Using Machine Learning for Cosmetics and Cosmetic Ingredients

Timothy E. H. Allen

MRC Toxicology Unit, University of Cambridge
Department of Chemistry, University of Cambridge

teha2@cam.ac.uk
Conflict of Interest Statement

- I have no real or perceived conflicts of interest with the research described in this presentation.
Outline/Objectives

- To introduce how Unilever is applying the principles of next generation risk assessment (NGRA)
- To discuss where predictive computational toxicology and machine learning fit in
- To show outline some of the computational approaches that have been developed and compare them
- To show why machine learning efforts are good for some tasks
- To identify where these research efforts are now going, and how that impacts their use
The World of Consumer Products

- Personal Care
- Foods & Refreshment
- Homecare
Use of Safety Information for Consumer Products

- Context...
  - Classification and labelling
    - Favours a cautious approach
    - Hazard based rules
    - Occupational focus
  - Screening/product development
    - Many potential lead chemicals
    - Often only hazard prediction methods are used
    - Performance of models is less critical
    - Exposure may be considered by the use of threshold-based approaches e.g., TTC, DST, EBW, EcoTTC
  - Risk Assessment including actual exposure
    - Requires a high degree of accuracy
    - Route and amount of exposure dictate the need for toxicology data
    - High level of scrutiny (internal and external)
Ab Initio NGRA Framework

Exposure Estimation

- Local and systemic exposure estimates
  - Use scenario
  - Consumer Habits and Practices
  - Applied Dose
  - ADME parameters
  - Internal Exposure (PBK)

Collate Existing Information

- Problem Formulation
  - Molecular Structure
  - In silico predictions
  - Literature

Tier 0

Tier 1

- In vitro Hazard Characterisation
  - Initial PoD identification
    - ToxTracker
    - SafetyScreen44®
    - BioMap® Diversity 8 Panel
    - Cell Stress Panel
    - HTTr - Temp0-Seq

- Concentration-Response analysis

- Determine Margin of Safety
  - PoD_{in vitro}
  - Sufficient data and high certainty
  - Insufficient data and/or low certainty

Tier 2

- Risk Assessment Conclusion
  - High risk or Low risk conclusion based on the margin of safety calculations.

Uncertainty

Mechanistic understanding


SOT FDA Colloquia on Emerging Toxicological Science Challenges in Food and Ingredient Safety
ICCR Principles of Risk Assessment without Animal Tests

- The overall goal is human safety risk assessment
- The assessment is exposure led
- The assessment is hypothesis driven
- The assessment is designed to prevent harm (i.e., distinguish between adaptation and adversity)
- Using a tiered and iterative approach
- Following an appropriate appraisal of existing information
- Using robust and relevant methods and strategies
- The logic of the approach should be transparently and explicitly documented
- Sources of uncertainty should be characterised and documented
Adverse Outcome Pathway

Chemical Properties
- Toxicant
  - Macromolecular Interactions
    - Receptor/Ligand Interaction
    - DNA Binding
    - Protein Oxidation
  - Cellular Responses
    - Gene Activation
    - Protein Production
    - Altered Signalling
    - Protein Depletion
  - Organ Responses
    - Altered Physiology
    - Disrupted Homeostasis
    - Altered Tissue Development or Function
  - Organism Responses
    - Lethality
    - Impaired Development
    - Impaired Reproduction
    - Cancer
  - Population Responses
    - Structure
    - Recruitment
    - Extinction

Anchor 1 (initiating event)

Anchor 2 (adverse outcomes at the organism- or population-level)

Structural Alerts

ChEMBL

Bowes 2012 Targets

Computational Tools

2D Fragments

SMILES

Performance Data

Novel Compound

0

No Hit

1

Hit

Structural Alerts

2D STRUCTURAL ALERTS

AA2AR ACTIVE

Novel Molecule

P(Ki) = 6.66 +/- 0.78
Range = 4.84 - 8.18
n = 41

Expanded Target List

- Acetylcholinesterase
- Serotonin 3a (5-HT3a) receptor
- Serotonin 3a (5-HT3a) receptor

- Adenosine A2a receptor
- Tyrosine-protein kinase LCK

- Alpha-2a adrenergic receptor
- Vasopressin V1a receptor
- Type-1 angiotensin II receptor

- Androgen receptor
- RAC-alpha serine/threonine-protein kinase
- Beta-secretase 1

- Beta-1 adrenergic receptor
- Cholinesterase
- Caspase-1

- Beta-2 adrenergic receptor
- Caspase-3
- Caspase-8

- Delta opioid receptor
- Muscarinic acetylcholine receptor M5

- Dopamine D1 receptor
- Inhibitor of nuclear factor κ-B kinase subunit α

- Dopamine D2 receptor
- Macrophage colony-stimulating fac. 1 receptor

- Dopamine transporter
- Casein kinase I isoform delta

- Endothelin receptor ET-A
- Endothelin B receptor

- Glucocorticoid receptor
- Neutrophil elastase

- Histamine H1 receptor
- Neutrophil elastase

- Mu opioid receptor
- Neutrophil elastase

- Muscarinic acetylcholine receptor M1
- Neutrophil elastase

- Muscarinic acetylcholine receptor M2
- Ephrin type-A receptor 2

- Muscarinic acetylcholine receptor M3
- Fibroblast growth factor receptor 1

- Norepinephrine transporter
- Peptidyl-prolyl cis-trans isomerase

- Serotonin 2a (5-HT2a) receptor
- Vascular endothelial growth factor receptor 1
- Vascular endothelial growth factor receptor 3

