Machine Learning for predictive modelling based on small biomedical and clinical data

Presenter: Torgyn Shaikhina

Supervisors: Dr Natasha Khovanova and Dr Mark Leeson

WARWICK CENTRE FOR PREDICTIVE MODELLING SEMINAR SERIES

10TH MAY 2016, COVENTRY, UK

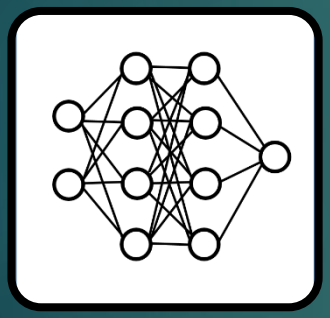
Content

- Machine Learning (ML) for Small Data
  - What is ML
  - Challenges with Biomedical and Clinical data
  - Our framework
- Case 1: Neural Networks in Hard Tissue Engineering
  - Background and Model
  - Results
  - Validation
- Case 2: Tree-based model in Kidney Transplantation
  - Background
  - Decision Tree model
  - Random Forest model
- Summary

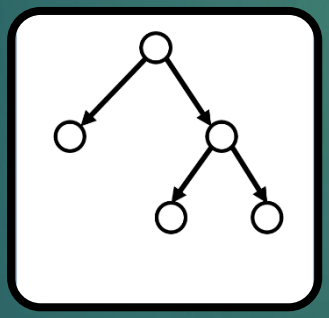
Machine Learning (ML)

- "Field of study that gives computers the ability to learn without being explicitly programmed“ (A. Samuel, 1959)

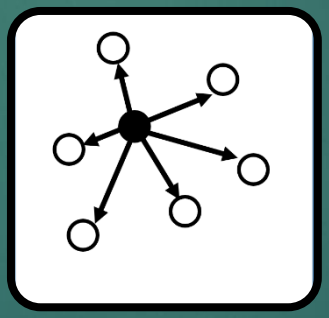
- **Original purpose (1950s)**: creating AI, simulating intelligence
- **Current use**: finding trends in large complex data, recognising patterns (speech, image and semantics)



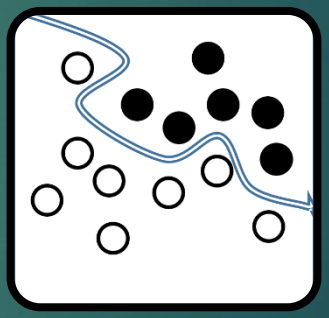
Artificial Neural Networks
- Deep learning



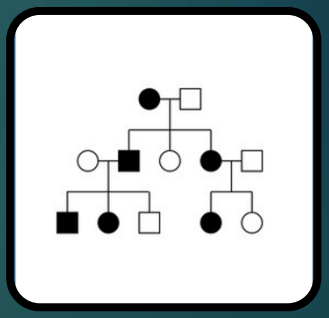
Tree-based methods
- Decision trees
- Random forests



Instance-based learning
- k-Nearest Neighbour



Kernel methods
- Support Vector machines



Genetic algorithms
- Evolutionary theory

ML and Biomedical Engineering

- **Machine Learning (ML) - indispensable tool in Bioinformatics [1,2]**
- **Still a relatively slow take up in biomedical engineering and healthcare**

Why Biomedical/Clinical data are small?

Cost

- Expensive experiments
- Living tissue
- Clinical data

Standards

- Institutional bias
- Ethical approvals
- International data transfer

"Big Data"?

- Too few events
- Exclusion criteria
- Missing data
- Class imbalance

Why is ML rarely viewed in the context of "**Small Data**"?
(<10 observations per predictor variable)

Insufficient training data

Insufficient test data

High volatility

Generalisation issues

ML and Small Data: making it work

Method of multiple runs

- What does it do?

