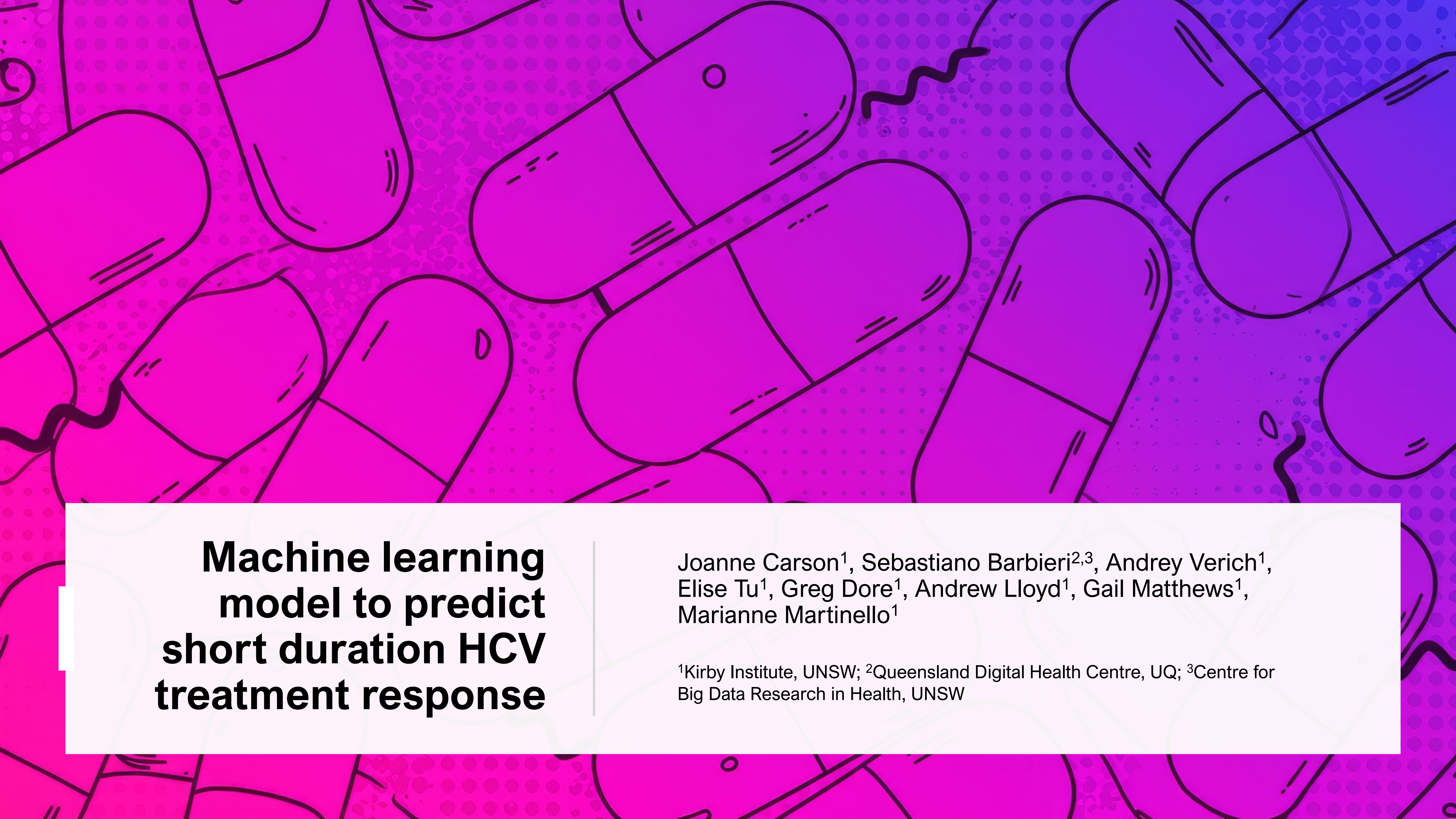


Machine learning model to predict short duration HCV treatment response

Joanne Carson[1], Sebastiano Barbieri[2,3], Andrey Verich[1], Elise Tu[1], Greg Dore[1], Andrew Lloyd[1], Gail Matthews[1], Marianne Martinello[1]

[1]Kirby Institute, UNSW; [2]Queensland Digital Health Centre, UQ; [3]Centre for Big Data Research in Health, UNSW

Real people, real action, real results

- **Real people:** Populations key to HCV elimination may face challenges completing standard duration (8-12 week) direct acting antivirals (DAAs)

- **Real action:** What if we could we deliver personalized short-duration (4-6 weeks) treatment that is highly effective? **Reverse the paradigm- match treatment to people, not people to protocols?**

- **Real results:** We aimed to develop a machine learning model to predict response to short-duration (4–6 weeks) HCV treatment using baseline clinical factors.

