Identifying Depression using classification trees

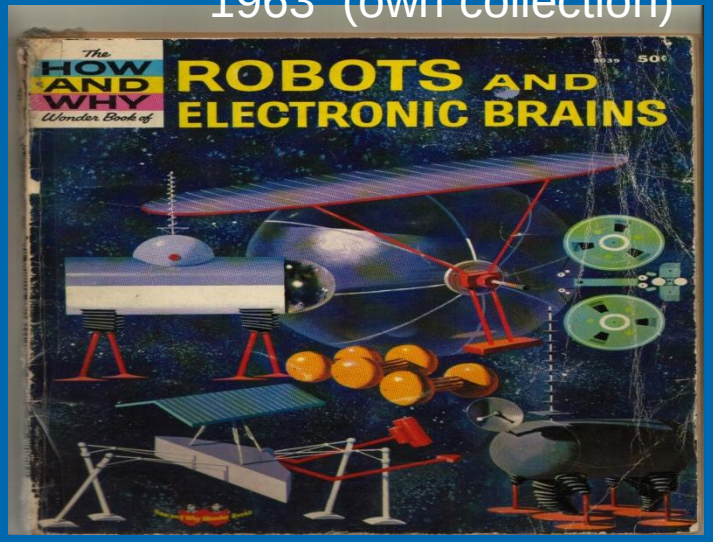
Dr Dean P. McKenzie BA(Hons), PhD

Senior Research Fellow & NHMRC Postdoctoral Fellow

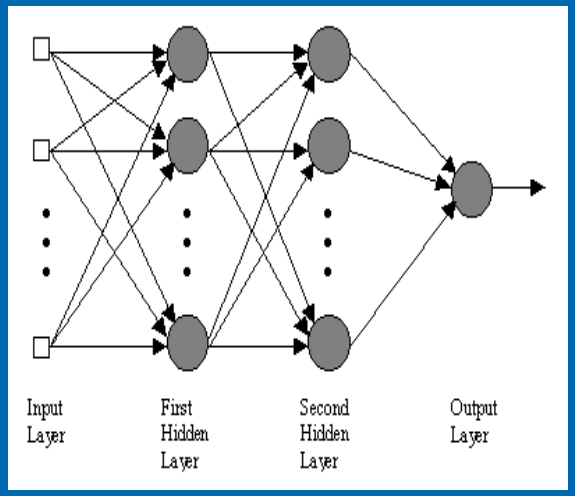
Department of Epidemiology & Preventive Medicine, ***Monash University,*** *Alfred Hospital, Prahran* ***&*** *Centre for Adolescent Health,* ***Murdoch Childrens Research Institute,*** *Royal Children's Hospital, Parkville*

Machine Learning

1963 (own collection)