Small data → many ML models

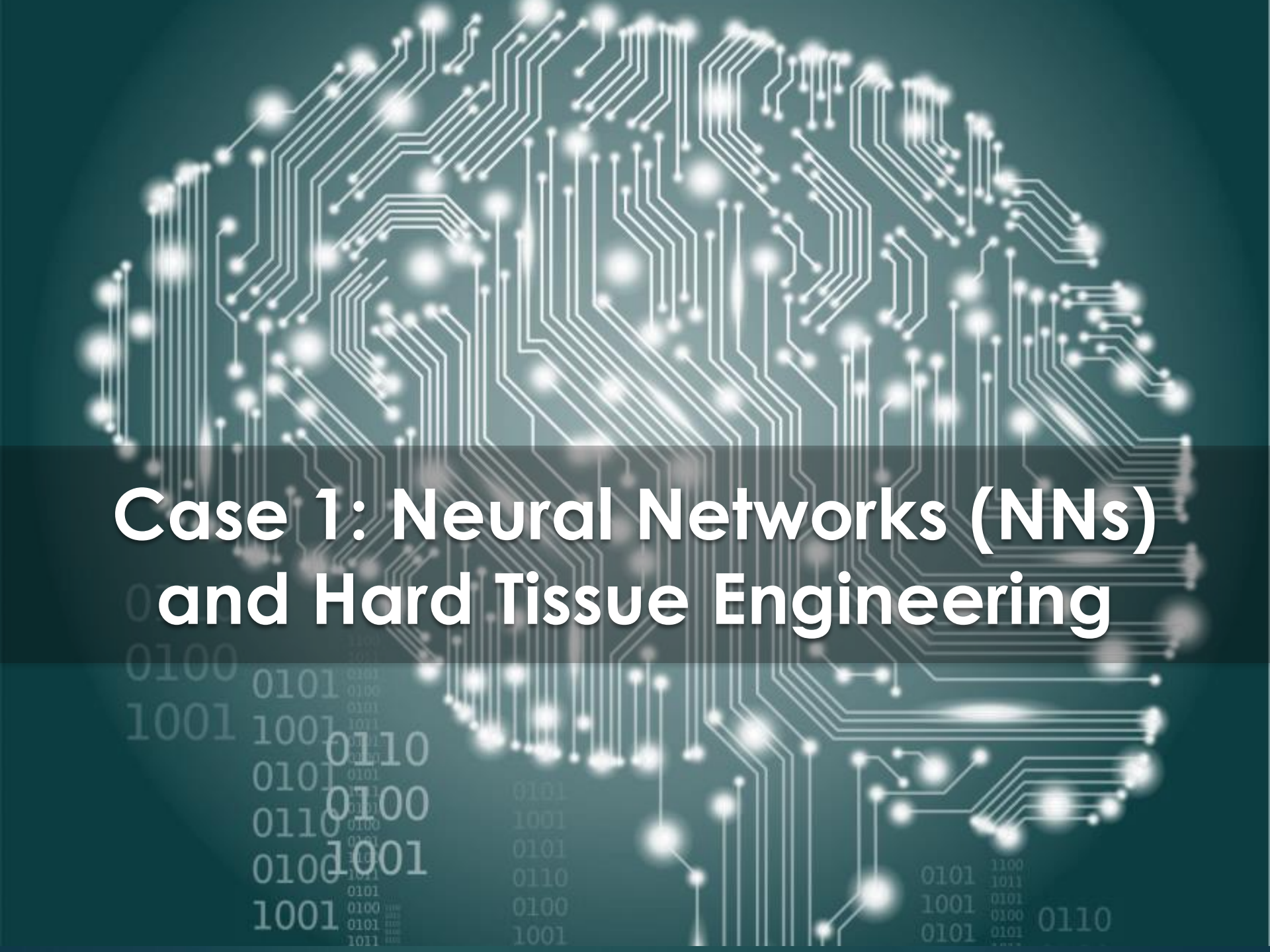
Reduces volatility

Increases consistency

- How does it work?
  - "Run" = thousands of ML models trained in parallel
  - A range of well- and poorly-performing models
  - Allows for iterative design optimisation
  - Performance measured collectively across the run
  - Use output of the best-performing model