- Serotonin transporter
- Tyrosine-protein kinase FYN

- Vasopressin V1a receptor
- Histone deacetylase 3

- Type-1 angiotensin II receptor
- Insulin-like growth factor 1 receptor

- RAC-alpha serine/threonine-protein kinase
- Insulin receptor

- Beta-secretase 1
- Vascular endothelial growth factor receptor 2

- Cholinesterase
- Leukotriene B4 receptor 1

- Caspase-1
- Tyrosine-protein kinase Lyn

- Caspase-3
- Mitogen-activated protein kinase 1

- Caspase-8
- Mitogen-activated protein kinase 9

- Muscarinic acetylcholine receptor M5
- MAP kinase-activated protein kinase 2

- Inhibitor of nuclear factor κ-B kinase subunit α
- Hepatocyte growth factor receptor

- Macrophage colony-stimulating fac. 1 receptor
- Matrix metalloproteinase-13

- Casein kinase I isoform delta
- Matrix metalloproteinase-2

- Endothelin B receptor
- Matrix metalloproteinase-3

- Neutrophil elastase
- Matrix metalloproteinase-9

- Ephrin type-A receptor 2
- Serine/threonine-protein kinase NEK2

- Fibroblast growth factor receptor 1
- P2Y purinoreceptor 1

- Peptidyl-prolyl cis-trans isomerase
- Serine/threonine-protein kinase PAK 4

- Vascular endothelial growth factor receptor 1
- Phosphodiesterase 4A

- Tyrosine-protein kinase FYN
- Phosphodiesterase 5A

- Histone deacetylase 3
- PIP_3 3-kinase catalytic subunit α

- Insulin-like growth factor 1 receptor
- Peroxisome proliferator-activated receptor γ

- Insulin receptor
- Protein Tyr phosphatase non-receptor type 1

- Vascular endothelial growth factor receptor 2
- Protein Tyr phosphatase non-receptor type 11

- Leukotriene B4 receptor 1
- Protein Tyr phosphatase non-receptor type 2

- Tyrosine-protein kinase Lyn
- RAF proto-oncogene Ser/Thr-protein kinase

- Mitogen-activated protein kinase 1
- Retinoic acid receptor alpha

- Mitogen-activated protein kinase 9
- Retinoic acid receptor beta

- MAP kinase-activated protein kinase 2
- Rho-ass. coiled-coil-containing protein kinase I

- Hepatocyte growth factor receptor
- Ribosomal protein S6 kinase alpha-5

- Matrix metalloproteinase-13
- NAD-dependent protein deacetylase sirtuin-2

- Matrix metalloproteinase-2
- NAD-dependent protein deacetylase sirtuin-3

- Matrix metalloproteinase-3
- Proto-oncogene tyrosine-protein kinase Src

- Matrix metalloproteinase-9
- Substance-K receptor

- Serine/threonine-protein kinase NEK2
- Thromboxane A2 receptor

- P2Y purinoreceptor 1
- Tyrosine-protein kinase receptor TEK
Neural Networks

[Diagram of neural network with nodes labeled (a), (b), (c), and (d) connected through hidden layers.]

Chemistry

Biology

SOT FDA Colloquia on Emerging Toxicological Science Challenges in Food and Ingredient Safety
## Average Model Performance

<table>
<thead>
<tr>
<th></th>
<th>Training Data</th>
<th>Validation Data</th>
<th>Test Data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SE</td>
<td>SP</td>
<td>ACC</td>
</tr>
<tr>
<td><strong>AVERAGE</strong></td>
<td>92.1</td>
<td>96.5</td>
<td>95.8</td>
</tr>
<tr>
<td><strong>SD</strong></td>
<td>8.8</td>
<td>4.2</td>
<td>3.1</td>
</tr>
</tbody>
</table>
Model Performance vs. Total Compounds

Test MCC vs Total Compounds

MCC

Number of Compounds

KCNH2

MAPK1

SOT FDA Colloquia on Emerging Toxicological Science Challenges in Food and Ingredient Safety
Positive Probability Curve

Adenosine A2a Receptor Positive Probability Curve

- 99.41%
- 85.92%
- 82.61%
- 81.25%
- 66.67%
- 64.71%
- 63.16%
- 31.58%
- 25.93%
- 2.45%
## Comparative Model Performance