1950's, ~1986

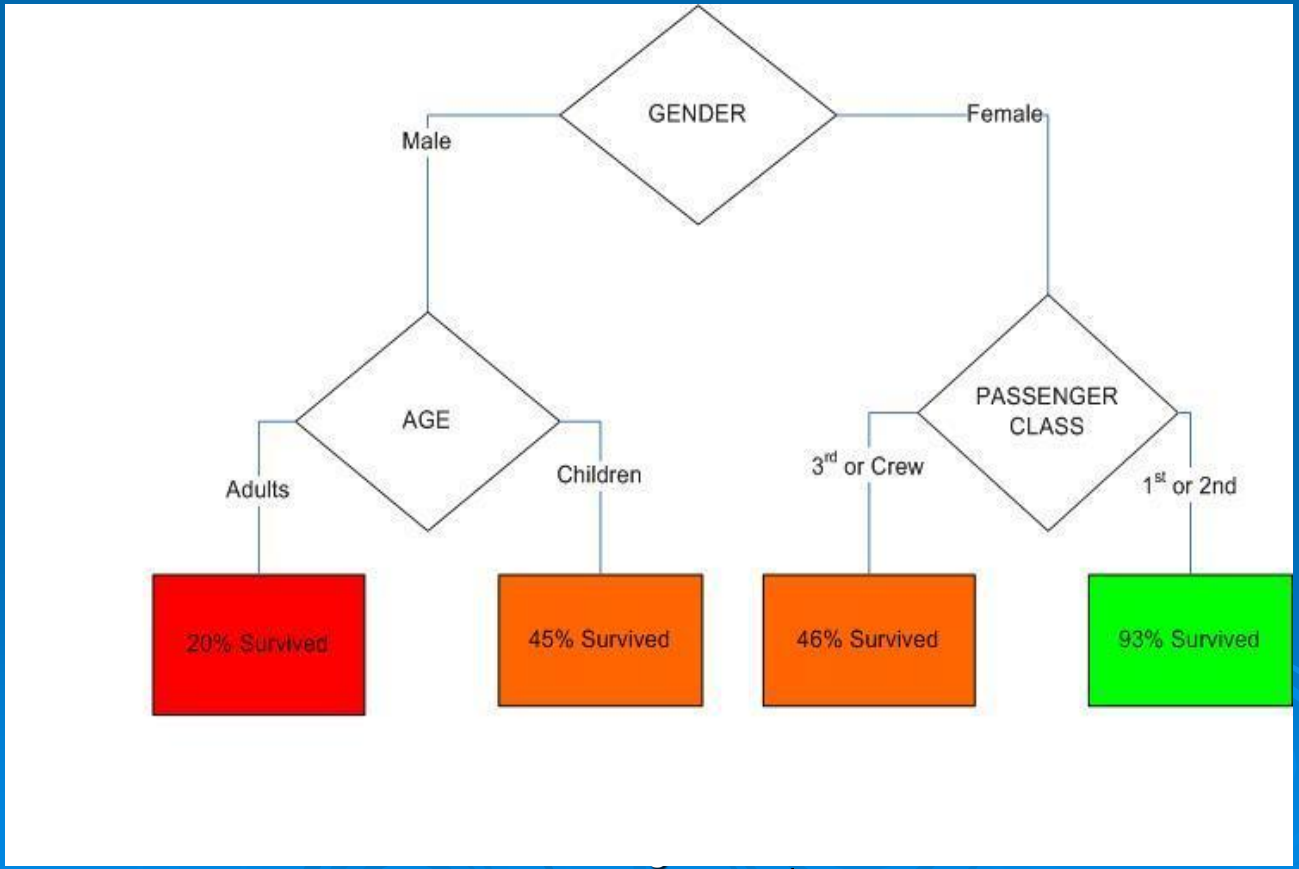


'*The study of machine learning focuses on developing computation methods for discovering new knowledge from data*'

Raynor 1999

Machine -> *Mechos* (Gr) 'Expedient', *Machina* (Lat) 'Device', *Mechonah* (Hebr) 'Foundation, Plan'

RMS/SS Titanic Sinking April 1912 (only 33% of 2223 survived)



50 years of Decision Trees

Originally developed in

Australia early 1950's (cluster analysis in botany, eg not prediction) (David Goodall)