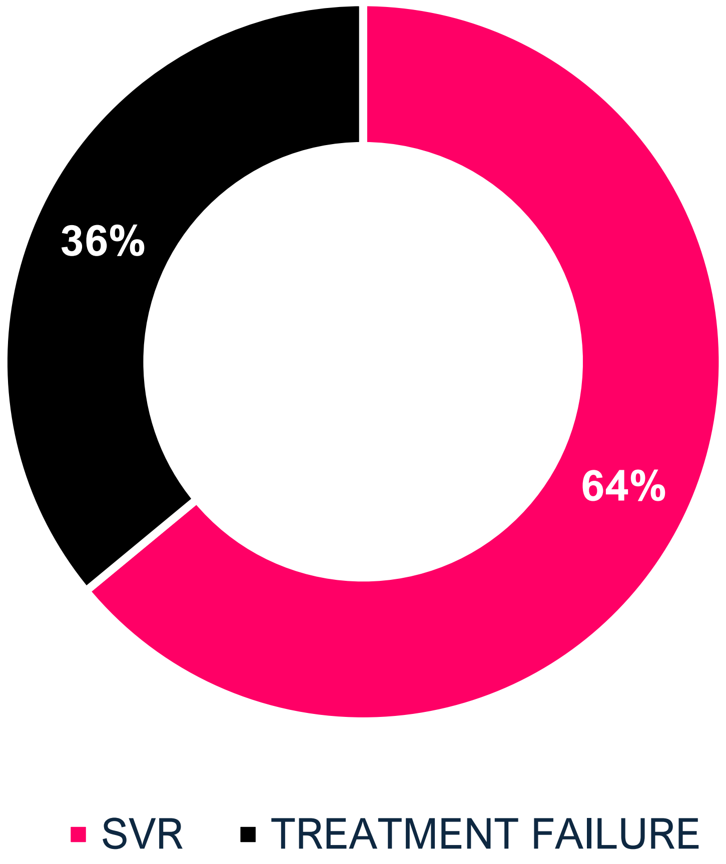
Data sources

To build machine learning models you need high quality data – sourcing this is more challenging than you might think!