<table>
<thead>
<tr>
<th>Target Gene</th>
<th>SE</th>
<th>ΔSA</th>
<th>ΔRF</th>
<th>SP</th>
<th>ΔSA</th>
<th>ΔRF</th>
<th>ACC</th>
<th>ΔSA</th>
<th>ΔRF</th>
<th>MCC</th>
<th>ΔSA</th>
<th>ΔRF</th>
</tr>
</thead>
<tbody>
<tr>
<td>AChE</td>
<td>88.4</td>
<td>6.6</td>
<td>-3.2</td>
<td>85.3</td>
<td>-4.6</td>
<td>7.5</td>
<td>87.0</td>
<td>1.6</td>
<td>1.6</td>
<td>0.737</td>
<td>0.024</td>
<td>0.030</td>
</tr>
<tr>
<td>ADORA2A</td>
<td>97.6</td>
<td>2.8</td>
<td>-0.5</td>
<td>93.2</td>
<td>2.0</td>
<td>4.2</td>
<td>96.1</td>
<td>2.5</td>
<td>1.1</td>
<td>0.912</td>
<td>0.054</td>
<td>0.024</td>
</tr>
<tr>
<td>ADRA2A</td>
<td>91.3</td>
<td>11.2</td>
<td>2.0</td>
<td>93.9</td>
<td>-1.2</td>
<td>-1.2</td>
<td>92.7</td>
<td>4.3</td>
<td>0.2</td>
<td>0.853</td>
<td>0.084</td>
<td>0.004</td>
</tr>
<tr>
<td>AR</td>
<td>66.5</td>
<td>-0.8</td>
<td>1.1</td>
<td>99.1</td>
<td>1.4</td>
<td>1.2</td>
<td>90.5</td>
<td>0.8</td>
<td>1.2</td>
<td>0.749</td>
<td>0.026</td>
<td>0.036</td>
</tr>
<tr>
<td>ADRB1</td>
<td>92.7</td>
<td>7.0</td>
<td>0.3</td>
<td>89.5</td>
<td>-2.5</td>
<td>1.5</td>
<td>91.2</td>
<td>2.5</td>
<td>0.9</td>
<td>0.823</td>
<td>0.046</td>
<td>0.017</td>
</tr>
<tr>
<td>ADRB2</td>
<td>72.9</td>
<td>-2.3</td>
<td>-3.3</td>
<td>89.9</td>
<td>2.2</td>
<td>1.4</td>
<td>81.6</td>
<td>0.0</td>
<td>-0.9</td>
<td>0.639</td>
<td>0.004</td>
<td>-0.014</td>
</tr>
<tr>
<td>OPRD1</td>
<td>97.1</td>
<td>1.1</td>
<td>-1.2</td>
<td>81.0</td>
<td>-0.3</td>
<td>4.5</td>
<td>92.4</td>
<td>0.7</td>
<td>0.4</td>
<td>0.813</td>
<td>0.018</td>
<td>0.011</td>
</tr>
<tr>
<td>DRD1</td>
<td>77.4</td>
<td>0.9</td>
<td>-3.7</td>
<td>96.6</td>
<td>1.5</td>
<td>4.3</td>
<td>89.2</td>
<td>1.4</td>
<td>1.2</td>
<td>0.773</td>
<td>0.030</td>
<td>0.028</td>
</tr>
<tr>
<td>DRD2</td>
<td>98.3</td>
<td>1.9</td>
<td>-1.0</td>
<td>84.8</td>
<td>5.7</td>
<td>7.8</td>
<td>96.1</td>
<td>2.5</td>
<td>0.5</td>
<td>0.855</td>
<td>0.091</td>
<td>0.021</td>
</tr>
<tr>
<td>SLC6A3</td>
<td>89.9</td>
<td>1.3</td>
<td>-3.1</td>
<td>94.5</td>
<td>1.9</td>
<td>4.9</td>
<td>91.8</td>
<td>1.5</td>
<td>0.2</td>
<td>0.837</td>
<td>0.032</td>
<td>0.010</td>
</tr>
<tr>
<td>EDNRA</td>
<td>93.8</td>
<td>-0.6</td>
<td>-3.4</td>
<td>95.6</td>
<td>1.7</td>
<td>4.8</td>
<td>94.6</td>
<td>0.4</td>
<td>0.4</td>
<td>0.893</td>
<td>0.010</td>
<td>0.009</td>
</tr>
<tr>
<td>NR3C1</td>
<td>74.5</td>
<td>2.3</td>
<td>0.7</td>
<td>96.9</td>
<td>0.1</td>
<td>0.6</td>
<td>90.1</td>
<td>0.7</td>
<td>0.6</td>
