Using Machine Learning to Predict Length of Stay and Discharge Destination for Hip-Fracture Patients

Mahmoud Elbattah, Owen Molloy
mahmoud.elbattah@nuigalway.ie

National University of Ireland, Galway

Background:
Hip Fracture Care in Ireland

Our Focus: Hip Fracture Care in Ireland

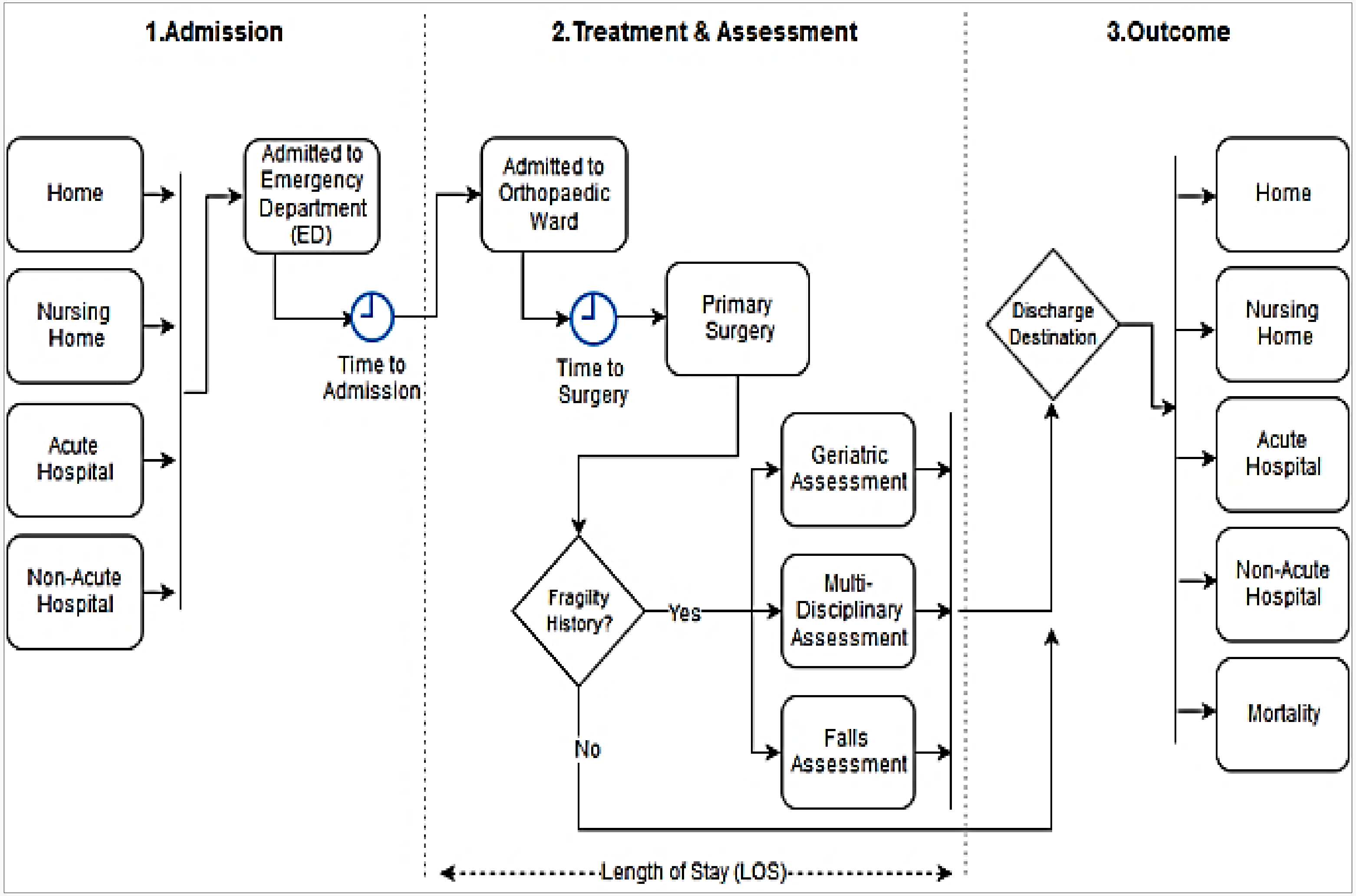


Figure1. Elderly Patient Journey

OÉ Gaillimh
NUI Galway

Study Objectives

Questions of Interest

1) How to predict the inpatient length of stay (LOS)?