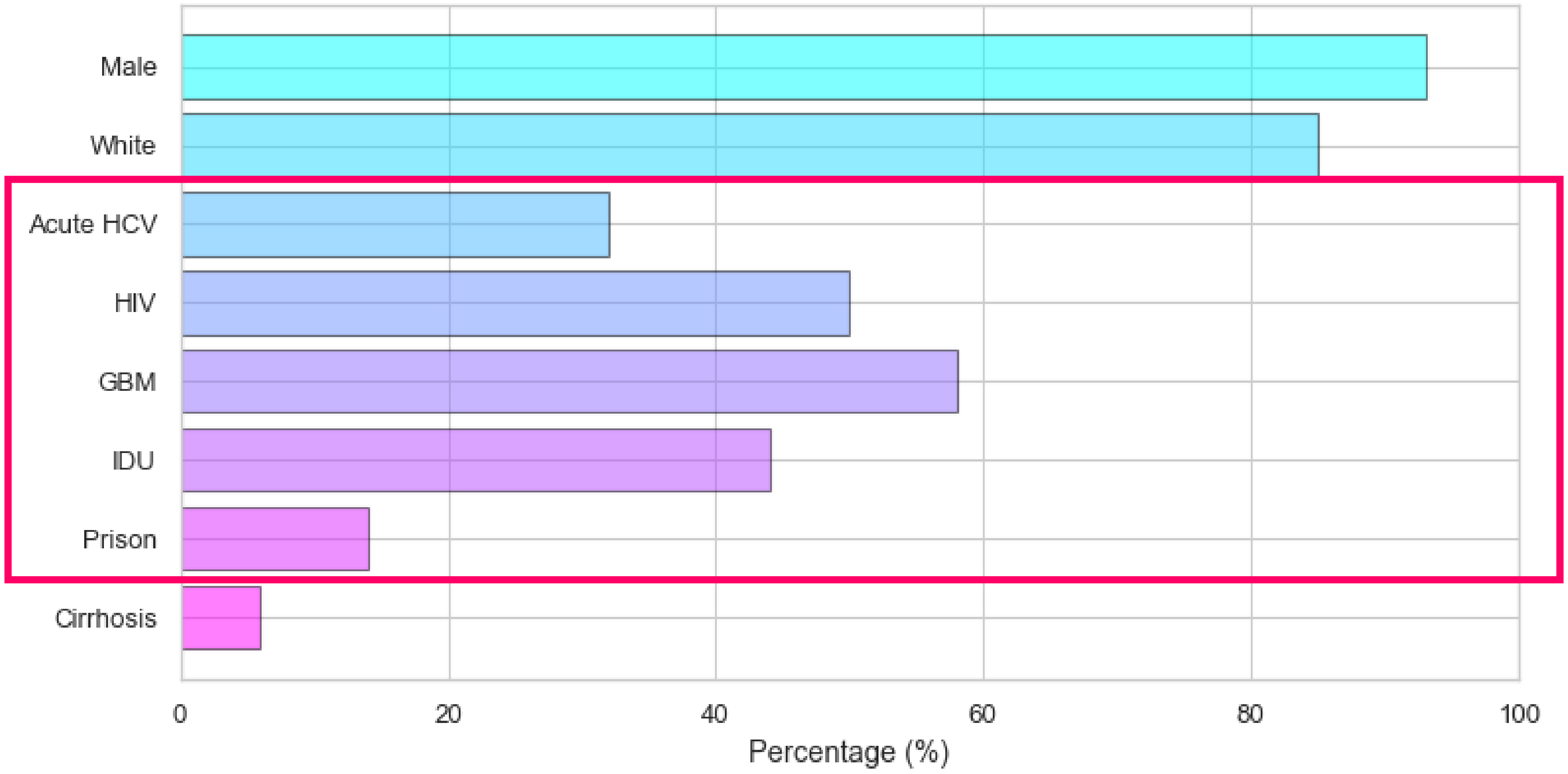
9 studies→ 264 short duration with baseline, clinical, treatment and outcome data.

DAAs were included as classes due to the small sample:

- **NS5A+NS5B**: sofosbuvir-velpatasvir (n=134); sofosbuvir-ledipasvir (n=17); sofosbuvir+daclastasvr (n=11)
- **PI+NS5A**: glecaprevir-pibrentasvir (n=70)
- **PI+NS5A+NS5B:** ombitasvir-paritaprevir-ritonavir+dasabuvir (n=30); glecaprevir-pibrentasvir+sofosbuvir (n=1); grazoprevir-elbasvir +sofosbuvir (n=1)



Population characteristics



Includes key populations who may be at risk of HCV or disruptions to treatment.

Developing models

Seven different machine learning models were evaluated to find the one that performed best.

Nested cross-validation was employed to optimize hyperparameters and assess performance – *reduces risk of data leakage, overfitting & provides more realistic estimates of performance*

Generated synthetic failure data based on the hardest examples to classify (training folds only) – *helps with small samples & imbalanced data.*

Python code used to develop models available at: https://github.com/jojocarson/hcv_tx_response