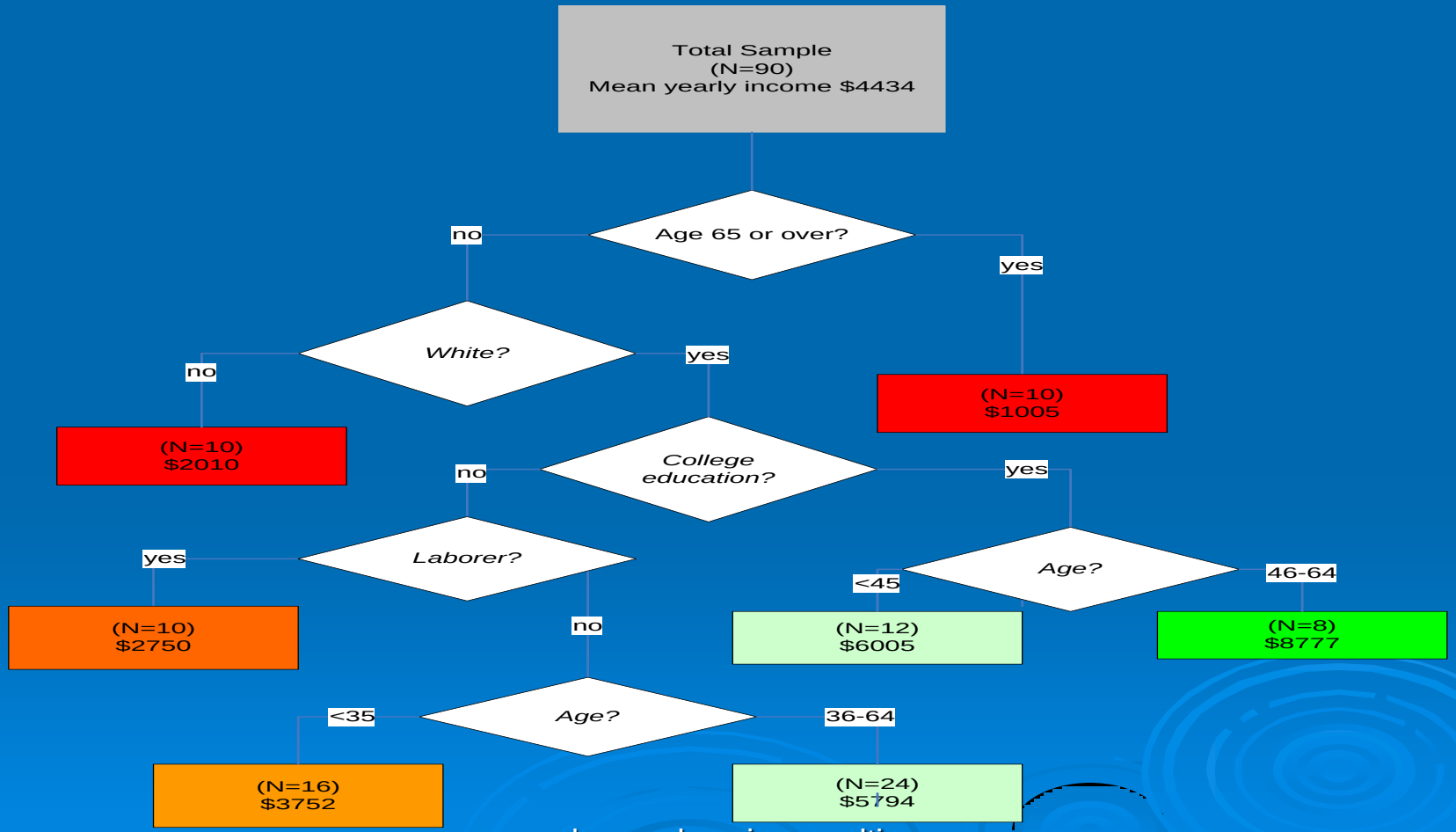
England mid 1950's (Australian, **William Belson**, BBC)

USA early 1960's by social survey researchers (**Morgan**, **Sonquist**)

Automatic Interaction Detection (AID)

also in USA / Australia by Cognitive Psychologists / Computer Scientists (**Hunt**, Marin, Stone, **Quinlan, Australian , of ID3/C4.5 fame**)

AID Income Analysis USA 1963



Early AID : too 'powerful'?