<td>0.760</td>
<td>0.018</td>
<td>0.015</td>
</tr>
<tr>
<td>KCNH2</td>
<td>84.4</td>
<td>15.7</td>
<td>-8.6</td>
<td>70.9</td>
<td>-11.5</td>
<td>17.4</td>
<td>79.1</td>
<td>5.0</td>
<td>1.6</td>
<td>0.558</td>
<td>0.059</td>
<td>0.036</td>
</tr>
<tr>
<td>HRH1</td>
<td>95.2</td>
<td>8.0</td>
<td>-0.6</td>
<td>88.4</td>
<td>-5.3</td>
<td>0.7</td>
<td>92.0</td>
<td>1.7</td>
<td>0.0</td>
<td>0.840</td>
<td>0.032</td>
<td>-0.001</td>
</tr>
<tr>
<td>OPRM1</td>
<td>94.8</td>
<td>1.2</td>
<td>-1.1</td>
<td>94.5</td>
<td>1.7</td>
<td>3.2</td>
<td>94.7</td>
<td>1.4</td>
<td>0.6</td>
<td>0.889</td>
<td>0.030</td>
<td>0.014</td>
</tr>
<tr>
<td>CHRM1</td>
<td>96.6</td>
<td>6.4</td>
<td>0.9</td>
<td>83.3</td>
<td>-2.9</td>
<td>0.7</td>
<td>91.7</td>
<td>2.9</td>
<td>0.8</td>
<td>0.821</td>
<td>0.062</td>
<td>0.019</td>
</tr>
<tr>
<td>CHRM2</td>
<td>93.9</td>
<td>2.9</td>
<td>0.0</td>
<td>94.5</td>
<td>1.1</td>
<td>3.2</td>
<td>94.2</td>
<td>1.9</td>
<td>1.8</td>
<td>0.883</td>
<td>0.039</td>
<td>0.035</td>
</tr>
<tr>
<td>CHRM3</td>
<td>91.9</td>
<td>3.4</td>
<td>-3.2</td>
<td>93.8</td>
<td>1.8</td>
<td>5.8</td>
<td>92.8</td>
<td>2.8</td>
<td>0.8</td>
<td>0.854</td>
<td>0.053</td>
<td>0.017</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Target Gene</th>
<th>SE</th>
<th>ΔSA</th>
<th>ΔRF</th>
<th>SP</th>
<th>ΔSA</th>
<th>ΔRF</th>
<th>ACC</th>
<th>ΔSA</th>
<th>ΔRF</th>
<th>MCC</th>
<th>ΔSA</th>
<th>ΔRF</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLC6A2</td>
<td>94.9</td>
<td>4.2</td>
<td>-0.2</td>
<td>92.5</td>
<td>-1.0</td>
<td>1.0</td>
<td>93.9</td>
<td>2.0</td>
<td>0.3</td>
<td>0.875</td>
<td>0.039</td>
<td>0.008</td>
</tr>
<tr>
<td>HTR2A</td>
<td>99.1</td>
<td>2.4</td>
<td>-0.4</td>
<td>88.2</td>
<td>-0.4</td>
<td>6.1</td>
<td>96.7</td>
<td>1.8</td>
<td>1.0</td>
<td>0.901</td>
<td>0.051</td>
<td>0.030</td>
</tr>
<tr>
<td>HTR3A</td>
<td>89.4</td>
<td>5.7</td>
<td>-1.0</td>
<td>98.2</td>
<td>-0.4</td>
<td>0.4</td>
<td>95.8</td>
<td>1.3</td>
<td>0.0</td>
<td>0.893</td>
<td>0.034</td>
<td>0.000</td>
</tr>
<tr>
<td>SLC6A4</td>
<td>98.4</td>
<td>2.7</td>
<td>-0.3</td>
<td>89.0</td>
<td>2.1</td>
<td>6.3</td>
<td>96.3</td>
<td>2.5</td>
<td>1.2</td>
<td>0.892</td>
<td>0.070</td>
<td>0.037</td>
</tr>
<tr>
<td>LCK</td>
<td>95.5</td>
<td>3.7</td>
<td>0.2</td>
<td>79.8</td>
<td>-2.6</td>
<td>-0.9</td>
<td>92.3</td>
<td>2.5</td>
<td>0.0</td>
<td>0.763</td>
<td>0.054</td>
<td>-0.002</td>
</tr>
<tr>
<td>AVPR1A</td>
<td>93.9</td>
<td>1.9</td>
<td>0.6</td>