2) How to predict the patient's discharge destination?

Significance of the Study

Why predicting inpatient LOS and discharge destination is important?

LOS:

- A significant measure of patient outcomes [1-3].
- A valid proxy to measure the consumption of hospital resources.
- Reported as the main component of the overall cost of hip fracture care [4].

Discharge Destination:

- Having a strategic importance in order to estimate the needed capacity of long-stay care facilities such as nursing homes.

Significance of the Study: A Bigger Picture

Patient-Focused Perspective

Machine Learning

Predict LOS and
Discharge Destination

+

Population-Driven Perspective

Simulation Modeling

Modeling Projected
Flow of Elderly Patients

Methodology

Overview

- Training a regression model for predicting the LOS.
- Training a multi-class classifier for predicting the discharge destination.

Data Description

- Irish Hip Fracture Database (IHFD).
- Patient records in the year 2013.
- Patients aged 60 and over.
- 38 data fields such as gender, age, type of fracture, date of admission, and LOS.

Data Anomalies: Outliers

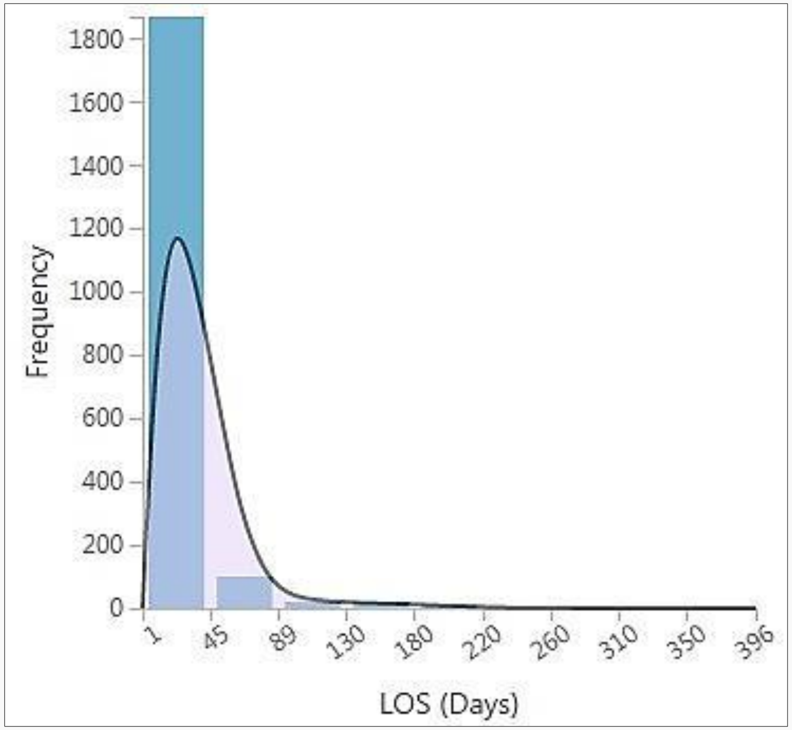


Figure 2. Histogram and probability density of the LOS variable. The outliers can be observed when the LOS becomes longer than 40 days.