- Ad hoc 'stopping rules' based on group size, number of end groups and increase in explained variance
- Designed to be used with large datasets and cross-validated
- When misused was very controversial, **could and did find 'relationships' in totally random data!**

(Einhorn, *Alchemy in the Behavioural Sciences*, 1972)

Classification and Regression Trees (CART)

- 'Blue Book' by Breiman et al 1984
  Psychiatric data – 1991 alcohol risk
  1993 Schizophrenia diagnosis, McKenzie, **McGorry** et al
- helping tree-building become acceptable for the wider Statistical Community
- applications in *Addiction*, *American Journal of Epidemiology*, *American Journal of Psychiatry*,
- CART introduced important concept of backward **Pruning**

Tree Pruning (according to Yates)

- *If the pruning be too severe the tree will **grow to wood** instead of fruit'.*

(or as Einstein (never?) said 'everything should be as simple as possible, but not simpler')

*'if the branches are left too thick they ….encourage a great growth of leaves, **but very little fruit'.***

Arthur Yates & Co, Sydney 1956

CART Pruning / Cross-validation

- CART firstly grows large tree
- then prunes back taking into account **'bang'** (error rate) for **'buck'** (number of nodes)
- Final tree is chosen using cross-validation, eg build tree on 90%, test on 10%, repeat 10 times
- Very conservative (simulations and real data,)

- Extra safeguard : develop tree on 75% of sample, test on separate 25% , if performance not significantly different then combine

Case study

Pessimism, worthlessness, anhedonia and thoughts of death identify DSM-IV major depression in hospitalized medically ill

DP McKenzie, DM Clarke, AB Forbes & MR Sim

Monash University

Published in ***Psychosomatics, July 2010***

D E P R E S S I O N

- From the stem of *deprimere*, to **press down**

- ***'A pathological state of excessive melancholy, characterized by a mood of hopelessness, with feelings of inadequacy, and sometimes physical symptoms'***

Background : Depression in Hospitalised Medically Ill

- Depression common in hospitalised medically ill, **20 - 30%**
- difficult to identify as sadness and loss of interest often thought by medical staff / visitors to be part of illness / hospital experience
- DSM-IV etc view depression as homogenous (*Parker, 2006*), yet evidence of heterogeneity in presentation, course & treatment (*Rush, 2007*)

Background : Subtypes of depression

- Monash team led by Professor David Clarke looked at depression symptoms in the medically ill, using factor analysis & latent class cluster analysis *(Clarke et al, 2000; 2003)*
- We found major depression to be highly prominent in 2 classes

  1 scoring highly on **demoralization** factor (hopelessness, helplessness)

  1 scoring highly on **anhedonia** factor (loss of interest or pleasure)

Goal

- Identify specific symptoms of demoralization and anhedonia associated with DSM-IV major depression, **to aid understanding, screening/diagnosis & treatment**
- Examine overall relationships with logistic regression, uncover **possible subgroups of patients**, using **classification and regression trees** (**CART)**

Depression in medically ill: Design

- 312 Monash Medical Centre, Clayton hospital patients, GHQ-36 scores approximating those patients generally referred to consultation-liaison psychiatry
- Mean age 47.5, 61% female
- Current (past-month) DSM-IV major depression 19%