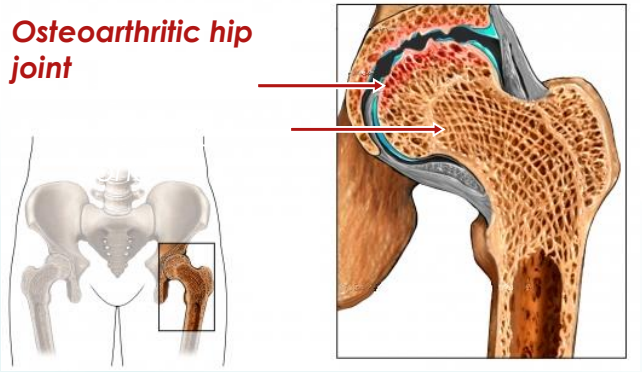
Surrogate data

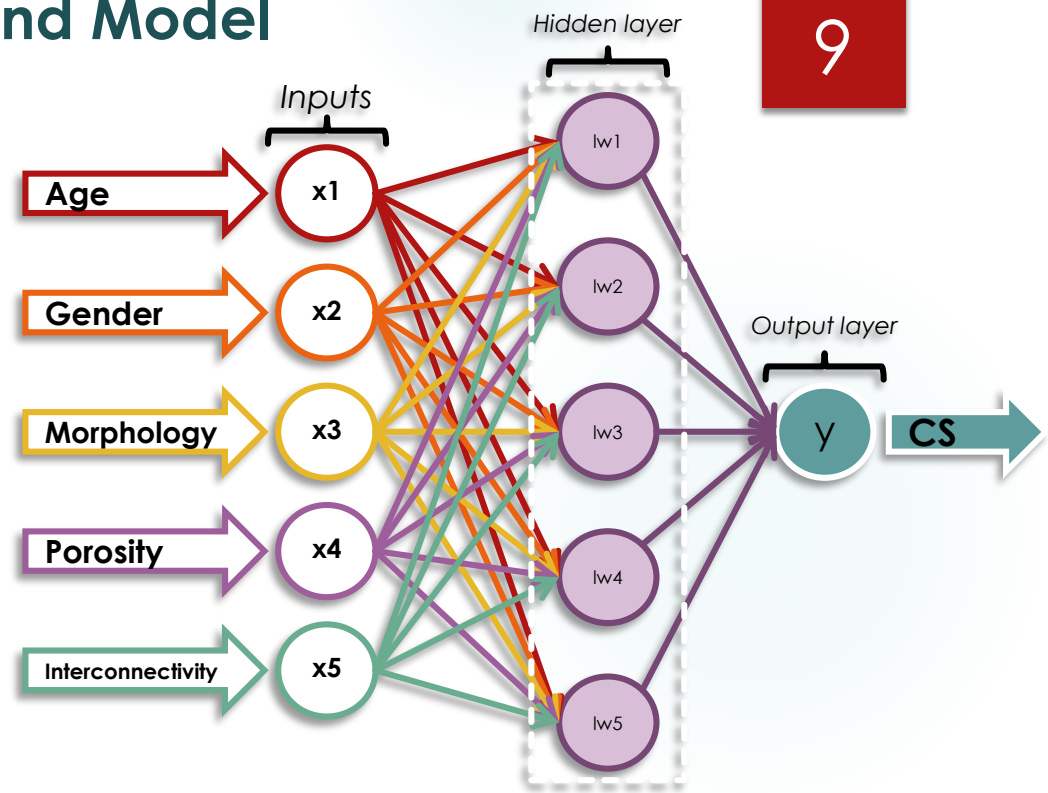
- What does it do?
  - Quantifies random effects due to small data
  - Model validation despite insufficient number of **test** samples
- How does it work?
  - Generate synthetic samples that mimic the real dataset
  - Train and test ML model on surrogate data
  - Highest performing surrogate model = lowest performance threshold for real data models


Case 1: Neural Networks (NNs)
and Hard Tissue Engineering

Case study 1: Background and Model

Regression task: to predict compressive strength (CS) of trabecular bone in severe osteoarthritis



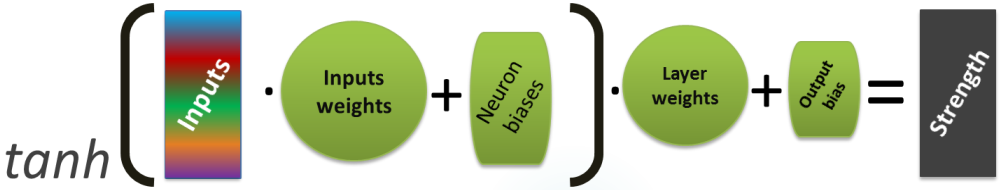


The model

- 35 trabecular bone samples [4]
- 20% reserved for tests
- Feedforward backpropagation NN with 5 input features and 1 output [5]
- Multiple run of 2000 NNs

$$y = \tanh\left[\bar{x} \cdot IW + \overline{b_{(1)}}\right] \cdot \overline{lw'} + b_{(2)}$$

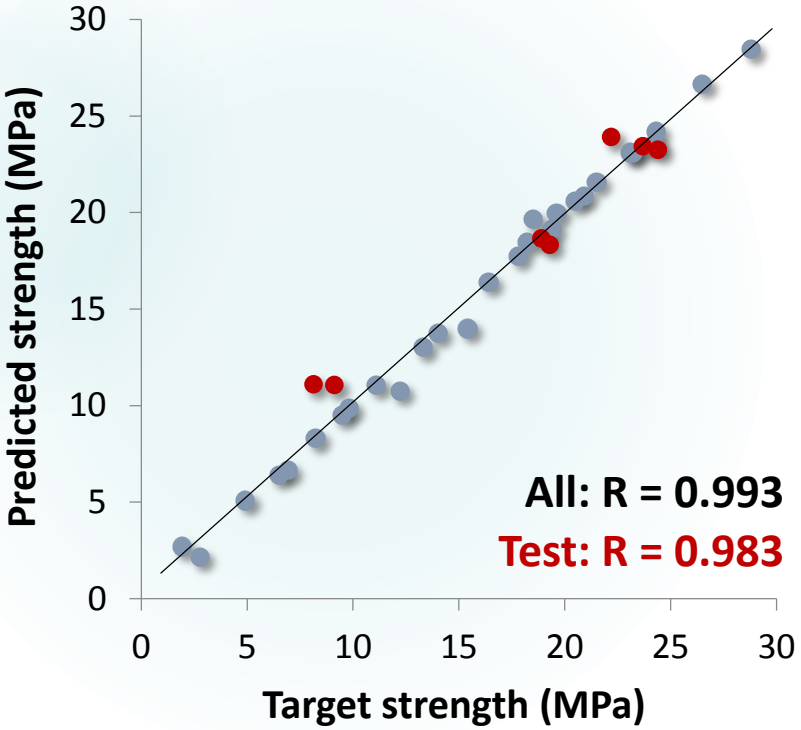
output = tanh(*inputs* × *input weights* + *biases*) × *layer weights* + *output bias*