Data Anomalies: Imbalances

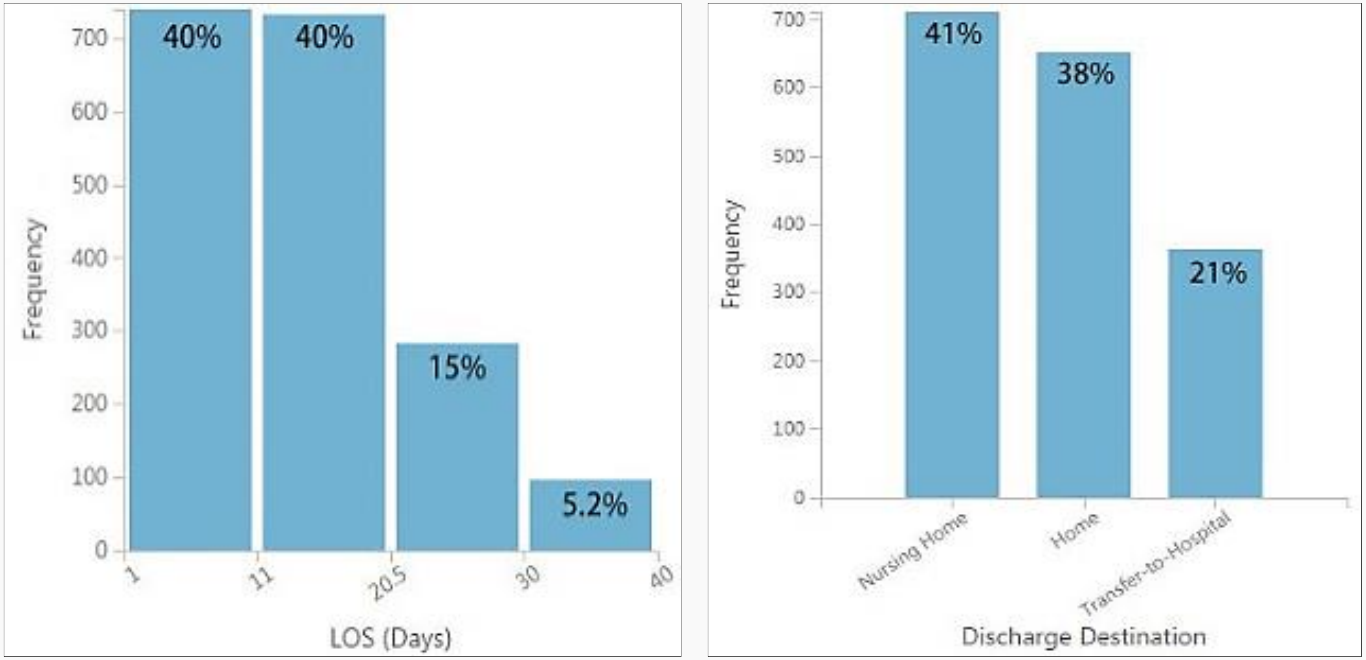


Figure 3. The imbalanced training samples, where figures (a) and (b) plot histograms of inpatient LOS and discharge destination respectively.

Feature Selection

LOS Regression Model:

- Hospital Admitted to
- Age
- ICD-10 Diagnosis
- Fracture Type
- Patient Gender
- Fragility History

Discharge Destination Classifier:

- Hospital Admitted to
- Age
- LOS
- Residence Area
- Patient Gender

* The features were decided based on the permutation importance method [5]

Learning Algorithm: Random Forests

Google

Scholar

Export

Leo Breiman 1928-2005

Random forests

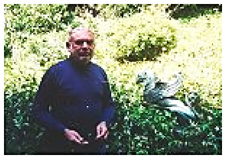
[PDF] from springer.com

Authors	Leo Breiman
Publication date	2001/10/1
Journal	Machine learning
Volume	45
Issue	1
Pages	5-32
Publisher	Springer Netherlands
Description	Abstract Random forests are a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest. The generalization error for forests converges as to a limit as the number of trees in the forest becomes large. The generalization error of a forest of tree classifiers depends on the strength of the individual trees in the forest and the correlation between them. Using a random selection of features to split each node yields error rates ...
Total citations	Cited by 23543

Scholar articles
Random forests
L Breiman - Machine learning, 2001
Cited by 23543 - Related articles - All 87 versions

Source: https://scholar.google.com/citations?user=mXSv_1UAAAAj

Paying Tribute to Leo Breiman (1928-2005)



Leo Breiman 1928-2005

Professor of Statistics, UC Berkeley
Data Analysis, Statistics, Machine Learning
Verified email at stat.berkeley.edu - Homepage

