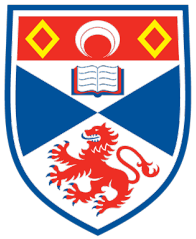


On Predicting the Outcomes of Chemotherapy Treatment in Breast Cancer

Agastya Silvina, Juliana Bowles, and Peter Hall

Outline

- Introduction
- Related Work
- Data Analysis
- Models Creation
- Result
- Conclusion

Introduction

- Cancer
  - is a mutation caused by an abnormal reproduction of cells.
  - can occur in different organs (e.g., **breast**, lungs, bone, etc.)
- Treatments vary from surgery with **chemotherapy** and/or radiotherapy (i.e. usually take a long time and in sequence)
  - However some treatments are toxic and expensive
- We **compared several different techniques (Markov Model, HMM, RF, RNN)** applied to the same data set to **predict the toxicity outcome of different treatment options**

Related Work

- Many ongoing research looks at prediction of cancer treatment outcomes
  - Bayesian Logistic Regression (Subramani et al.)
  - Random Forest (Hui-Ling Chen et al.)
  - SVM (Nguyen et al.)
- We have added **HMM** and **RNN** (common in different fields like NLP) to explore what their benefits might be

NHS

predict
breast cancer

Outline

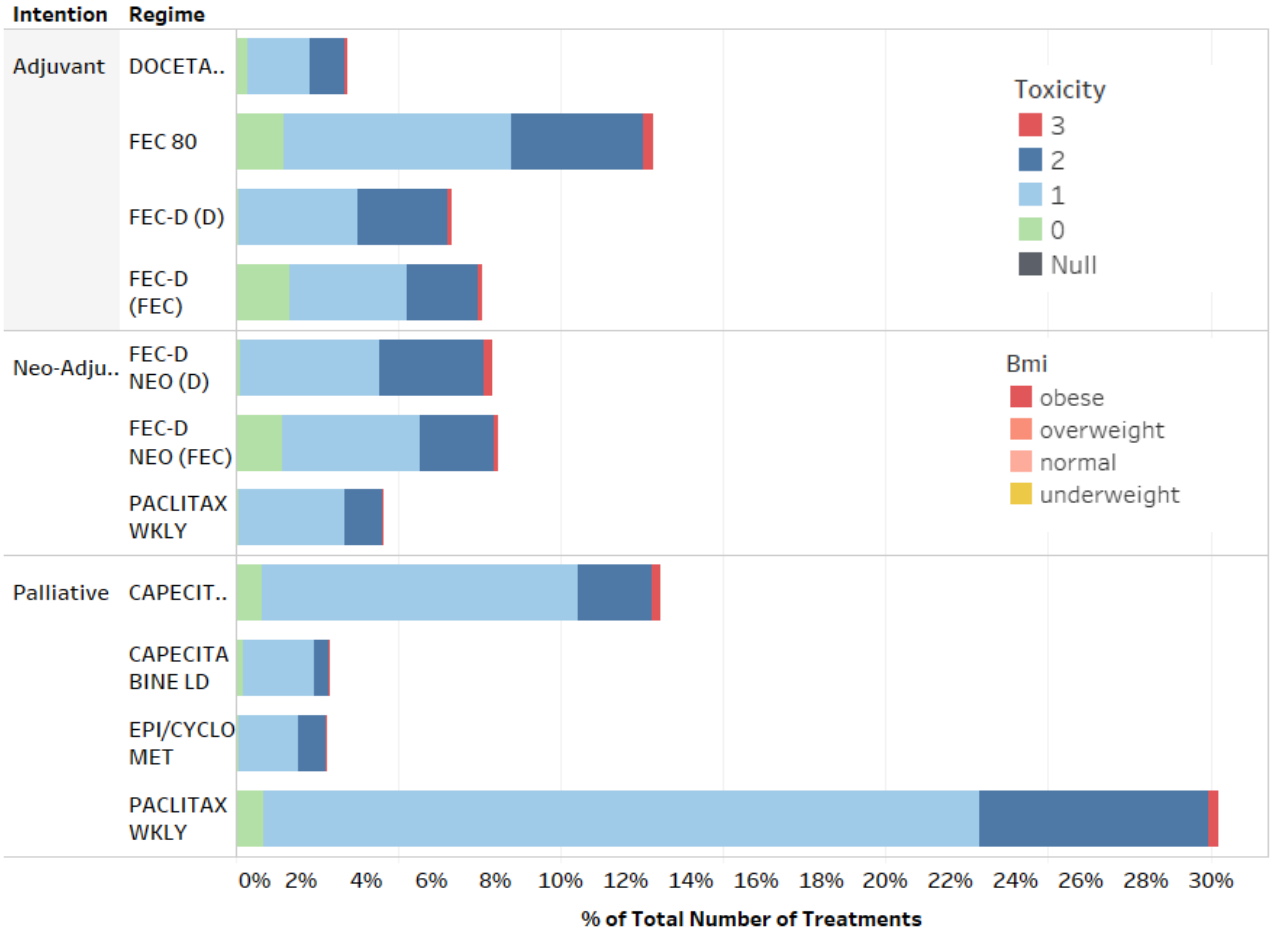
- Introduction
- Related Work
- **Data Analysis**
- Models Creation
- Result
- Conclusion