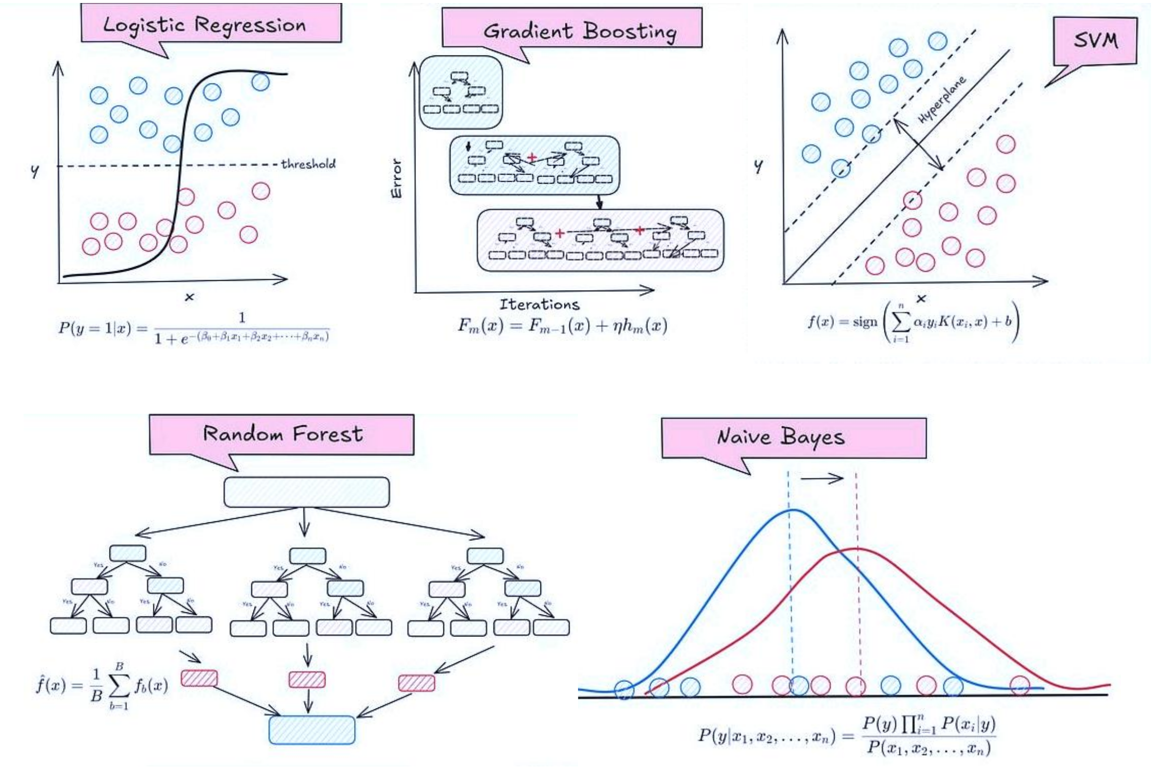
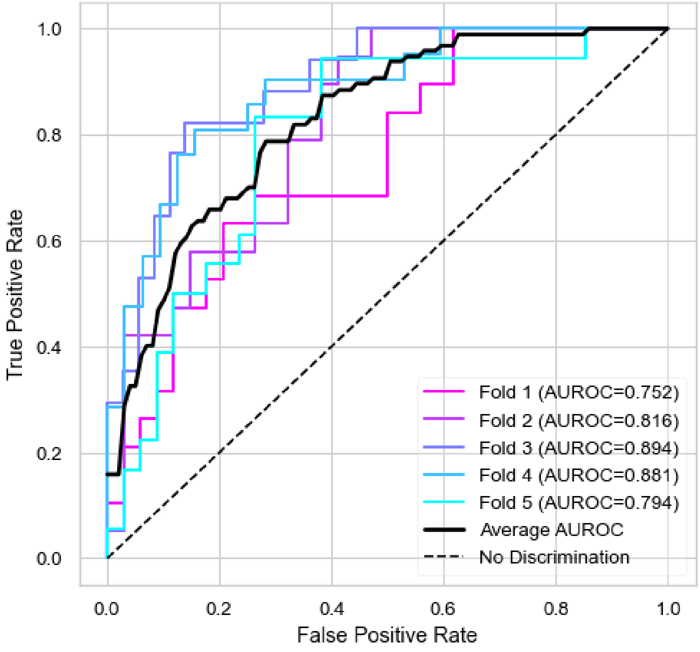
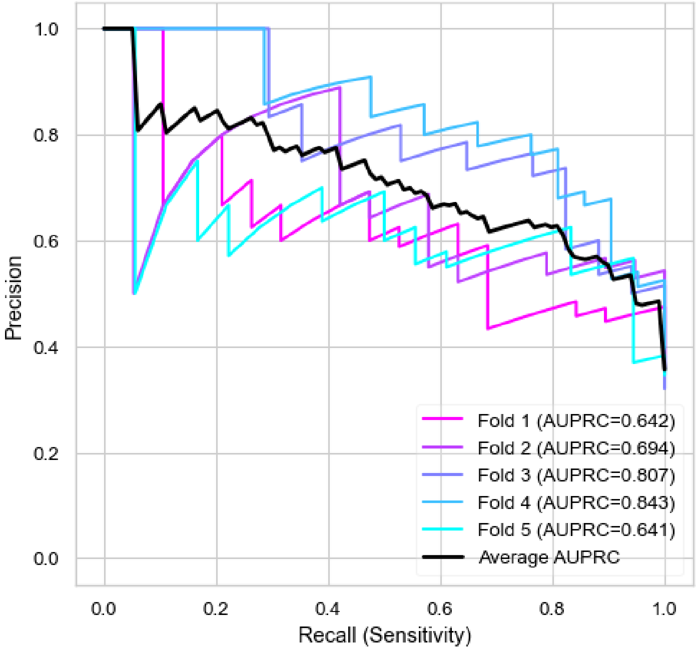