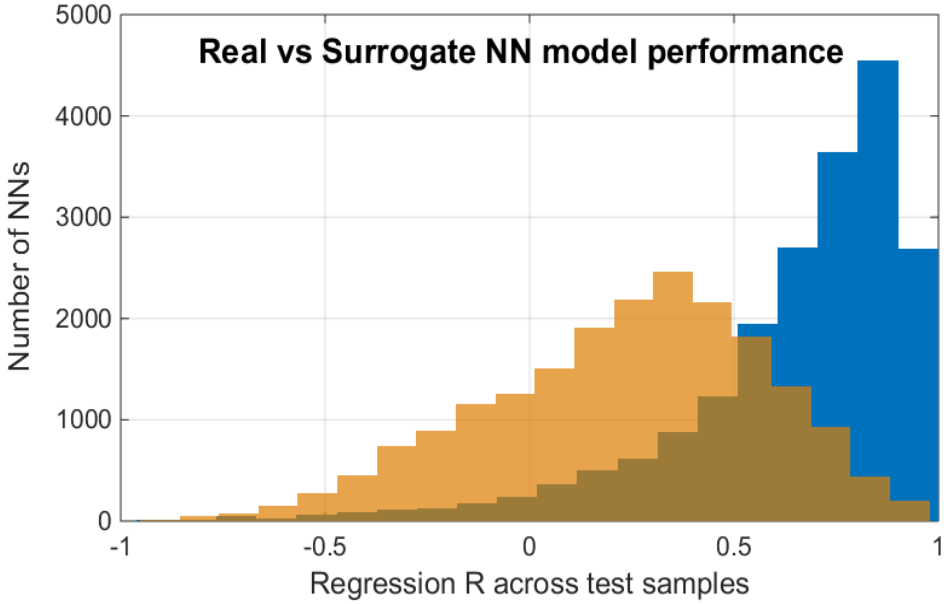


Case study 1: Results

NN performance

- Regression between actual and predicted CS:
  - across all samples, R = 99.3%
  - across 7 test samples, R = 98.3%
- Standard error = 0.85 MPa