Data Analysis

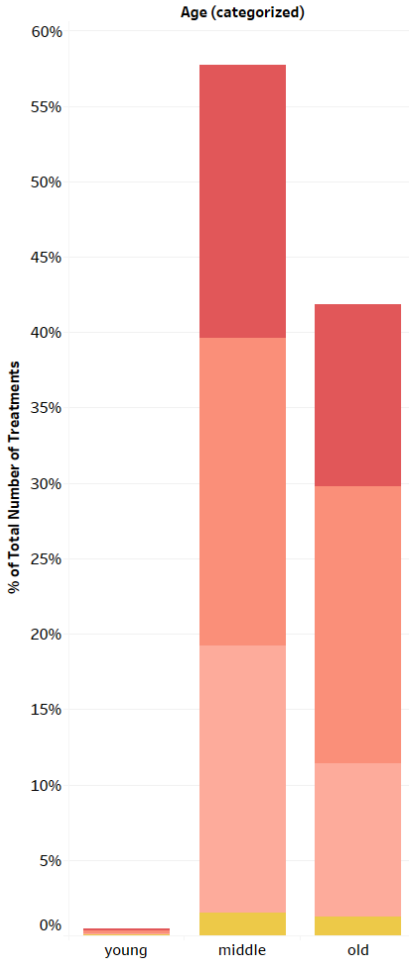
- We use a data extraction from an oncology department in Scotland
  - 3 years (2014 - 2016)
  - Includes various observations concerning breast cancer treatments (e.g. **intention, regime, cycles**), recorded side effects (here, **toxicity level**), and patient characteristics (e.g. **age, BMI, performance status**).

Intention	Number of Treatments	Number of Patients
Adjuvant	1209	205
Neo-adjuvant	1855	382
Palliative	2752	213

