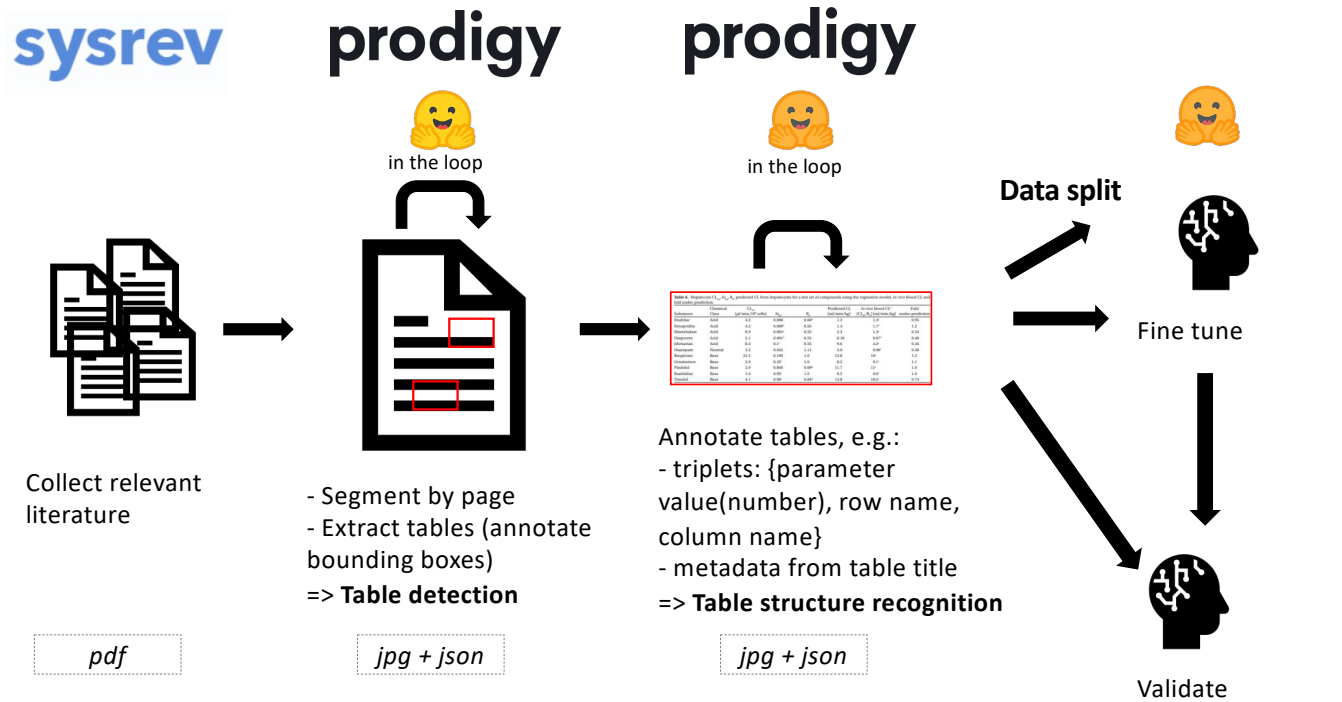
Aim

Design a workflow to collect PBPK model parameters and validation data in a structured way. We propose to finetune AI models to extract these data from text, figures, tables and databases. Extracted data will include:

- Physicochemical properties
- *In vitro* ADME parameters
- *In vivo* ADME parameters

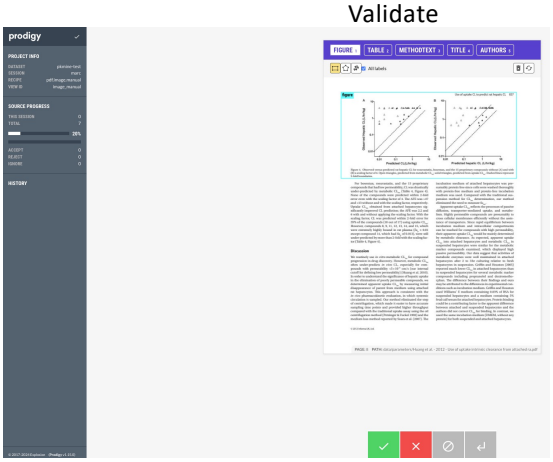
Output

Parameters in a format that maximises interoperability with PBK tools and software (e.g. PKSim, R).



Getting training data

Use annotation software **Prodigy** (explosion.ai) on PDF files collected in the paper: Louisse, J. et al (2020). Towards harmonization of test methods for in vitro hepatic clearance studies. *Toxicology in vitro; 63*, 104722.



Candidate models

TableTransformerModel
https://huggingface.co/docs/transformers/model_doc/table-transformer
class transformers.TableTransformerModel

TrOCRProcessor
https://huggingface.co/docs/_doc/trocr
class transformers.TrOCRProcessor

Abbreviations

ADME: Absorption, Distribution, Metabolism, Excretion
PBPK: Physiologically Based PharmacoKinetic modelling
TrOCR: Transformer Optical Character Recognition