Figure. Examples of some of the machine learning architectures evaluated.

Best performing model

WINNER!! **Elastic Net** regularized logistic regression with L1 & L2 penalties.





A simple Elastic Net performed better than more complex models.

AUROC: **83%**

Shows how well the model can tell apart people who will and won't respond to short-duration DAAs.

AUPRC: **73%**

Shows how well the model finds true short duration failures without too many false alarms.

Abbreviations: AUROC, Area Under Receiver Operator Characteristics; AUPRC, Area Under Precision Recall Curve

Feature importance

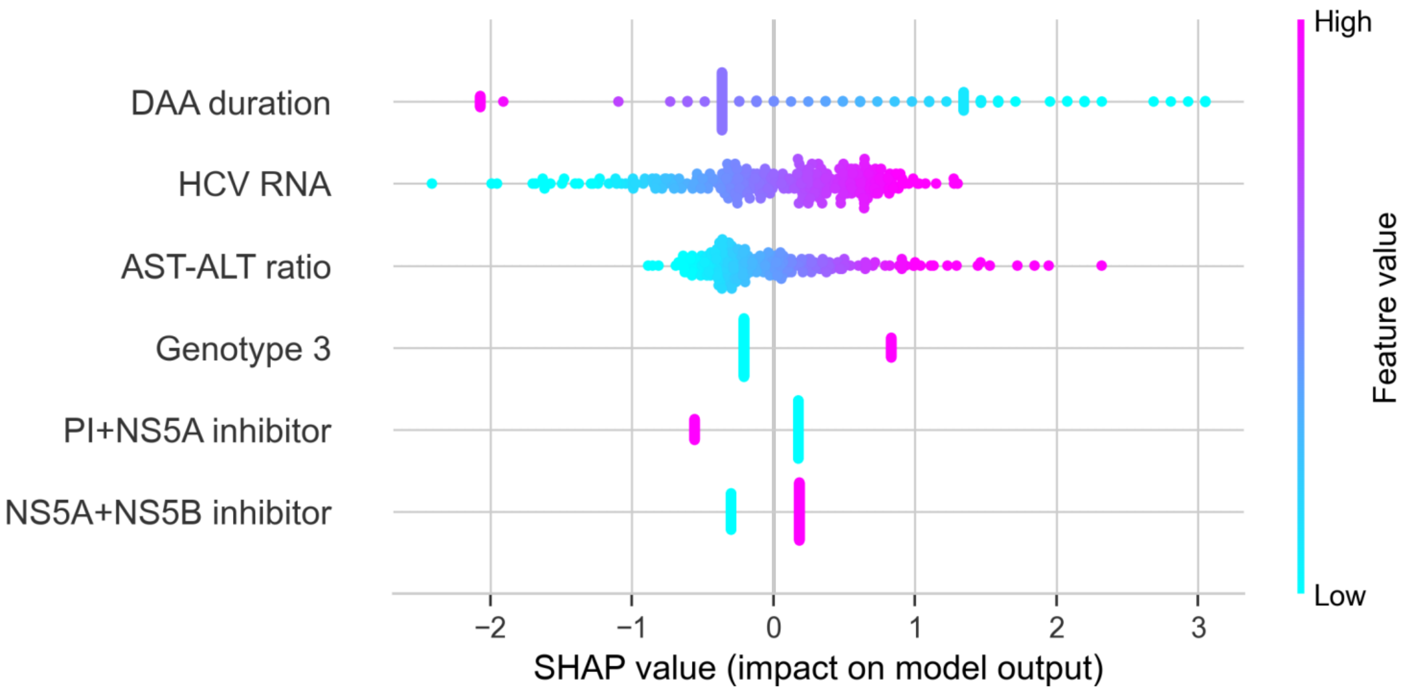


Figure. SHAP (SHapley Additive exPlanations) plot of feature contributions to the Elastic Net model's predictions of short-duration DAA response.