<td>99.3</td>
<td>1.5</td>
<td>4.8</td>
<td>97.2</td>
<td>1.6</td>
<td>3.2</td>
<td>0.941</td>
<td>0.034</td>
<td>0.068</td>
</tr>
<tr>
<td>AGTR1</td>
<td>87.3</td>
<td>0.0</td>
<td>3.4</td>
<td>99.3</td>
<td>1.0</td>
<td>1.9</td>
<td>94.5</td>
<td>0.6</td>
<td>2.6</td>
<td>0.888</td>
<td>0.014</td>
<td>0.054</td>
</tr>
<tr>
<td>AKT1</td>
<td>95.4</td>
<td>1.0</td>
<td>-1.4</td>
<td>91.3</td>
<td>4.4</td>
<td>6.6</td>
<td>94.1</td>
<td>2.1</td>
<td>1.1</td>
<td>0.864</td>
<td>0.049</td>
<td>0.028</td>
</tr>
<tr>
<td>BACE1</td>
<td>92.0</td>
<td>-1.1</td>
<td>-5.7</td>
<td>93.5</td>
<td>5.7</td>
<td>14.1</td>
<td>92.5</td>
<td>0.9</td>
<td>0.1</td>
<td>0.827</td>
<td>0.028</td>
<td>0.016</td>
</tr>
<tr>
<td>BCHE</td>
<td>85.6</td>
<td>7.1</td>
<td>-0.9</td>
<td>93.6</td>
<td>0.6</td>
<td>4.7</td>
<td>90.4</td>
<td>3.3</td>
<td>2.5</td>
<td>0.799</td>
<td>0.067</td>
<td>0.048</td>
</tr>
<tr>
<td>CASP1</td>
<td>69.1</td>
<td>5.3</td>
<td>-1.9</td>
<td>94.7</td>
<td>-0.1</td>
<td>-0.1</td>
<td>86.5</td>
<td>1.5</td>
<td>-0.8</td>
<td>0.680</td>
<td>0.039</td>
<td>-0.018</td>
</tr>
<tr>
<td>CASP3</td>
<td>84.8</td>
<td>3.4</td>
<td>3.7</td>
<td>94.9</td>
<td>-1.1</td>
<td>-1.1</td>
<td>91.0</td>
<td>0.6</td>
<td>0.7</td>
<td>0.809</td>
<td>0.013</td>
<td>0.015</td>
</tr>
<tr>
<td>CASP8</td>
<td>86.8</td>
<td>-4.9</td>
<td>-2.5</td>
<td>95.4</td>
<td>-2.1</td>
<td>-2.8</td>
<td>92.1</td>
<td>-4.1</td>
<td>-4.1</td>
<td>0.832</td>
<td>-0.060</td>
<td>-0.059</td>
</tr>
<tr>
<td>CHRM5</td>
<td>87.4</td>
<td>3.6</td>
<td>-4.2</td>
<td>95.3</td>
<td>1.5</td>
<td>4.0</td>
<td>92.3</td>
<td>2.3</td>
<td>0.9</td>
<td>0.835</td>
<td>0.048</td>
<td>0.015</td>
</tr>
<tr>
<td>CHUK</td>
<td>88.8</td>
<td>5.7</td>
<td>-3.3</td>
<td>97.4</td>
<td>0.4</td>
<td>0.0</td>
<td>95.2</td>
<td>1.7</td>
<td>-0.9</td>
<td>0.871</td>
<td>0.047</td>
<td>-0.024</td>
</tr>
<tr>
<td>CSF1R</td>
<td>94.3</td>
<td>5.9</td>
<td>-2.4</td>
<td>97.0</td>
<td>0.8</td>
<td>3.1</td>
<td>95.5</td>
<td>3.7</td>
<td>0.0</td>
<td>0.910</td>
<td>0.070</td>
<td>0.001</td>
</tr>
<tr>
<td>CSNK1D</td>
<td>91.4</td>
<td>12.4</td>
<td>3.8</td>
<td>94.8</td>
<td>-1.6</td>
<td>2.0</td>
<td>93.4</td>
<td>4.4</td>
<td>2.8</td>
<td>0.864</td>
<td>0.085</td>
<td>0.056</td>
</tr>
<tr>
<td>EDNRB</td>
<td>96.1</td>
<td>2.0</td>
<td>-0.5</td>
<td>94.4</td>
<td>-0.4</td>
<td>-0.7</td>
<td>95.1</td>
<td>0.6</td>
<td>-0.6</td>
<td>0.899</td>
<td>0.013</td>
<td>-0.012</td>
</tr>
</tbody>
</table>