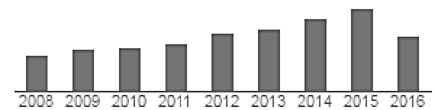
Follow

Title 1–20	Cited by	Year
Classification and Regression Trees L Breiman, JH Friedman, RA Olshen, CJ Stone CRC Press, New York	~~31603~~ *	1999
Classification and regression trees L Breiman Chapman & Hall/CRC	31603	1984
Random forests L Breiman Machine learning 45 (1), 5-32	23543	2001
Bagging predictors L Breiman Machine learning 24 (2), 123-140	14954	1996
Estimating optimal transformations for multiple regression and correlation L Breiman, JH Friedman Journal of the American Statistical Association, 580-598	1688	1985
Arcing classifier (with discussion and a rejoinder by the author) L Breiman The annals of statistics 26 (3), 801-849	1486	1998
Prediction games and arcing classifiers L Breiman Technical Report 504, Statistics Department, University of California at ...	~~1486~~ *	1997
Statistical Modeling: The Two Cutures L Breiman	1458 *	2003
Statistical modeling: The two cultures (with comments and a rejoinder by the author)	~~1454~~	2001

Google Scholar

Get my own profile

Citation indices	All	Since 2011
Citations	86161	44366
h-index	49	33
i10-index	78	48



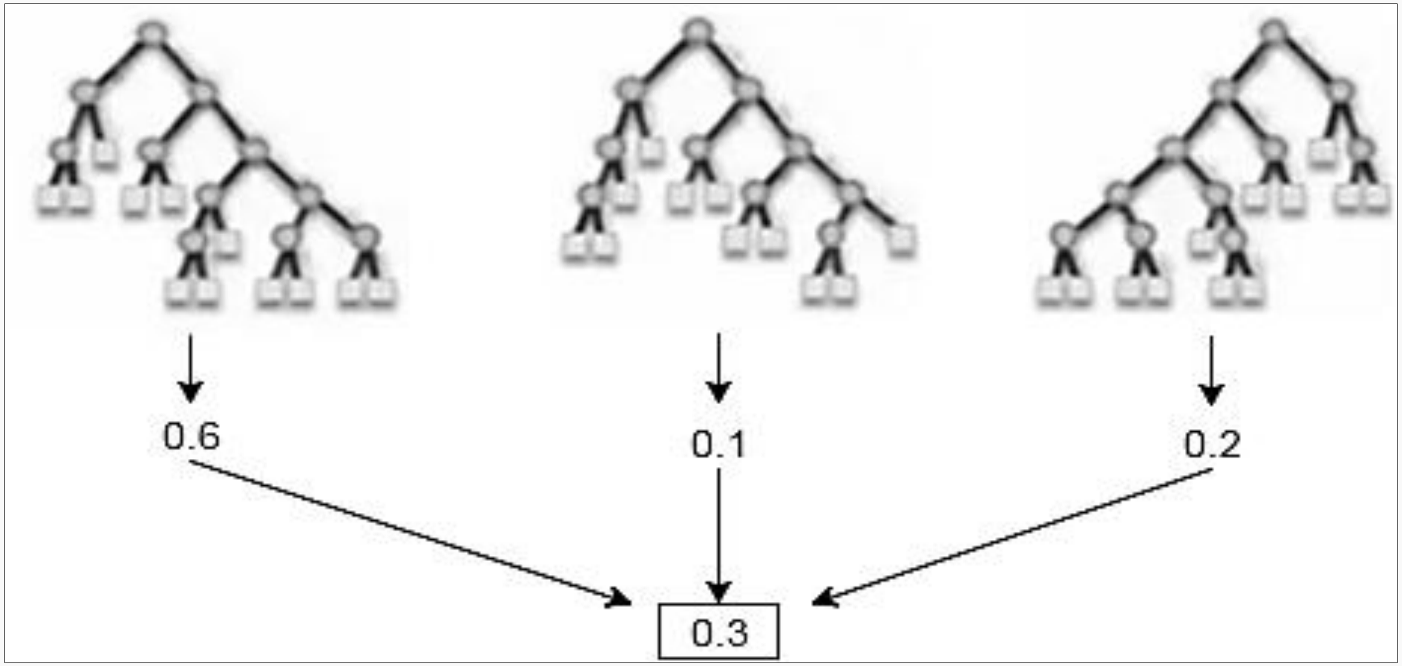
Learning Algorithm: Random Forests

- A Random Forest is a classifier consisting of a collection of tree-structured classifiers $\{h(x, \Theta_k), k = 1,...\}$ where the $\{\Theta_k\}$ are independent identically distributed random vectors and each tree casts a unit vote for the most popular class at input x.