Predictors of treatment failure (+)
- High HCV RNA,
- High AST–ALT ratio
- Genotype 3
- NS5A+NS5B inhibitors

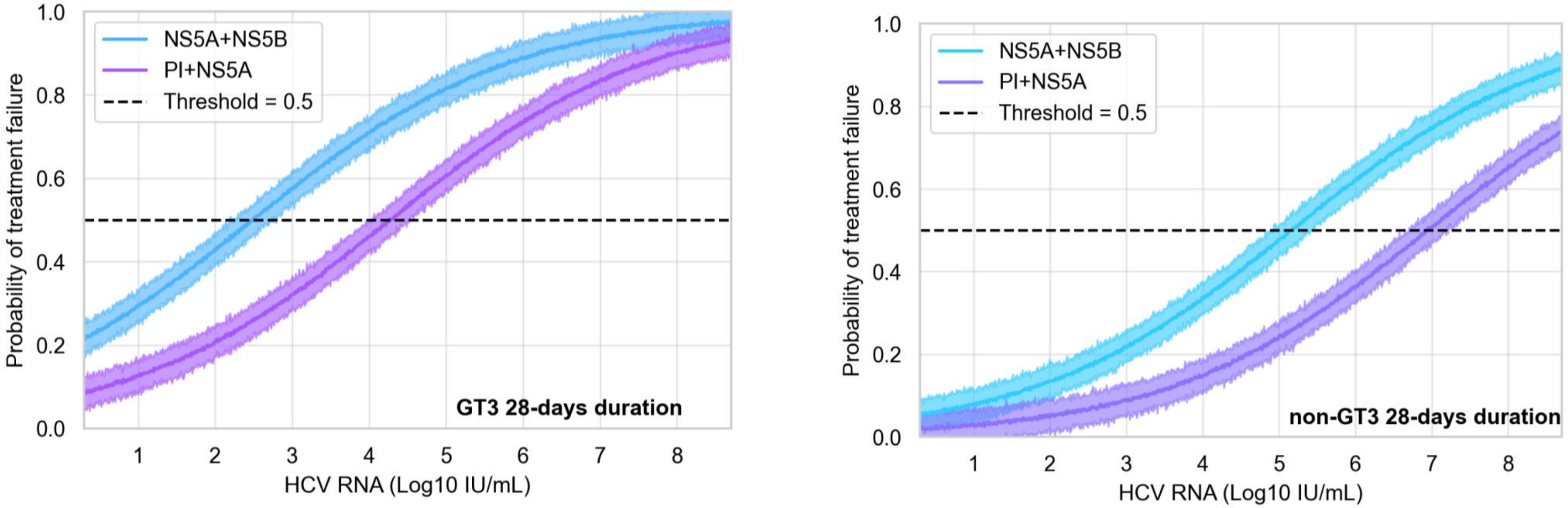
Predictors of SVR (+)
- Longer DAA durations
- PI+NS5A inhibitors

Including additional factors (25+ considered) did not improve performance.

HCV RNA cutoffs predicting failure

HCV RNA cutoffs predictive of short-duration treatment failure were higher for PI+NS5A vs. NS5A+NS5B.