## Comparative Model Performance

<table>
<thead>
<tr>
<th>Target Gene</th>
<th>SE</th>
<th>ΔSA</th>
<th>ΔRF</th>
<th>SP</th>
<th>ΔSA</th>
<th>ΔRF</th>
<th>ACC</th>
<th>ΔSA</th>
<th>ΔRF</th>
<th>MCC</th>
<th>ΔSA</th>
<th>ΔRF</th>
</tr>
</thead>
<tbody>
<tr>
<td>ELANE</td>
<td>90.9</td>
<td>-0.4</td>
<td>-5.2</td>
<td>93.8</td>
<td>0.2</td>
<td>5.8</td>
<td>92.1</td>
<td>-0.1</td>
<td>-0.8</td>
<td>0.839</td>
<td>-0.001</td>
<td>-0.012</td>
</tr>
<tr>
<td>EPHA2</td>
<td>87.2</td>
<td>0.0</td>
<td>-0.7</td>
<td>99.6</td>
<td>1.1</td>
<td>1.8</td>
<td>95.4</td>
<td>0.7</td>
<td>1.0</td>
<td>0.899</td>
<td>0.018</td>
<td>0.023</td>
</tr>
<tr>
<td>FGFR1</td>
<td>96.8</td>
<td>5.9</td>
<td>0.8</td>
<td>92.0</td>
<td>-2.8</td>
<td>1.4</td>
<td>95.1</td>
<td>2.8</td>
<td>1.0</td>
<td>0.892</td>
<td>0.054</td>
<td>0.021</td>
</tr>
<tr>
<td>FKBP1A</td>
<td>88.1</td>
<td>-3.0</td>
<td>-6.9</td>
<td>97.4</td>
<td>-0.4</td>
<td>1.5</td>
<td>94.8</td>
<td>-1.1</td>
<td>-0.9</td>
<td>0.869</td>
<td>-0.028</td>
<td>-0.025</td>
</tr>
<tr>
<td>FLT1</td>
<td>91.6</td>
<td>1.6</td>
<td>-3.0</td>
<td>99.0</td>
<td>0.9</td>
<td>2.5</td>
<td>96.2</td>
<td>1.2</td>
<td>0.4</td>
<td>0.919</td>
<td>0.024</td>
<td>0.008</td>
</tr>
<tr>
<td>FLT4</td>
<td>91.9</td>
<td>10.7</td>
<td>0.0</td>
<td>96.5</td>
<td>-0.4</td>
<td>1.2</td>
<td>94.5</td>
<td>4.3</td>
<td>0.6</td>
<td>0.888</td>
<td>0.086</td>
<td>0.014</td>
</tr>
<tr>
<td>FYN</td>
<td>77.8</td>
<td>5.1</td>
<td>-3.0</td>
<td>98.2</td>
<td>0.4</td>
<td>1.5</td>
<td>92.7</td>
<td>1.6</td>
<td>0.3</td>
<td>0.809</td>
<td>0.044</td>
<td>0.006</td>
</tr>
<tr>
<td>GSK3B</td>
<td>96.8</td>
<td>12.2</td>
<td>-1.1</td>
<td>78.3</td>
<td>-5.2</td>
<td>2.5</td>
<td>90.4</td>
<td>6.2</td>
<td>0.1</td>
<td>0.785</td>
<td>0.121</td>
<td>0.001</td>
</tr>
<tr>
<td>HDAC3</td>
<td>94.1</td>
<td>2.3</td>
<td>0.3</td>
<td>94.8</td>
<td>1.5</td>
<td>2.7</td>
<td>94.5</td>
<td>2.0</td>
<td>1.6</td>
<td>0.890</td>
<td>0.039</td>
<td>0.032</td>
</tr>
<tr>
<td>IGF1R</td>
<td>94.6</td>
<td>-1.0</td>
<td>-2.1</td>
<td>95.5</td>
<td>1.3</td>
<td>5.1</td>
<td>94.9</td>
<td>-0.2</td>
<td>0.3</td>
<td>0.889</td>
<td>-0.003</td>
<td>0.011</td>
</tr>
<tr>
<td>INSR</td>
<td>91.7</td>
<td>1.7</td>
<td>-3.1</td>
<td>98.9</td>
<td>1.2</td>
<td>1.9</td>
<td>95.5</td>
<td>1.4</td>
<td>-0.5</td>
<td>0.912</td>
<td>0.028</td>
<td>-0.007</td>
</tr>
<tr>
<td>KDR</td>
<td>97.4</td>
<td>2.8</td>
<td>-0.3</td>
<td>73.1</td>
<td>-9.7</td>
<td>9.2</td>
<td>93.5</td>
<td>0.8</td>
<td>1.2</td>
<td>0.748</td>
<td>0.005</td>
<td>0.058</td>
</tr>
<tr>
<td>LTB4R</td>
<td>91.4</td>
<td>4.3</td>
<td>2.2</td>
<td>98.8</td>
<td>0.0</td>
<td>1.2</td>
<td>96.8</td>
<td>1.2</td>
<td>1.5</td>
<td>0.918</td>
<td>0.030</td>
<td>0.038</td>
</tr>
<tr>
<td>LYN</td>
<td>89.1</td>
<td>10.9</td>
<td>-1.8</td>
<td>98.0</td>
<td>2.7</td>
<td>2.7</td>
<td>95.4</td>
<td>5.2</td>
<td>1.4</td>
<td>0.888</td>
<td>0.127</td>
<td>0.030</td>
</tr>
<tr>
<td>MAPK1</td>
<td>43.5</td>
<td>-4.4</td>
<td>2.3</td>
<td>99.2</td>
<td>4.0</td>
<td>-0.2</td>
<td>79.3</td>
<td>1.0</td>
<td>0.6</td>
<td>0.557</td>
<td>0.043</td>
<td>0.013</td>
</tr>
<tr>
<td>MAPK9</td>
<td>95.5</td>
<td>3.5</td>
<td>-2.9</td>
<td>97.7</td>
<td>2.2</td>
<td>7.9</td>
<td>96.5</td>
<td>2.9</td>
<td>2.0</td>
<td>0.931</td>
<td>0.058</td>
<td>0.040</td>
</tr>
<tr>
<td>MAPKAPK2</td>
<td>86.9</td>
<td>5.1</td>
<td>-3.3</td>
<td>94.1</td>
<td>1.4</td>
<td>2.5</td>
<td>91.0</td>
<td>3.0</td>
<td>0.0</td>
<td>0.816</td>
<td>0.062</td>
<td>-0.001</td>
</tr>
<tr>
<td>MET</td>
<td>97.7</td>
<td>4.3</td>
<td>-1.0</td>
<td>91.9</td>
<td>3.2</td>
<td>8.4</td>
<td>95.9</td>
<td>4.0</td>
<td>1.9</td>
<td>0.904</td>
<td>0.091</td>
<td>0.045</td>
</tr>
</tbody>
</table>