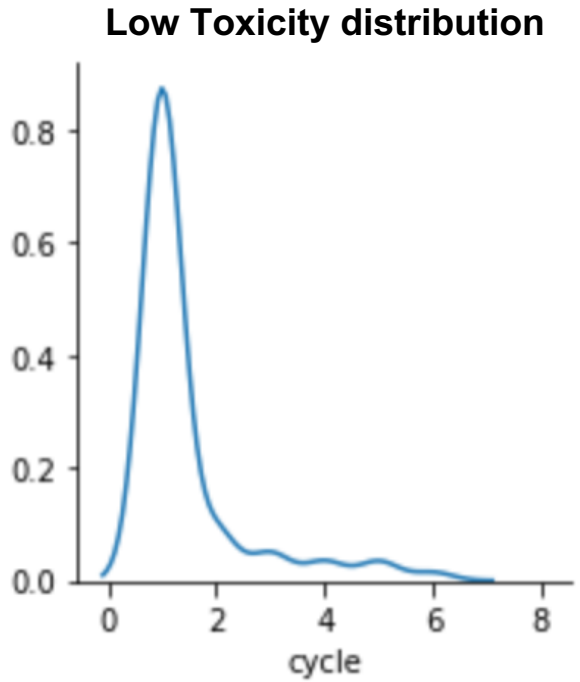
Toxicity proportion

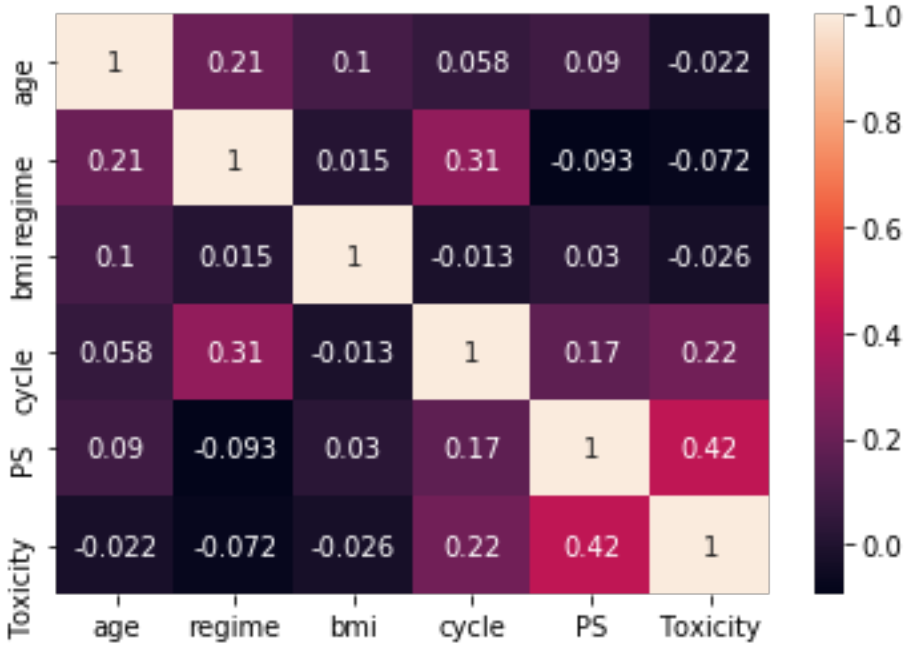


Age - BMI



Features correlation

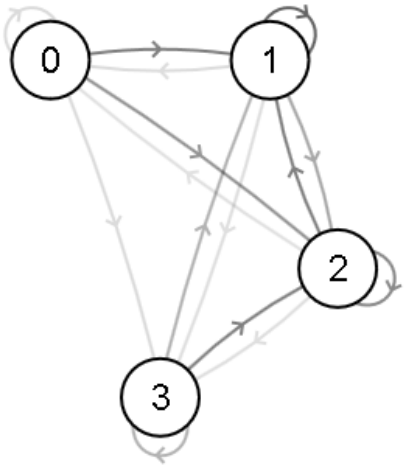




Outline

- Introduction
- Related Work
- Data Analysis
- **Models Creation**
- Result
- Conclusion

Markov Model (MM)



The Adjuvant therapy Markov Chain

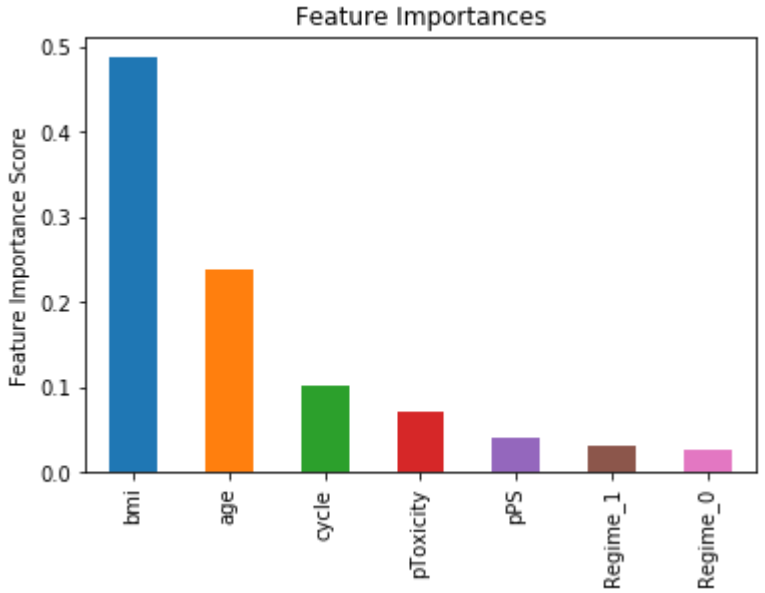
- A stochastic model with the assumption that a future state only depends on the current state.
- ***0, 1, 2, 3*** denote the toxicity state(i.e. No toxicity, Low, Medium, High)

Hidden Markov Model (HMM)

- Based on augmenting a Markov chain to observe the hidden states of events
- Our HMM components:
  - States: **T0, T1, T2, T3**
  - Transitions: from **T0 to T1**, from **T1 to T3**, etc
  - Observations: **cycle, age, BMI, regime** (categorised and coded)**.** For example, 1-2-3-1 denotes the observation for an overweight patient who gets the FEC-D (D) in their first cycle and is aged less than 50 years
  - Emissions: the probability of the observations generated from the toxicity state