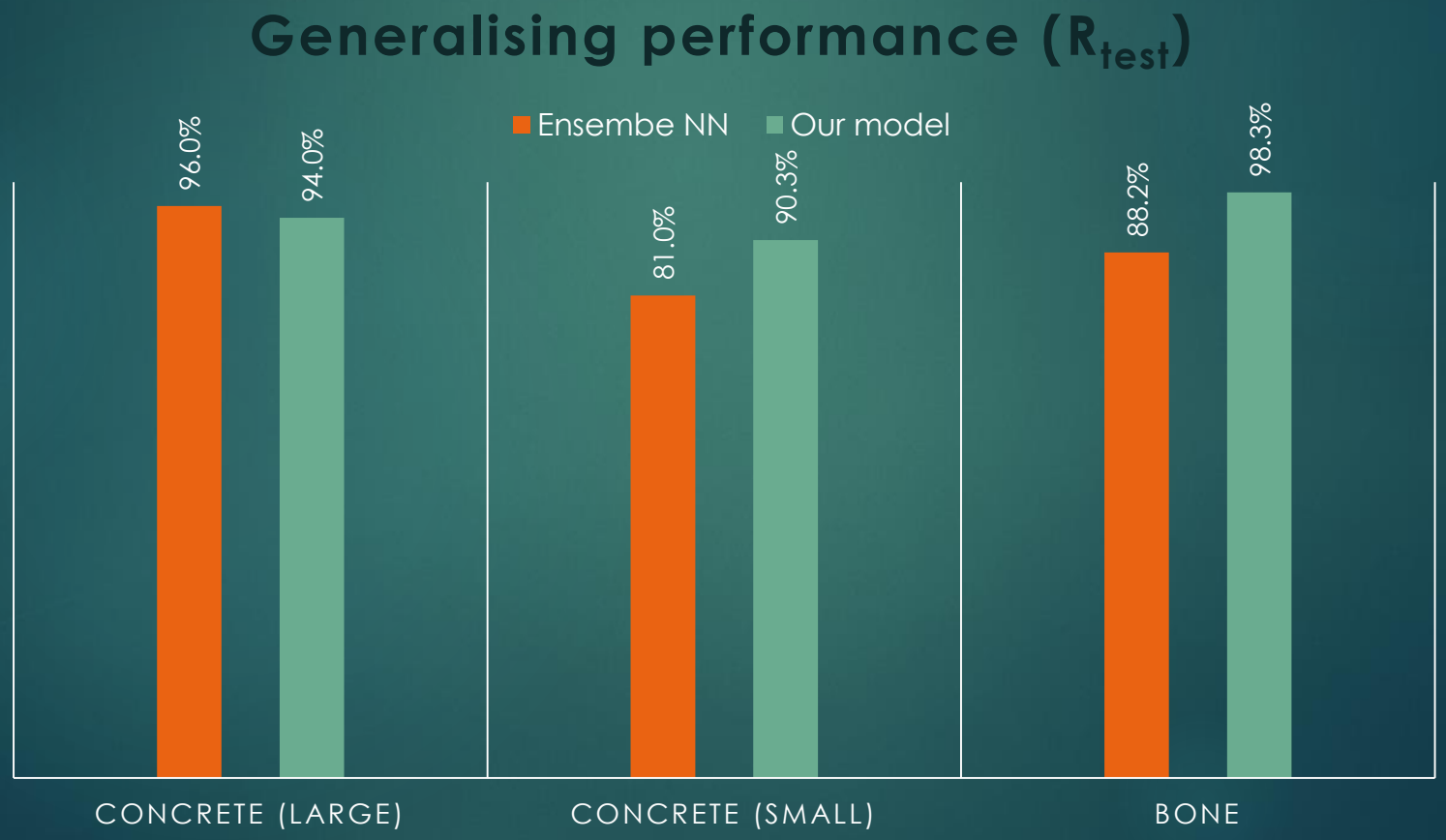
Validation with surrogates

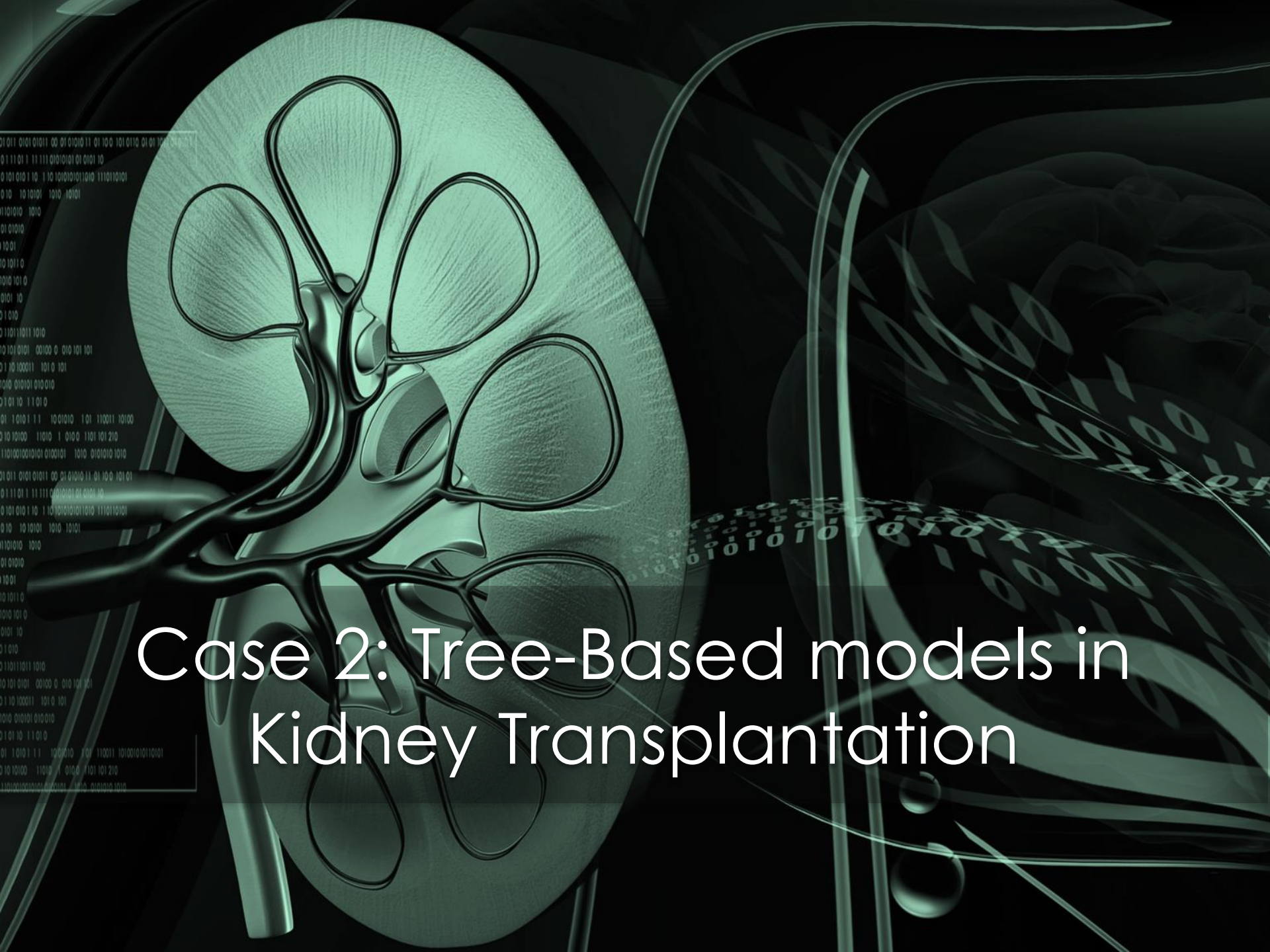
- Wilcoxon rank test (200000 NNs)
  - Hypothesis rejected ($p<0.000001$)
- Mean values
  - Surrogates, R = 0.33
  - Real data, R = 0.68
- Surrogate threshold R = 0.87