4-weeks duration

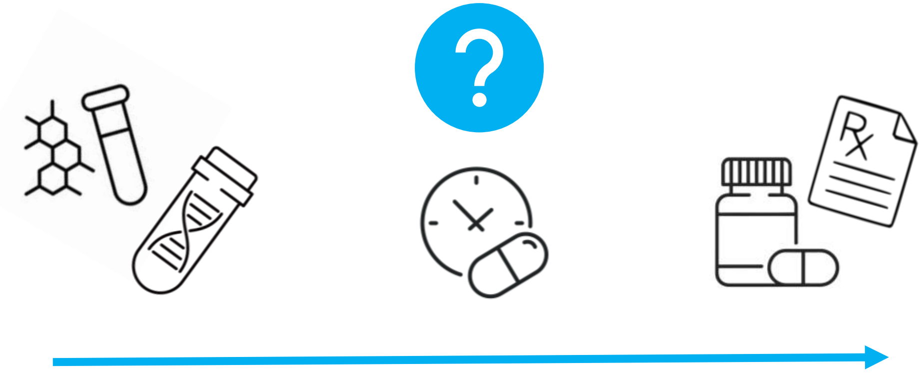


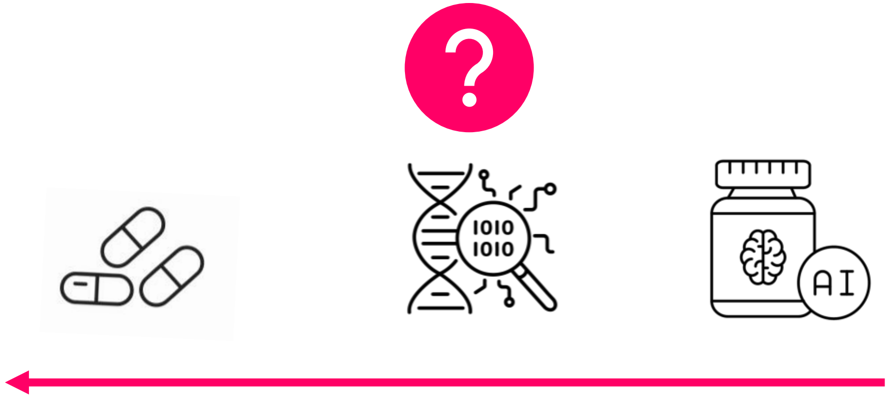
Figures. HCV RNA cutoffs predictive of short duration treatment failure for genotypes, DAA combinations class and durations were estimated with AST-ALT fixed at the mean. Confidence intervals were bootstrapped (n=1000).
Abbreviations: GT, genotype. *non-GT3 = 85% GT1; 15% GT2,4,6

Clinical utility

Forward prediction of response to short duration DAAs to personalise therapy

Effective delivery of short DAA durations could improve treatment uptake, cost-effectiveness, and health system efficiency





Backward prediction of response to incomplete DAAs to guide retreatment

HCV RNA+ following incomplete standard duration therapy - unclear if reinfection or treatment failure. Increasingly common.

Threshold optimisation

Standard threshold balances sensitivity **(81%)** and specificity **(76%).**

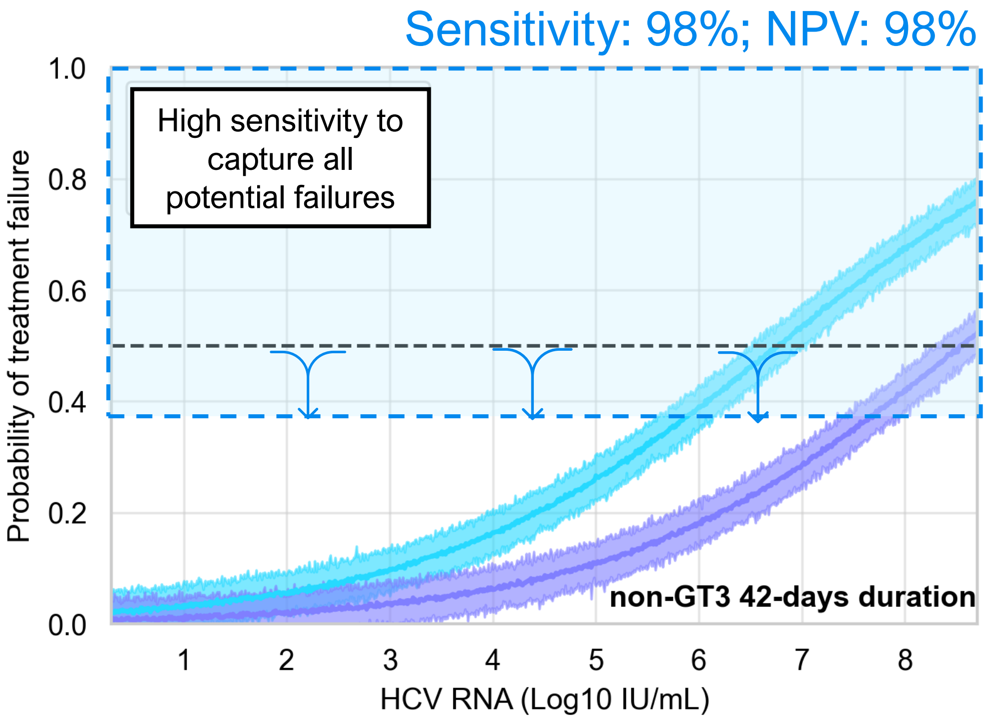
Optimizes specificity **(87%)** to identify true short-duration failure (vs. reinfection) if viraemic post- incomplete treatment.