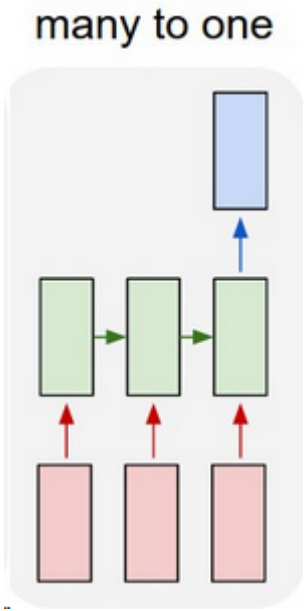
Random Forest

- An ensemble of decision trees for solving classification problems.
- We created three RF models for each treatment intention (i.e., adjuvant, neoadjuvant, palliative)
    - Predictors: **age, BMI, Regime, cycle, previous toxicity,** and previous performance status.
    - Outcome: **Patients' toxicity**



Recurrent Neural Network (RNN)

- A class of NN where connections between nodes form a directed graph along a temporal sequence.
- Implemented using **tensorflow LSTM** module.
- Used similar features as for our RF model. However, **we do not use** the previous **performance status** and **previous toxicity fields.**



Model Comparison

Markov Model

- A Stochastic model
- **Has no state memory**

Random Forest

- Ensemble learning
- **Has one state memory** with previous toxicity field

Hidden Markov model (HMM)