Comparison with NN ensembles

NN ensembles – powerful extension to NNs:

- Performed with 96% generalising accuracy on large-data concrete model (2% improvement)
- Underperformed with small-data





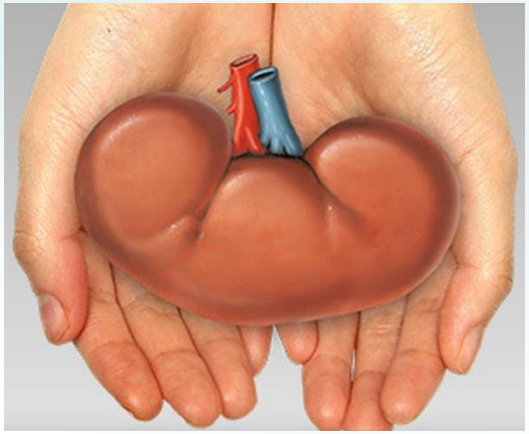
Case 2: Tree-Based models in Kidney Transplantation

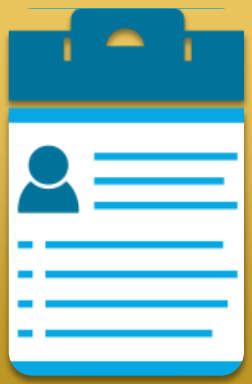
Case study 2: Background of the task

Classification task: to predict rejection (R/NR) of kidney transplants in early (<30 days)post-transplant period

Secondary tasks:

- Identify risk factors
- Dangerous antibody subclasses (IgG1-4)
- Harmful antibody levels





Dataset [6,7] - 80 patients

- **15 predictors**
  - age, gender, tissue mismatches, antibody levels, dialysis, etc.
- **Well balanced** : 46 R and 34 NR
- **Issues:** missing data, heterogeneity
- **Computational intensity**: 60960 splits