Potential confounders

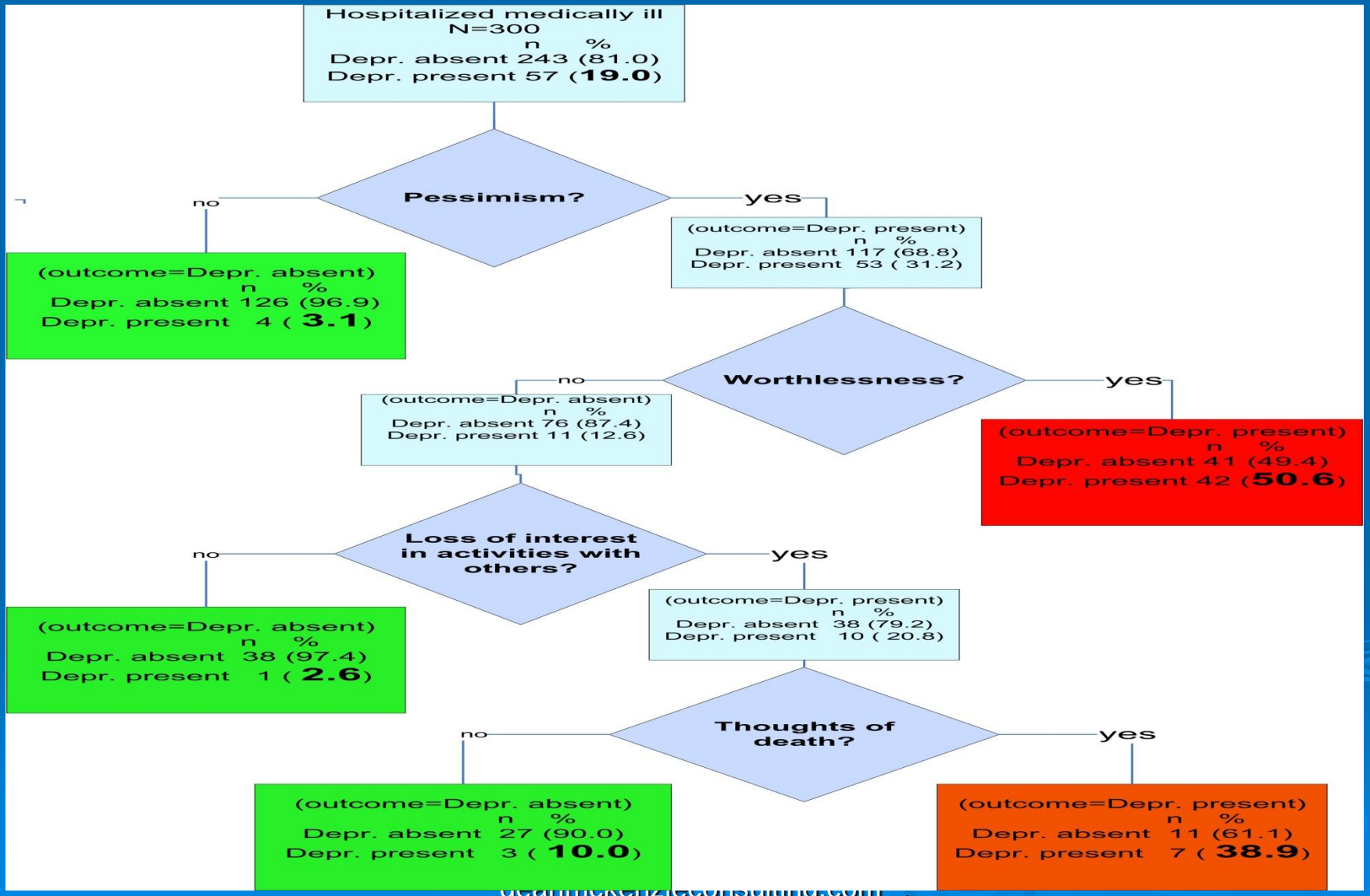
Age, gender, education, marital status, past psych history, illness severity

(controlled for using logistic regression, in initial model and subsequent analysis of CART subgroups)

Results : Logistic regression

Symptom	Prevalence (%)	Adjusted OR	95% CI
Despondency	75.3	2.7	0.5 - 14.6
Pessimism	**56.7**	**8.3**	**1.9 - 35.0**
Hopelessness	45.3	0.7	0.2 - 2.4
Helplessness	49.7	1.1	0.4 - 3.2
Worthlessness	**39.0**	**3.6**	**1.4 - 9.3**
Loss confidence	38.3	2.4	0.8 - 6.7
Thoughts death	**42.3**	**3.2**	1.2 - 8.4
Less interest in others	**55.7**	**3.6**	**1.1 - 12.0**

Results : CART decision tree



Summary

- Same 4 items important in logistic regression & CART
- CART suggests two subtypes of DSM-IV depression in hospitalised medically ill:
- 1/ combination of **pessimism & worthlessness**
- 2/ combination of **pessimism, loss of interest in others & thoughts of death**
- CART tree: sensitivity 86% & specificity 79%

Conclusions & Discussion #1

- Although not in DSM-IV criteria, **pessimism highly associated** with DSM-IV depression
- **Loss of interest in others**, **worthlessness** & **thoughts of death** also important
- (Worthlessness associated with in-hospital mortality (Furlanetto, 2000)