- Statistical Markov Model
- **Has one state memory**
- Based on POS Tagging

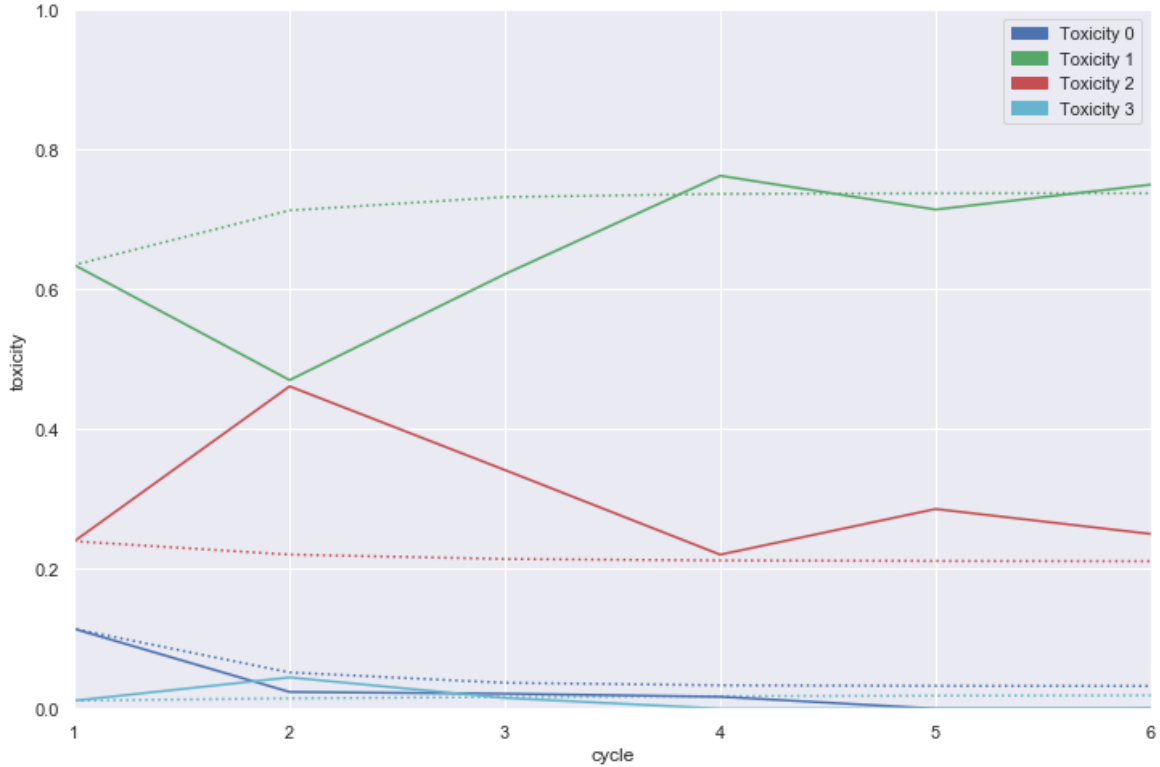
Recurrent Neural Network

- Sequential ANN
- **Memorises all states**

Outline

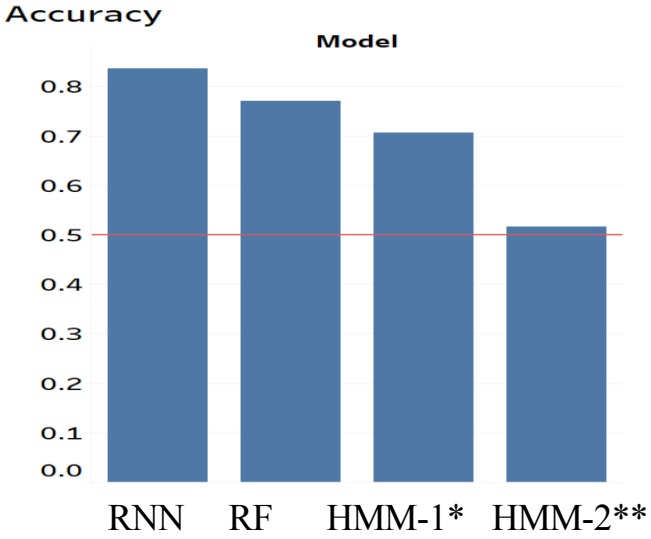
- Introduction
- Related Work
- Data Analysis
- Models Creation
- **Result**
- Conclusion

Markov Model

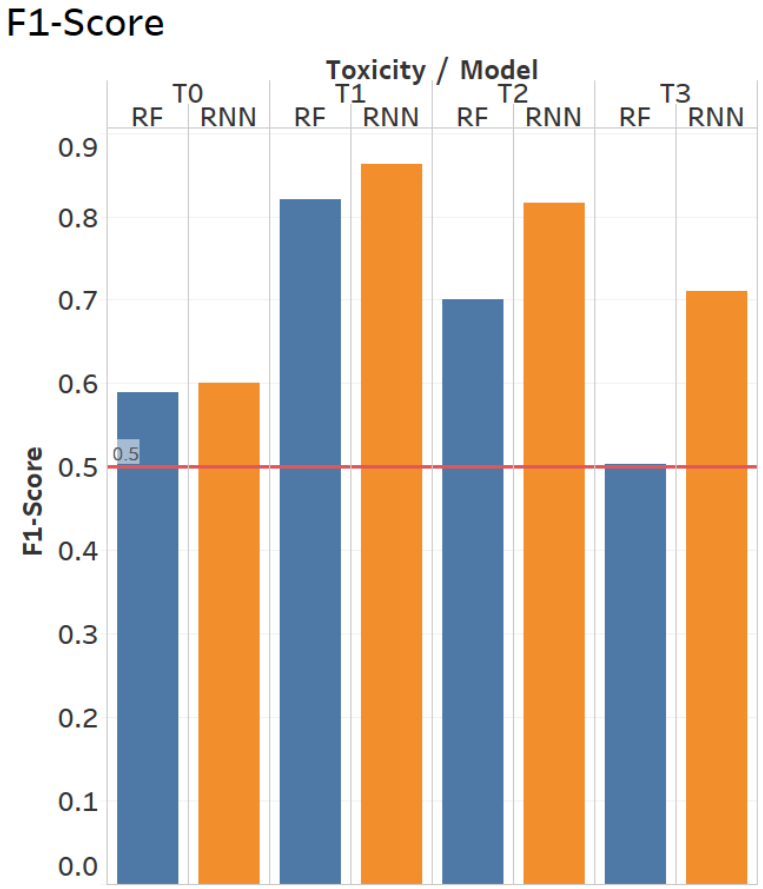


Palliative treatments distribution

Classifier



- **RNN**: Recurrent Neural Network
- **RF**: Random Forest
- ***HMM-1**: Hidden Markov Model-mid treatments
- ****HMM-2**: Hidden Markov Model-init & end cycle (i.e., cycle = 0 or cycle = end of treatment)



Outline

- Introduction
- Related Work
- Data Analysis
- Models Creation
- Result
- **Conclusion**

Conclusion

- Our classifiers can predict the toxicity outcome of the chemotherapy outcomes with around **0.8 - 0.85** accuracy.
- The RNN model performed better than all other models because it considers all the history of patients' treatments
- In contrast to the MM, the classifiers are more tailored for an individual patient.
    - Both the MM and the classifiers are a complement to each other.

