Gendered Data in Falls Prediction using Machine Learning

Presenter: Leeanne Lindsay
(Ulster University)

Authors:
Prof. Sonya Coleman
Dr. Dermot Kerr
Prof. Brian Taylor
Dr. Anne Moorhead

Ulster University

- BSc in Information Technology (First Class Honours)

- Studying PhD in Social and Computer Science

Overview

- Background
- TILDA Dataset
- Health Risk Factors for Older Adults
- Machine Learning
- Methodology
- Key Findings & Results
- Questions

The Irish Longitudinal Study on Ageing (TILDA) Dataset

- Over 8,500 ageing individuals took part in the questionnaire.

- Data collected in waves once every two years, focusing on adults aged 50+.

- Information focuses on their health and healthcare, pensions, housing and accommodation, mobility issues, education and their employment.

Context: Risk Factors for Older Adults

- Communication around the term ‘risk’ is vital for health and social care professionals, individuals and families.
- Risks can be positive or negative. Risks regarding the elderly are generally negative risks such as falling.
- Adults over the age of 65 years may be considered as a vulnerable population prone to having falls which may have huge consequences.

(Alzheimer's Society 2007, Stevenson & Taylor 2016)

Ulster University

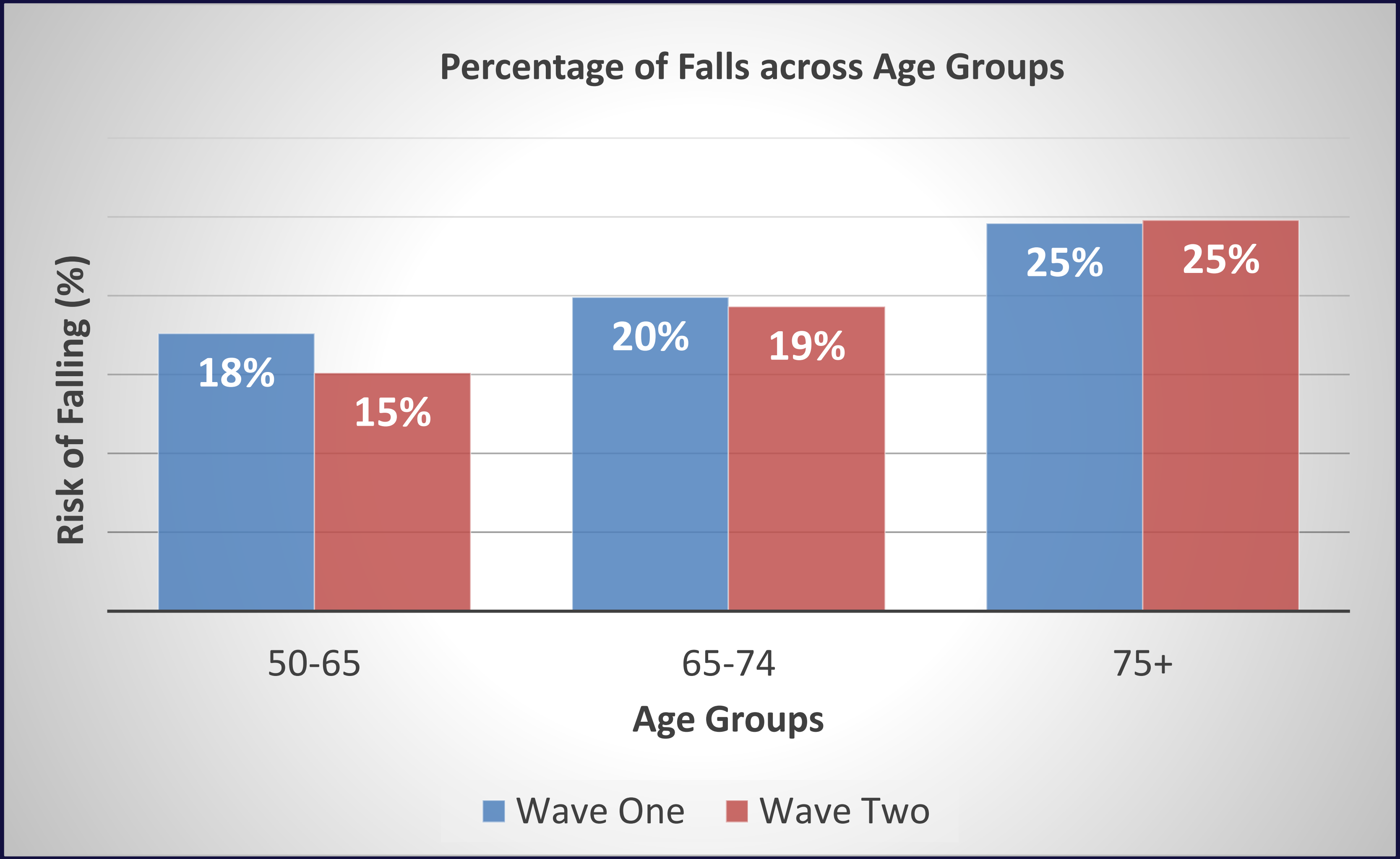
Risks of Falls

- **Cognitive decline in adults over 65 unfortunately relates to older adults falling which can also lead to recurrent falls.**

- **This study was using a machine learning approach to see if their were any gender differences in males and females using seven health risk factors to predict the risk of falls.**

Section 2

Risks of Falls in TILDA



Section 2

Background to Machine Learning

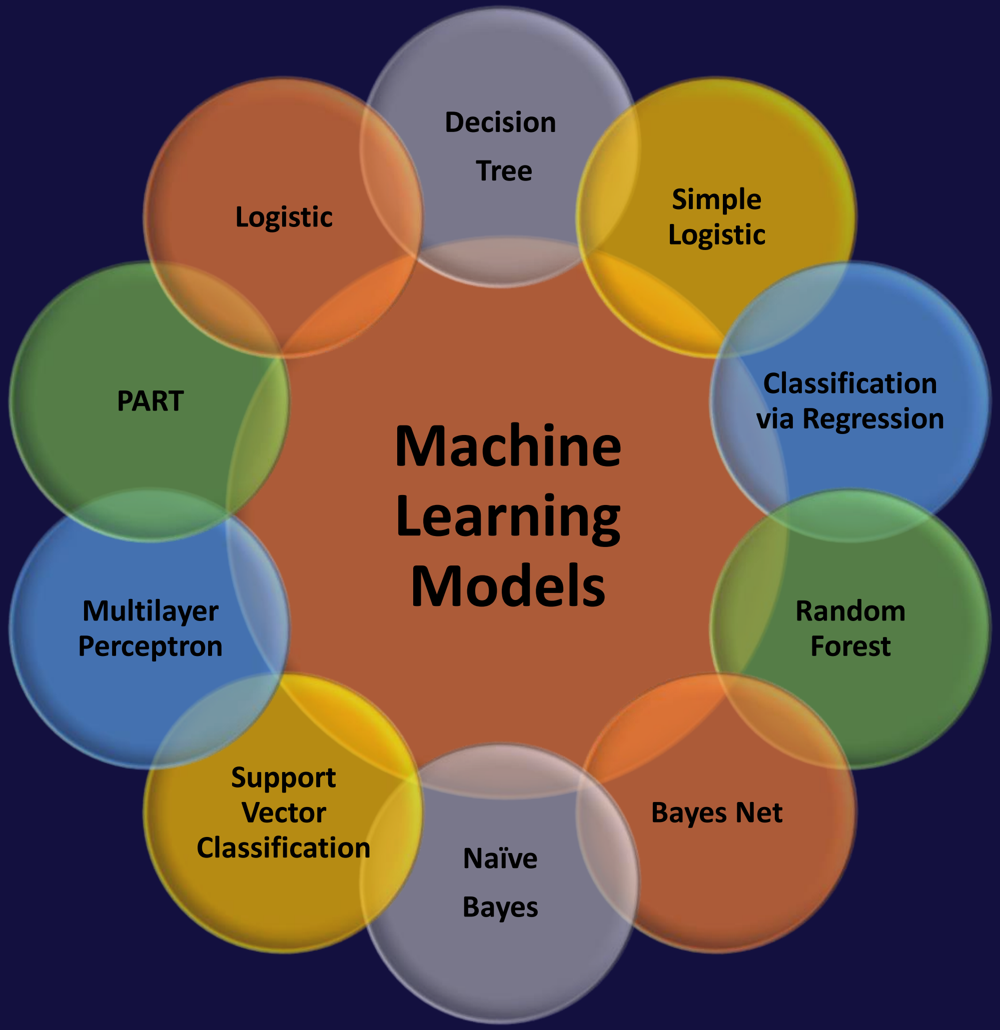
- … Analyse, design and develop systems with the intent to learn from the data.

