Daily Log Project M.Sc. ECMM451

Student Name: Souradeep Sen (700054986)

Project Name: Survival Analysis of Heart Failure Patients

Internal Supervisor: Dr Ayah Helal

External Supervisor: Professor Krasimira Tsaneva

Daily Log of Activity

Date	Activity	Outcomes	Comments
3.2.23	First meeting with Prof Tsaneva	Requested to work on health data project. Given some papers to read using CPRD data and ML methods. CNN method vs statistical method (Hidden Markov Model) considered.	It may be interesting to look at the problem from two perspectives – one from deep learning, another from a more statistical angle.
10.2.23	Sent mail to Prof Tsaneva regarding updates after reading papers	Started reading books on survival analysis and Cox prop hazards model – they will form the base for supervised layers.	Ask for some papers on HMM. Capacity building needs to be done for maths knowledge.
16.3.23	Reached out to Prof Kadam (medical school) for CPRD data	Prof Kadam will share sample data soon. Actual data to come along at a later date.	Need to follow up with Prof Kadam about timelines.
22.3.23	First meeting with both Dr Helal and Prof Tsaneva	Brief updates on where I am with the project. Asked how to get access to data.	Ethics platform is not working. Reported to Dr Helal and requested help.
14.4.23	Completed MIMIC-III training; signed DUA	Waiting to be granted full access to the MIMIC-III data on GCP/AWS (yet to decide which to use).	Looking at sample MIMIC-III data in the meantime.
15.4.23		Got access to entire MIMIC-III data.	Need to learn how to access it via local server/ BigQuery/ AWS
24.4.23	Sent first draft to of proposal to		

	Prof Kadam, Prof Tsaneva and Dr Helal		
26.4.23		Received feedback from Dr Helal.	Need to incorporate feedback.
28.4.23	Incorporated feedback and resent to supervisors		
28.4.23	Filled out ethics form		Need to get some help from Dr Helal regarding the form filling. Unsure about all the fields.
29.4.23	Ethics form submitted for first triage		
9.5.23	Additional information requested by Joanna Parsons on WorkTribe; forwarded to supervisors		
15.5.23	Email exchange with Prof Kadam and Prof Bailey (medical school)	CPRD data is apparently from a cancer cohort.	May need to drop CPRD in favour of MIMIC-III.
Gap due to exams			
29.5.23	Level set with supervisors after exams ended; Create ethics form for MIMIC-IV data	Inform of the challenges with ethics approval and CPRD data acquisition. Decide to utilize open-source MIMIC-IV data, instead of CPRD data.	
30.5.23	Set up access to MIMIC-III via BigQuery		Further exploration needed.
31.5.23	Learned about venv, pytest		Will use testing and virtual environments in my projects to

			demonstrate better software practices.
1.6.23	Learned about MIMIC-IV		Covers data till 2019. Will be more useful.
3.6.23	Set up MIMIC-IV on local postgres	Able to explore data from within pgAdmin 4 (the DBMS manager that comes with postgres).	Connect with Python required
4.6.23	Using SQLAlchemy to connect postgres server to Python	Able to explore data from notebooks/ code editor.	Begin exploration of data
8.6.23	Meeting with Dr Helal. Discuss primary vs secondary aims of project and related prioritization	Decide to drop secondary aim of exploring Hidden Markov Models for now.	
9.6.23	Discuss data storage options with Prof Bailey	CPRD - Ethics letter received from Prof Bailey. Submitted to WorkTribe.	
15.6.23	Access to Research Data Storage is sought. Guidance is requested from the Information Governance team		
16.6.23	RDS access is set up	Can view the CPRD data now.	The data requires exploration.
20.6.23		Ethics Approval committee requests additional information, including correspondence with Information Governance team.	
21.6.23	Basic exploration carried out on MIMIC-IV data	Patient cohort created. Including all patients with a Heart Failure Diagnosis.	Need to get patient covariates such as medication, lab tests, BMI, smoking status, weight etc.

22.6.23	Capacity building with PyTorch	Learning to use modularized code to build blocks for a neural network.	
23.6.23	Simple feed-forward PyTorch model set up using dummy data		Need to find how to feed patient covariates (time-static and time-varying) to a neural network. Need questions to the following: What should be the format of the output? What loss function should be used?
26.6.23	Literature review searching for suitable loss functions	This paper has a suitable loss function - Kvamme, Håvard, and Ørnulf Borgan. 2019. "Continuous and Discrete-Time Survival Prediction with Neural Networks." arXiv. http://arxiv.org/abs/1910.06724.	
29.6.23	Discussed feedback for proposal with supervisors	Decision is made to introduce uncertainty quantification as part of the project.	Need to explore methods pertaining to uncertainty quantification.
1.7.23	Begin writing thesis		
4.7.23	Test out off-the-shelf fitters		Random Survival Forest is very powerful, even with default parameters. Cox Proportional Hazards model also has high discriminatory power and great calibration for MIMIC-IV data. Off-the-shelf neural network models are not that great.
10.7.23	Discussion about CPRD data with Prof Bailey	Death date is not effectively captured in CPRD data	This rules out CPRD as a potential choice of data for the project, as date of death is important for survival analysis

15.7.23	Application approved for Triage by supervisor		
17.7.23	Application staged for review by Ethics committee		Further correspondence from Information Governance team is required.
20.7.23	Capacity building for uncertainty quantification of neural networks	Literature review yields Gaussian approximation to Neural Nets as an effective route Eldan, Ronen, Dan Mikulincer, and Tselil Schramm. 2021. "Non-Asymptotic Approximations of Neural Networks by Gaussian Processes." arXiv. http://arxiv.org/abs/2102.08668. Lee, Jaehoon, Yasaman Bahri, Roman Novak, Samuel S. Schoenholz, Jeffrey Pennington, and Jascha Sohl-Dickstein. 2018. "Deep Neural Networks as Gaussian Processes." arXiv. http://arxiv.org/abs/1711.00165.	Try to search for a less mathematically involved method – Monte Carlo dropout may be suitable.
23.7.23	Added MC dropout to implementation		Try to add explanation capability to implementation
26.7.23	Added SHAP to Time-Invariant Survival		
General experimentation with own models and off-the-shelf models			
4.8.23	Finish writing first draft		Share with supervisors for comments
5.8.23	Run experiments with all models (neural network, tree-based, traditional)	Time-Variant methodology is beating out other models in terms of c-index	It is the worst-performing in terms of Brier score
7.8.23	Share draft with supervisors		Await feedback

8.8.23	Application approved for Triage by supervisor		
9.8.23	Application approved by ethics committee		
10.8.23	Create a presentation for the thesis		
11.8.23	Presentation submission		
13.8.23	Final run of experiments	Containerization of experiments and example code	
14.8.23	Received feedback from supervisor		
17.8.23	Incorporation of feedback and submission		