- The common element in all of these procedures is that for the *k*th tree, a random vector Θ_k is generated, independently of the past random vectors $\Theta_1, \ldots, \Theta_{k-1}$ but with the same distribution; and a tree is grown using the training set and Θ_k , resulting in a classifier $h(\mathbf{x}, \Theta_k)$ where **x** is an input vector.

Source: Breiman, L., 2001. Random forests. *Machine learning*, *45*(1), pp.5-32.

Learning Algorithm: Random Forests



$$p(c|v) = \frac{1}{T}\sum_{t=1}^{T} p_t(c|v)$$

, where $p_t(c|v)$ denotes the posterior distribution obtained by the t-th tree.

Experimental Results

Experimental Results:
Regression Accuracy Based on 10-Fold Cross-Validation (LOS)

Algorithm	Relative Absolute Error (≈)
Random Forests	0.26
Boosted Decision Tree	0.34
Neural Network	0.55
Linear Regression	0.93

Experimental Results:
Classification Accuracy Based on 10-fold Cross-Validation (Discharge Destination)

Algorithm	Precision (≈)	Recall (≈)	Overall Classification Accuracy (≈)
Random Forest	0.88	0.87	0.88
Neural Network	0.71	0.72	0.72
Logistic Regression	0.61	0.60	0.62

Full-Text Paper

Conference Paper | Full-text available

Using Machine Learning to Predict Length of Stay and Discharge Destination for Hip-Fracture Patients

September 2016

DOI: 10.1007/978-3-319-56994-9_15

Conference: Proceedings of SAI Intelligent Systems Conference (IntelliSys) 2016 · At: London, United Kingdom

Project: Machine Learning Applied to the Irish Hip Fracture Database (IHFD)

Mahmoud Elbattah · Owen Molloy

- https://www.researchgate.net/publication/319198340_Using_Machine_Learning_to_Predict_Length_of_Stay_and_Discharge_Destination_for_Hip-Fracture_Patients
- https://link.springer.com/chapter/10.1007/978-3-319-56994-9_15

References

1. O'Keefe, G.E., Jurkovich, G.J. and Maier, R.V., 1999. Defining excess resource utilization and identifying associated factors for trauma victims. Journal of Trauma and Acute Care Surgery, 46(3), pp.473-478.
2. Englert, J., Davis, K.M. and Koch, K.E., 2001. Using clinical practice analysis to improve care. Joint Commission Journal on Quality and Patient Safety, 27(6), pp.291-301.
3. Guru, V., Anderson, G.M., Fremes, S.E., O'Connor, G.T., Grover, F.L., Tu, J.V. and Consensus, C.C.S.Q.I., 2005. The identification and development of Canadian coronary artery bypass graft surgery quality indicators. The Journal of thoracic and cardiovascular surgery, 130(5), pp.1257-e1.
4. Johansen, A., Wakeman, R., Boulton, C., Plant, F., Roberts, J. and Williams, A., 2013. National Hip Fracture Database: National Report 2013. Clinical Effectiveness and Evaluation Unit at the Royal College of Physicians.
5. Altmann, A., Toloşi, L., Sander, O. and Lengauer, T., 2010. Permutation importance: a corrected feature importance measure. Bioinformatics, 26(10), pp.1340-1347.
6. Breiman, L., 2001. Random forests. *Machine learning*, *45*(1), pp.5-32.
7. Brailsford, S. and Vissers, J., 2011. OR in healthcare: A European perspective. *European journal of operational research*, *212*(2), pp.223-234.

THANK YOU!

mahmoud.elbattah@nuigalway.ie