- … previously used type 2 diabetes and breast cancer diagnosis.

- … Support Vector Machines (SVM), Decision Trees (J48) and Random Forest.

(Parodi 2009)

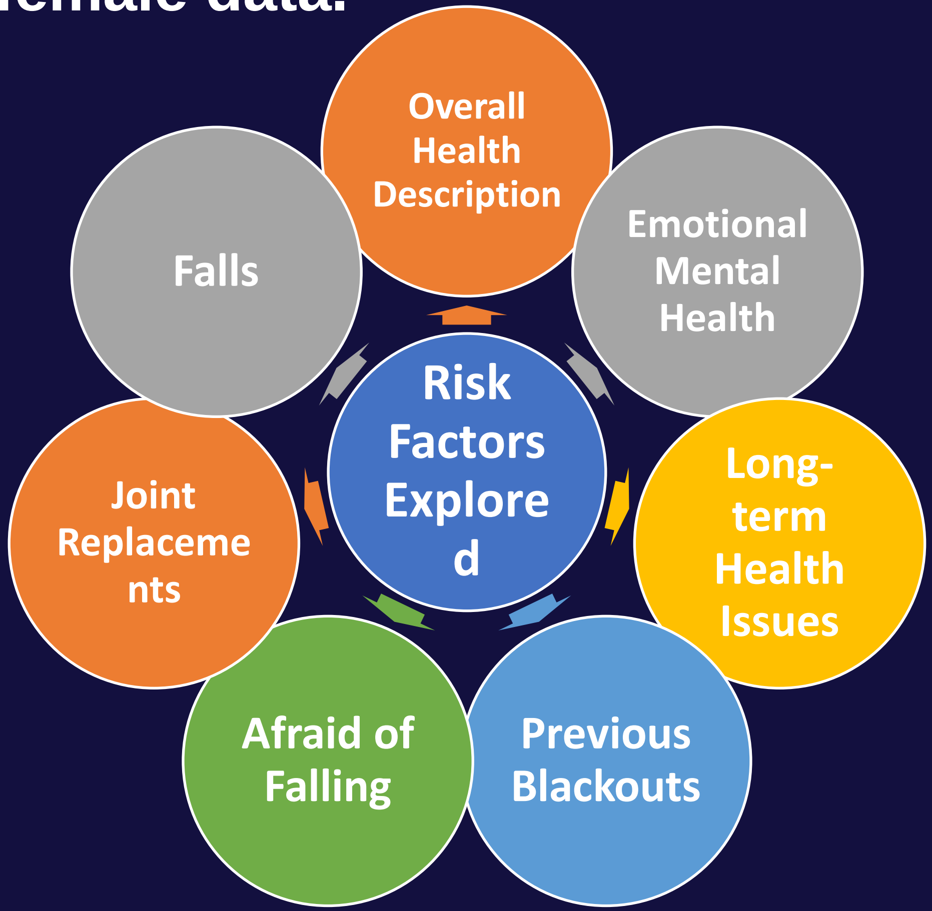
Machine Learning Models



Section 3

Methodology

This study focusses on predicting the likelihood of falls in older adults by performing experiments and analyses on separate male and female data.