Conclusions & Discussion #2

worthlessness associated with suicidality (Spijker, 2010) and, in adolescents, persistent depression (Wilcox,2004) (ditto low self-worth; McKenzie/Toumbourou/Patton et al, Journal of Affective Disorders, 2011);

anhedonia persists post cognitive therapy (*Taylor et al, 2009*)

Specific symptoms may facilitate targetted screening / treatment

eg cognitive therapy for demoralization, drug therapy for anhedonia (suggested by Donald Klein 1964)

Regression & CART : Synergy

- Regression generally gives good results, but generates global models, may miss interactions or relationships that hold for some subgroups not others

- *Eg Swan et al, Addiction 2004, used CART and logistic regression together, to find different subgroups for males and females in terms of predictors of smoking cessation*

Tree Forests: Bagging & Boosting

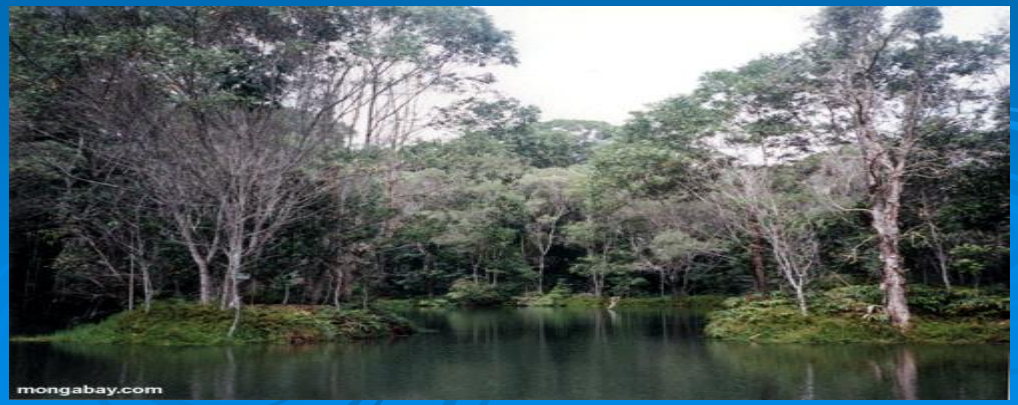
- CART etc extended in last 15 years to include
- *Bagging* : *Bootstrap Aggregating* : Parallel systems of many trees,–> *Random Forests*
- *Boosting* : Sequential systems of many trees, each one improving on the last
- *MultiBoost* (DU, Bagging & Boosting)

(strength in diversity : scope for Social Psychology)

Forests / Ensembles

- Classification Tree Forests or ‘Ensembles’ harder to interpret, but can be more powerful than single trees and give similar performance to Neural nets , (*although beware of the* ***Ford vs Holden*** *or* ***‘No Free Lunch’*** *syndrome*) but less parameters to set

- Large-scale problems often use Ensembles of Trees and or Ensembles of Neural nets, scope for smaller datasets, with validation



Preliminary additional results : depression dataset

	4 item CART	**50 bagged trees**	100 bagged trees
Sensitivity	86	**86**	87
Specificity	79	**84**	86
Positive predictive	48	**56**	61
Negative predictive	96	**96**	96

Quo Vadis : Where Next?

- Brief screens of depression, PTSD, that don't assume 'one size fits all' cut-offs
- Relationship between specific symptoms such as worthlessness and alcohol, profiling
- **Lots of Potential PhD Projects!!!!**

Acknowledgements

- National Health and Medical Research Council (funding of original study by Clarke et al) , and my PhD Scholarship, & Post-doc Fellowship)
- Monash Medical Centre, Clayton (Monash Health) Staff & Patients
- PhD supervisors *Malcolm Sim, David Clarke & Andrew Forbes*
- *Post-doc supervisors George Patton, Andrew Mackinnon; coauthors/collaborators David Clarke, John Toumbourou*

THANK YOU!