## Comparative Model Performance

| Target Gene | SE  | ΔSA | ΔRF | SP  | ΔSA | ΔRF | ACC | ΔSA | ΔRF | MCC | ΔSA | ΔRF |
|-------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| MMP13       | 94.9| 1.5 | -2.1| 94.7| 3.3 | 10.6| 94.8| 2.1 | 2.3 | 0.887| 0.045| 0.054|
| MMP2        | 95.5| 1.7 | 0.0 | 89.2| -3.6| 6.0 | 93.2| -0.2| 2.1 | 0.852| -0.006| 0.048|
| MMP3        | 94.7| 1.8 | -1.7| 92.4| 1.2 | 8.8 | 93.8| 1.5 | 2.1 | 0.868| 0.033| 0.047|
| MMP9        | 81.6| -0.2| -5.0| 88.8| -1.5| 6.9 | 84.6| -0.8| 0.0 | 0.696| -0.016| 0.011|
| NEK2        | 77.0| 9.4 | -2.7| 98.5| 0.7 | 1.0 | 94.0| 2.6 | 0.3 | 0.813| 0.085| 0.007|
| P2RY1       | 92.7| -2.4| -2.4| 100.0|1.5 | 3.4 | 97.7| 0.3 | 1.5 | 0.947| 0.007| 0.035|
| PAK4        | 89.4| 7.0 | -4.7| 99.3| 0.8 | 1.1 | 96.9| 2.2 | -0.3| 0.914| 0.064| -0.009|
| PDE4A       | 90.8| 2.9 | -1.1| 94.9| -0.4| 0.5 | 93.1| 1.0 | -0.3| 0.859| 0.020| -0.005|
| PDE5A       | 90.1| 3.4 | -4.8| 96.5| 2.8 | 7.3 | 93.0| 3.1 | 0.6 | 0.862| 0.062| 0.016|
| PIK3CA      | 98.9| 1.5 | -0.4| 93.3| -1.5| 4.6 | 97.2| 0.6 | 1.2 | 0.934| 0.014| 0.027|
| PPARG       | 69.5| -0.8| -2.9| 96.1| 3.4 | 2.0 | 86.0| 1.8 | 0.2 | 0.702| 0.041| 0.006|
| PTPN1       | 76.6| 9.5 | -3.9| 89.7| -2.6| 3.1 | 84.7| 2.0 | 0.4 | 0.673| 0.046| 0.005|
| PTPN11      | 64.8| 26.2| 14.8| 91.9| -3.4| -3.4| 85.7| 3.4 | 0.8 | 0.584| 0.153| 0.052|
| PTPN2       | 67.9| 2.5 | 2.5 | 96.0| -1.7| 2.3 | 90.1| -0.7| 2.4 | 0.687| -0.022| 0.069|
| RAF1        | 99.7| 4.2 | 0.0 | 95.2| -1.1| 1.5 | 97.7| 1.8 | 0.7 | 0.954| 0.038| 0.013|
| RARA        | 63.1| 1.9 | 1.9 | 99.4| 1.8 | 0.0 | 95.2| 1.9 | 0.3 | 0.742| 0.092| 0.013|
| RARB        | 85.9| 11.3| 5.6 | 99.9| 1.1 | 0.0 | 98.8| 1.9 | 0.5 | 0.913| 0.137| 0.033|
| ROCK1       | 94.3| 5.4 | -2.5| 92.0| -3.6| 1.1 | 93.2| 1.2 | -0.9| 0.864| 0.021| -0.018|