Section 4

Methodology & Model Accuracy

- The objective was to explore the models produced by using different machine learning algorithms comparing the models for both the male data and female data.

- This allowed for the model to predict if there were any gender differences in the data.

- To be useful in practice and delivery these computer models must be understandable and acceptable to health and social care professionals to use in their daily job.

Section 4

Results

- The same health risk factors were inputted into each of the machine learning models to establish the accuracy score of each machine learning model.

- The results are in the normal range for social sciences work with self-declared, qualitative data, although higher accuracy results are often expected in computer science.

Section 5

Full Dataset Results

WEKA Classifier	Correctly Classified %
Naïve Bayes	61
Support Vector Classification	60
PART	60
Random Forest	57
Decision Tree	59
Bayes Net	61
Logistic	60
Multilayer Perceptron	56
Simple Logistic	60
Classification via Regression	62

- **Machine learning algorithm performance using the full dataset excluding male/female *(n=3242)*.**

Section 5

Full Dataset Results (Male & Female)

WEKA Classifier	Correctly Classified %
Naïve Bayes	64
Support Vector Classification	63
PART	61
Random Forest	58
Decision Tree	63
Bayes Net	64
Logistic	64
Multilayer Perceptron	57
Simple Logistic	66
Classification via Regression	66

- **Machine learning algorithm performance using the full dataset <u>including male/female *(n=3242)*</u>.**
- **All results have improved slightly by adding male and female into the algorithms for the full dataset.**

Section 5

Male Dataset Results

WEKA Classifier	Correctly Classified %
Naïve Bayes	58
Support Vector Classification	59
PART	58
Random Forest	57
Decision Tree	58
Bayes Net	58
Logistic	59
Multilayer Perceptron	57
Simple Logistic	59
Classification via Regression	59

- **Machine learning algorithm performance using only the male dataset *(n=1364)*.**
- **Male dataset results have reduced however the number of records have also reduced.**

Section 5

Female Dataset Results

WEKA Classifier	Correctly Classified %
Naïve Bayes	61
Support Vector Classification	60
PART	59
Random Forest	59
Decision Tree	60
Bayes Net	61
Logistic	61
Multilayer Perceptron	59
Simple Logistic	59
Classification via Regression	60

- **Machine learning algorithm performance using only the female dataset *(n=1364)*.**
- **Female dataset results have improved slightly in comparison to the male dataset shown previous using the same number of records.**

Section 5

Ulster University

Reduced Dataset Results

WEKA Classifier	Correctly Classified %
Naïve Bayes	60
Support Vector Classification	59
PART	57
Random Forest	56
Decision Tree	60
Bayes Net	60
Logistic	61
Multilayer Perceptron	56
Simple Logistic	61
Classification via Regression	60

- **Machine learning algorithm performance using <u>a reduced dataset including male and female *(n=1621)*</u>.**
- **For fair comparison we re-ran the experiment in Table II using only 50% of the dataset to compare against Table III and IV.**

Section 5

Overall Comparison of Results

- No significant differences in model performances.
- Comparing Table II with Table V demonstrates a slight decrease in predictive accuracy due to using fewer records.
- Using male and female data as an input variable demonstrates gender differences in the data.
- The results are not significant enough to justify the use of individual models for gender due to the smaller data set; a single model with gender as an input is sufficient to classify the data.

Section 5

Ulster University

Summary

- The TILDA dataset was utilised along with a number of machine learning algorithms.

- A reduction in the size of the dataset lowers predictive accuracy as expected, but splitting the data into male and female gives slightly higher predictive accuracy with the female data outperforming the male.

- The slightly higher predictive accuracy of the female compared to the male data suggests that the risk factors used are slightly more relevant for females than males based on this data. For this data it is apparent that separating male and female was beneficial.

Section 6

Ulster University

Thank you for your attention.

Leeanne Lindsay
Lindsay-l@ulster.ac.uk