Future Work

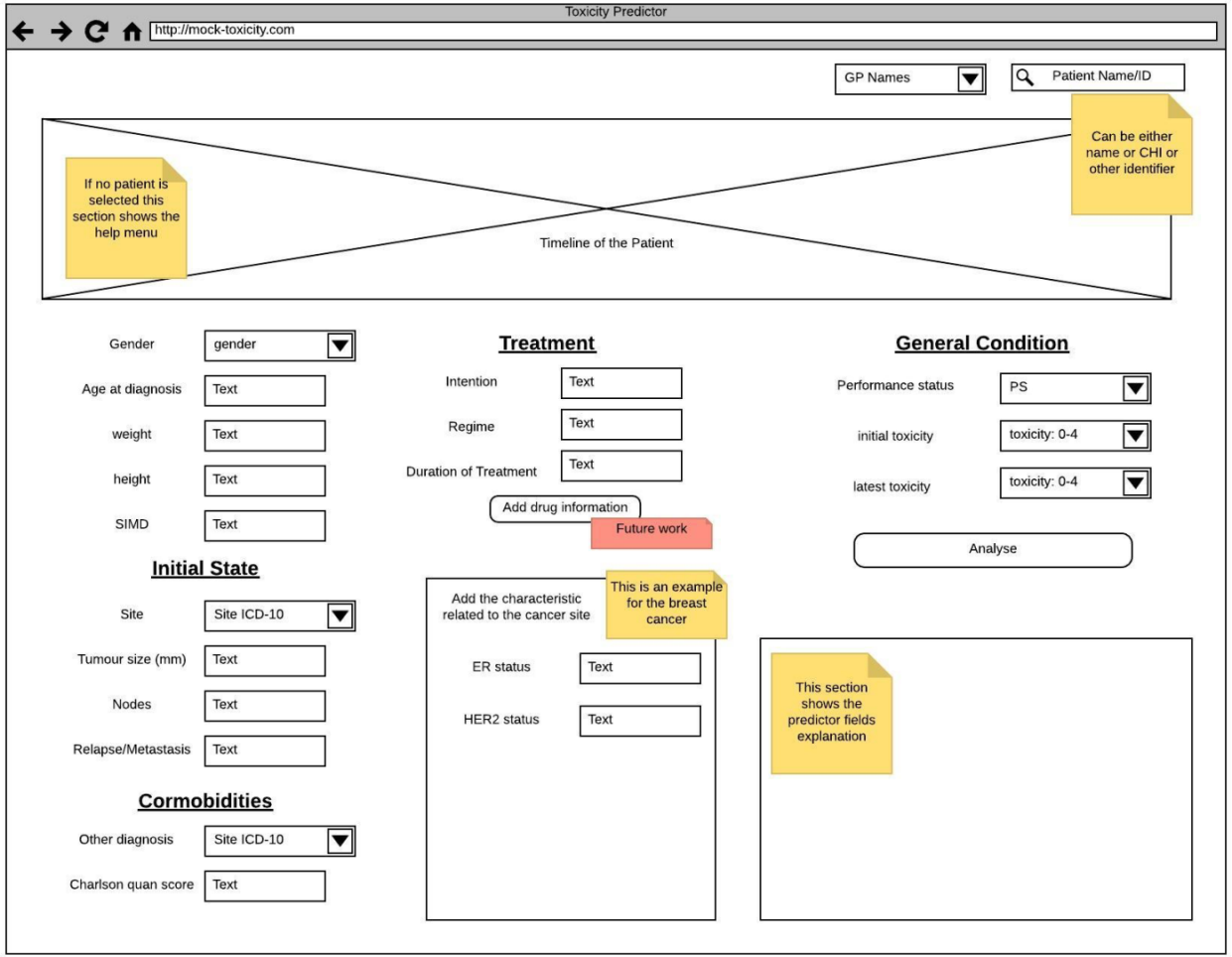
- Improve the accuracy further by integrating more data regarding the cancer characteristics and patients' comorbidity – usually ignored.
- Create a dashboard and/or reporting system which can be helpful to the clinical oncologist as a second opinion to decide which regimen is more suitable for an individual patient.

Future Work

Our aim is to use a synthetic data to further develop the application.

[Dashboard]

https://breast.predict.nhs.uk/tool



Open Issues and further work

- **Missing Values**
  - Solution**:** regression, removing some instances
- **Class imbalance**
  - Solution: duplication for some classes
- **Overfitting**
  - Solution: Cross Validation, using more data (will be provided by IBM - **SERUM** project)

Thank you

Q/A