### Comparative Model Performance

<table>
<thead>
<tr>
<th>Target Gene</th>
<th>SE</th>
<th>ΔSA</th>
<th>ΔRF</th>
<th>SP</th>
<th>ΔSA</th>
<th>ΔRF</th>
<th>ACC</th>
<th>ΔSA</th>
<th>ΔRF</th>
<th>MCC</th>
<th>ΔSA</th>
<th>ΔRF</th>
</tr>
</thead>
<tbody>
<tr>
<td>RPS6KA5</td>
<td>73.7</td>
<td>12.3</td>
<td>0.0</td>
<td>100.0</td>
<td>1.5</td>
<td>2.7</td>
<td>95.3</td>
<td>3.4</td>
<td>2.2</td>
<td>0.835</td>
<td>0.135</td>
<td>0.080</td>
</tr>
<tr>
<td>SIRT2</td>
<td>70.8</td>
<td>2.3</td>
<td>5.6</td>
<td>95.2</td>
<td>-0.6</td>
<td>-1.6</td>
<td>89.8</td>
<td>0.0</td>
<td>0.0</td>
<td>0.692</td>
<td>0.003</td>
<td>0.006</td>
</tr>
<tr>
<td>SIRT3</td>
<td>76.7</td>
<td>-2.4</td>
<td>-4.7</td>
<td>98.5</td>
<td>0.8</td>
<td>0.8</td>
<td>95.4</td>
<td>0.3</td>
<td>0.0</td>
<td>0.802</td>
<td>0.010</td>
<td>-0.005</td>
</tr>
<tr>
<td>SRC</td>
<td>94.7</td>
<td>4.6</td>
<td>-2.7</td>
<td>88.7</td>
<td>0.7</td>
<td>8.1</td>
<td>92.4</td>
<td>3.1</td>
<td>1.5</td>
<td>0.839</td>
<td>0.064</td>
<td>0.029</td>
</tr>
<tr>
<td>TACR2</td>
<td>87.6</td>
<td>-1.6</td>
<td>-4.4</td>
<td>100.0</td>
<td>2.4</td>
<td>2.4</td>
<td>96.0</td>
<td>1.1</td>
<td>0.2</td>
<td>0.910</td>
<td>0.029</td>
<td>0.008</td>
</tr>
<tr>
<td>TBXA2R</td>
<td>88.8</td>
<td>-0.4</td>
<td>-2.1</td>
<td>94.9</td>
<td>-0.4</td>
<td>1.4</td>
<td>93.0</td>
<td>-0.4</td>
<td>0.3</td>
<td>0.838</td>
<td>-0.010</td>
<td>0.003</td>
</tr>
<tr>
<td>TEK</td>
<td>89.9</td>
<td>2.5</td>
<td>-4.1</td>
<td>97.8</td>
<td>2.6</td>
<td>1.9</td>
<td>94.5</td>
<td>2.6</td>
<td>-0.6</td>
<td>0.887</td>
<td>0.053</td>
<td>-0.013</td>
</tr>
</tbody>
</table>

Neural Network Activation Similarity

A mix of chemical and biological similarity
Amiodarone (hERG)
Amiodarone (hERG)
Amiodarone (hERG)

A Experimental: Active
  NNAS: 1.060
  Tanimoto: 0.974

B Experimental: Inactive
  NNAS: 1.060*
  Tanimoto: 0.954

C Experimental: Active
  NNAS: 1.030*
  Tanimoto: 0.943

D Experimental: Active
  NNAS: 1.000*
  Tanimoto: 0.917

E Experimental: Active
  NNAS: 1.000*
  Tanimoto: 0.957

F Experimental: Active
  NNAS: 0.996
  Tanimoto: 0.215

G Experimental: Active
  NNAS: 0.997
  Tanimoto: 0.180

H Experimental: Inactive
  NNAS: 0.999
  Tanimoto: 0.157

I Experimental: Inactive
  NNAS: 0.915
  Tanimoto: 0.154

J Experimental: Active
  NNAS: 0.970
  Tanimoto: 0.153
Quantitative Predictions

- Dose-Response Relationships and Risk Assessment Procedures require Quantitative Information
- Adjustment of AR dataset to contain only quantitative activity values (p(Activity), 4880 values)
- Change of loss function to MSE
- Single output node with linear activation function
- Models evaluated using MSE and RMSE
Quantitative Predictions

Quantitative Predictions at AR

$y = 0.992x + 0.2712$

$R^2 = 0.4885$

RMSE = 0.9546
Quantitative Predictions

Quantitative Predictions at AR

RMSE = 2.239

RMSE = 1.066

RMSE = 0.866

RMSE = 0.811

RMSE = 0.662

RMSE = 0.452

RMSE = 0.316

RMSE = 0.391

RMSE = 0.489

RMSE = 0.635

Average RMSE = 0.913

$y = 0.992x + 0.2712$

$R^2 = 0.4885$
Summary

- Neural Networks are a class of Machine Learning Algorithms that can provide both binary and quantitative predictions.
- Structural Alerts, Random Forests and Neural Networks have been used to try and predict binary activity at Human MIEs.
- A combination of these models and understanding of their workings is key to highest performance and model use in toxicology decision making.
- Quantitative predictions help push this methodology closer to use in risk assessment, rather than just hazard identification.
References


Acknowledgements

• Professor Jonathan Goodman
• Professor Anne Willis
• Unilever
• Dr. Paul Russell, Dr. Steve Gutsell & colleagues at SEAC, Unilever
• Dr. Andrew Wedlake
• Maria Folia and Dr. Sam Piechota
• The Centre for Molecular Informatics
• The MRC Toxicology Unit
• St. John’s College