Model

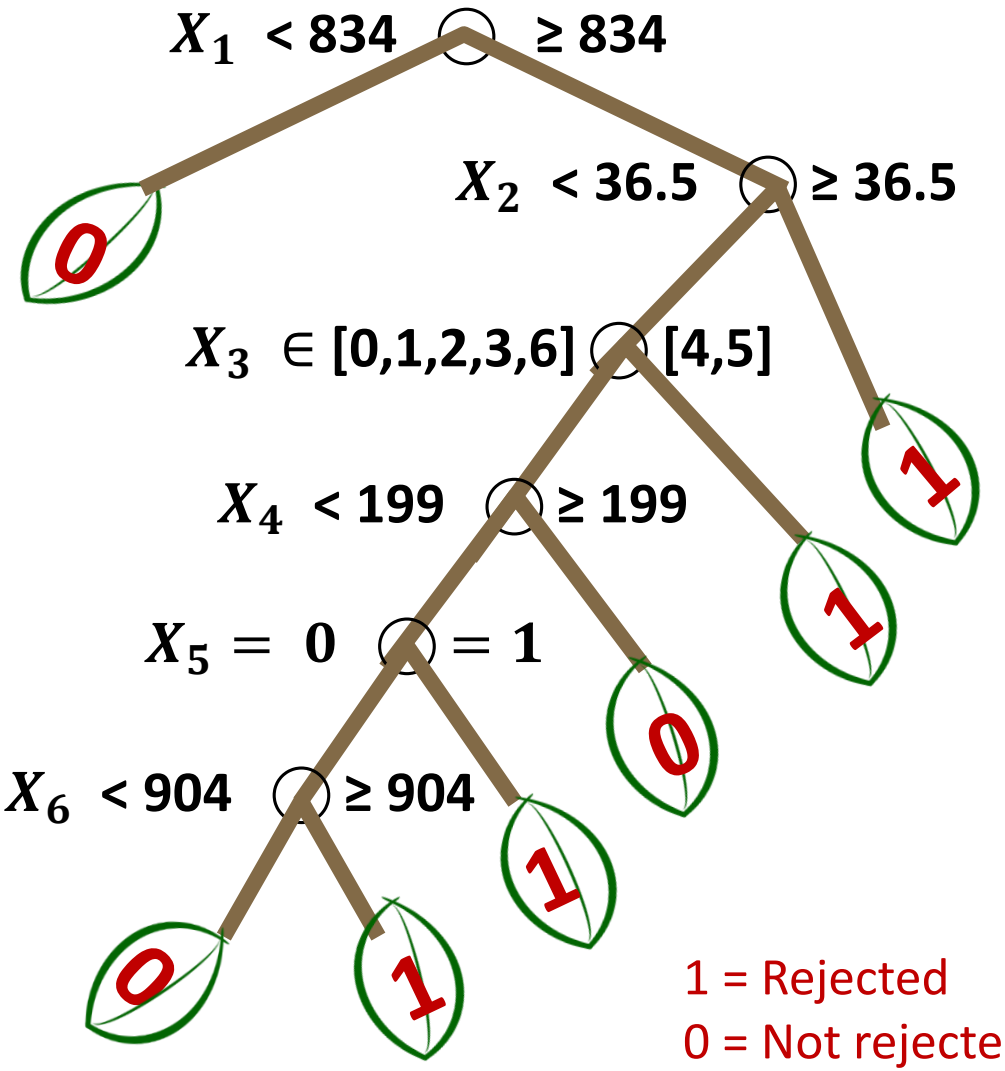
- Standard CART implemented in MATLAB
- Tree pruning: complexity penalised
- Multiple run of 600 DTs
- Random Forest (RF) extension to improve classifier robustness

Case study 2: DT model

- Only 6 (out of 15 possible) predictors were used by the DT:

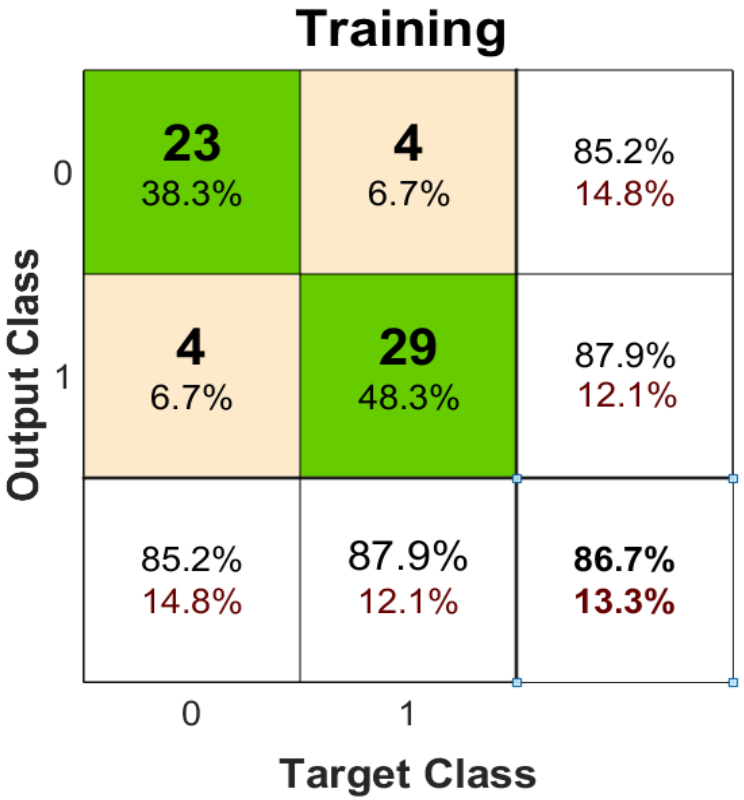
Risk levels:

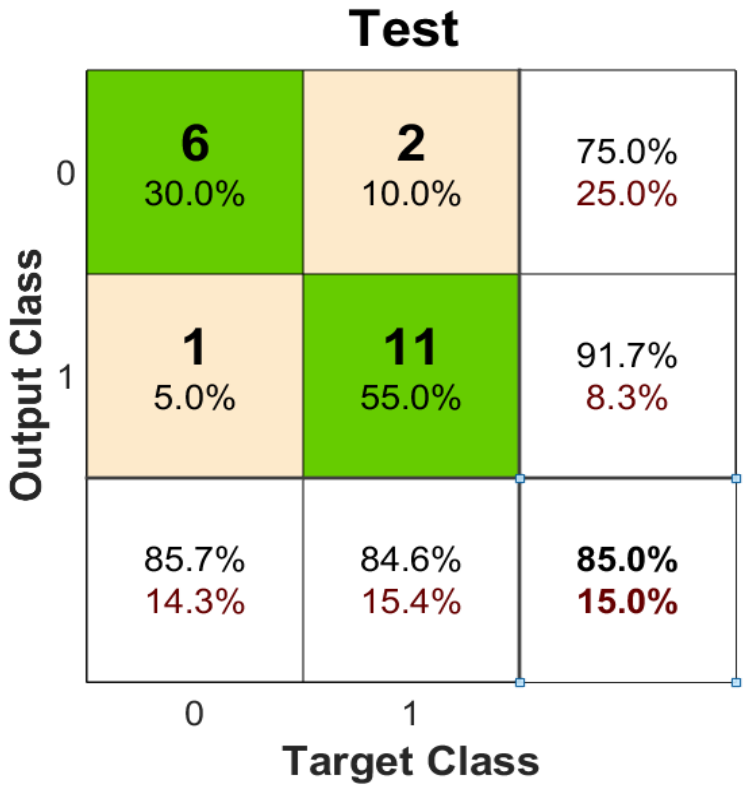
- X_1 highest antibody level
- X_2 total IgG4 ≥ 36.5
- X_3 tissue mismatches
- X_4 total IgG2
- X_5 delayed graft function
- X_6 total IgG1 ≥ 904



Case study 2: DT model results

Confusion matrices





Training accuracy = 86.7%
Sensitivity = 87.9%
Specificity = 85.2%

Generalising accuracy = 85.0%
Sensitivity = 84.6%
Specificity = 85.7%

Summary

- Machine learning is able to learn from small biomedical data:
  - Our framework produces well-generalising predictive models built with limited data, which outperform some state-of-the-art alternative models
  - How much data is enough remains a compromise

- **NN model for strength estimation in trabecular bone**
  - Non-destructive estimation of bone fracture risk
  - Highly-accurate (98.3%)
  - Compressive Strength accurate to 0.85 MPa

- **DT as a predictive tool in Kidney transplantation**:
  - Classified AMR/Non-AMR with 85% accuracy
  - Identified key risk factors
  - Estimated specific levels of antibodies

References

1. Inza, I. et al., 2010. Machine Learning: An Indispensable Tool in Bioinformatics. In R. Matthiesen, ed. *Bioinformatics Methods in Clinical Research*. Humana Press, pp. 25–48.
2. Campbell, C., 2014. Machine Learning Methodology in Bioinformatics. In N. Kasabov, ed. *Springer Handbook of Bio-/Neuroinformatics*. Springer Berlin Heidelberg, pp. 185–206.
3. I.-C. Yeh, "UCI Machine Learning Repository: Concrete Compressive Strength Data Set," *Machine Learning Repository*, University of California Irvine, Center of Machine Learning and Intelligent Systems, 2007. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Concrete+Compressive+Strength
4. Perilli, E. et al., 2007. Structural parameters and mechanical strength of cancellous bone in the femoral head in osteoarthritis do not depend on age. *Bone*, 41, pp.760–768.
5. Khovanova, N., et al., 2014. Neural networks for analysis of trabecular bone in osteoarthritis. *Bioinspired, Biomimetic and Nanobiomaterials*, 4(1), pp.90–100.
6. Lowe, D. et al., 2013. Significant IgG subclass heterogeneity in HLA-specific antibodies: Implications for pathogenicity, prognosis, and the rejection response. *Human Immunology*, 74, pp.666–672.
7. Khovanova, N., et al., 2015. Subclass analysis of donor HLA specific IgG in antibody incompatible renal transplantation reveals a significant association of IgG4 with rejection and graft failure. *Transplant International*, 28(12), pp. 1405–1415.

Acknowledgments

This work has been supported by EPSRC UK (EP/K02504X/1)

Our collaborators:

Drs. David Briggs and Dave Lowe
NHS Blood and Transplant Birmingham

Drs. Robert Higgins, Sunil Daga and Nithya Krishnan
University Hospitals Coventry and Warwickshire NHS Trust

Dr. Daniel Mitchell
Warwick Medical School, University of Warwick

WARWICK
THE UNIVERSITY OF WARWICK

EPSRC
Engineering and Physical Sciences Research Council

West Midlands (South)
NIHR Comprehensive Local Research Network

NHS
National Institute for Health Research

Thank you!