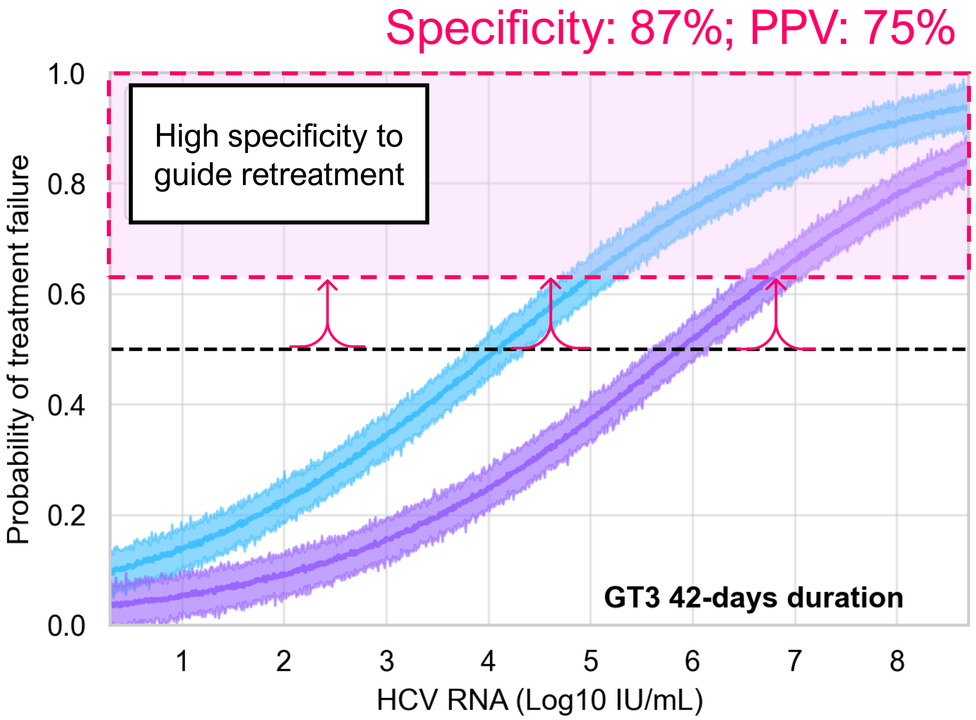
Metric, mean [SD]	Standard	Quadratic	Brier
Sensitivity	0.81 [0.09]	0.98 [0.03]	0.65 [0.11]
Specificity	0.76 [0.09]	0.51 [0.10]	0.87 [0.07]
PPV	0.67 [0.07]	0.53 [0.03]	0.75 [0.09]
NPV	0.88 [0.05]	0.98 [0.02]	0.83 [0.06]

Optimizes sensitivity **(98%)** to capture all potential short duration failures and avoid the costs of retreatment (where someone should have received standard duration)

Sensitive vs. Specific thresholds

What happens to HCV RNA cutoffs predictive of short-duration failure when we adjust our thresholds?





Abbreviations: GT, genotype; PPV, positive predictive value; NPV, negative predictive value.

vh2025

Key action 1: A simple machine learning model using readily available baseline clinical factors could help personalise DAA durations or guide retreatment decisions if a person is vireamic following incomplete treatment.

Key action 2: **Warning!!** This model does not meet the high standards required for clinic translation. We need larger, more diverse data samples to externally validate, calibrate, and increase model accuracy – while respecting data privacy, sovereignty and being transparent in how we use and keep data safe,

Key action 3: Rethink, refocus and reverse engineer solutions around people not the clinical protocols. This is where us numbers nerds need the knowledge and expertise of the community to set us in the right direction.

Acknowledgements

We thank all **study participants,** investigators, coordinators & data managers involved in the included studies:

1. REACT
2. TARGETED I
3. TARGETED II
4. TARGETED III
5. STRIVE4
6. REACH-C
7. STOP-C
8. CEASE
9. SMART-C

https://github.com/jojocarson/hcv_tx_response