World Academy of Science, Engineering and Technology
International Journal of Medical and Health Sciences
Vol:13, No:6, 2019

Classification of Health Risk Factors to Predict the Risk of Falling in Older Adults

L. Lindsay, S. A. Coleman, D. Kerr, B. J. Taylor, A. Moorhead

***Abstract*—**Cognitive decline and frailty is apparent in older adults leading to an increased likelihood of the risk of falling. Currently health care professionals have to make professional decisions regarding such risks, and hence make difficult decisions regarding the future welfare of the ageing population. This study uses health data from The Irish Longitudinal Study on Ageing (TILDA), focusing on adults over the age of 50 years, in order to analyse health risk factors and predict the likelihood of falls. This prediction is based on the use of machine learning algorithms whereby health risk factors are used as inputs to predict the likelihood of falling. Initial results show that health risk factors such as long-term health issues contribute to the number of falls. The identification of such health risk factors has the potential to inform health and social care professionals, older people and their family members in order to mitigate daily living risks.

***Keywords*—**Classification, falls, health risk factors, machine learning, older adults.

I. Introduction

INCREASED risk of falling has become an important issue among older adults often leading to loss of confidence in living at home, loss of mobility and reduced social contact as well as increased costs of health and social care services [1]. The Health Service Executive in Ireland, states that 30% of adults over the age of 65 years, who live at home within a community setting, will fall at least once per year [1]. Older adults who live in a residential nursing setting have a staggering 50% risk of falling. Everyone has the potential risk of having a fall, however those over 65 years old are known as a vulnerable group of people prone to falls due to, for example, cognitive decline [1]. Additionally, the fear of falling can be linked to someone having at least one previous fall, leading to more recurrent falls [2].

As risks are becoming high profile in the world of decision making, health and social care professionals are focusing more attention on risks [3]. Decisions are made every day by professionals regarding the safety and wellbeing of older adults, including those with dementia, as they need to communicate these potential risks to the individual and their carer's. Risks can be viewed as positive or negative. However, risks to the elderly are generally negative risks such as falling [4], burns [5], driving, wandering, forgetting about medication or taking too much medication [6]. Fig. 1 illustrates a number of categorized risk factors related to older adults and those with dementia.

Risk communication involves the sharing of information and opinions through professionals, carers, families and the individual with dementia [7]. This allows for better decisions to be made about the person in question within the shared decision-making process [8], [9]. Adults with dementia suffer from the risk of poor communication due to short-term memory loss, so it is vitally important for professionals to liaise closely with other family members and professionals to get to know the individual on a more personal level. They may be able to inform professionals of what risks they are currently taking like wandering and if the person becomes agitated there may be a way of calming them by taking the time to listen to the adult and respect that they are suffering with dementia [10]. Humans use their own knowledge and expertise to predict different outcomes every day. In this study, the focus is placed on estimating the negative risk of falling in older adults using computational intelligence techniques.

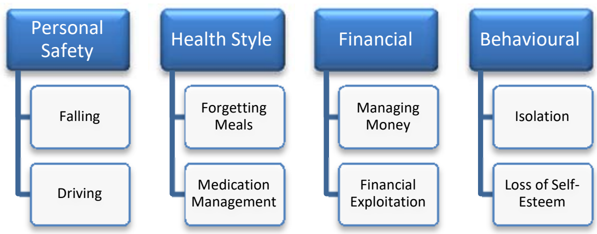

Fig. 1 Examples of risks associated with the ageing population

Computational intelligence incorporates algorithms such as Artificial Neural Networks (ANNs), Fuzzy Logic Systems (FLS), Genetic Algorithms (Gas) and many more. Machine learning is a methodology which helps to analyse, design and develop systems [11] with the intent to learn from data [12]. Machine learning (ML) algorithms can be supervised or unsupervised. Supervised can be broken down into classification or regression. Brownlee [13] describes the process as a teacher supervising a learning process where the learning will come to an end only when the algorithm outputs an acceptable variable [13]. Unsupervised learning deals only with input variables to learn from the data; there is no teacher and no target value. An example of a ML technique is a Support Vector Machine (SVM). This is an example of a nonlinear classification or regression approach to computational intelligence [14]. Additionally, decision trees

L. Lindsay and B. J. Taylor are with the Institute for Research in Social Sciences, Ulster University, N. Ireland, UK (e-mail: lindsay-l@ulster.ac.uk, bj.taylor@ulster.ac.uk).

S. A. Coleman and D. Kerr are with the School of Computing and Intelligent Systems, Ulster University, N. Ireland, UK (e-mail: sa.coleman@ulster.ac.uk, d.kerr@ulster.ac.uk).

A. Moorhead is with the School of Communication and Media, Ulster University, N. Ireland, UK (e-mail: a.moorhead@ulster.ac.uk).

are becoming a popular classification algorithm allowing results or calculations to be easily understood and therefore simple to interpret [15]. The classification via Regression method performs the classification using a regression method and for every class value a regression model is built [16].

ML is used within a number of healthcare areas. For example, a data mining approach was used to determine risk factors linked with type 2 diabetes using a decision tree and random forest approach [17]. Another example of the use of ML with health and social care data is breast cancer diagnosis [18]. Breast cancer diagnosis is typically based upon the decision of a radiologist by taking a mammography test. ML technology has now been introduced to collect and automatically analyse the data. Previously, SVMs were implemented to separate tumours and classify them into tumour types in the diagnosis of breast cancer. This previous research has been able to reduce the computation time without losing the accuracy of the diagnoses for the patient.

Considering the previous research in ML with health and social care data, this research combines various ML approaches with fall data to predict the likelihood of falling. This paper is organised as follows: in Section II the dataset used is presented and the proposed methodology is described in Section III. Results are presented in Section IV and the paper is concluded in Section V with future work proposed.

II. TILDA Dataset

The aim of this study is to explore the potential of risk analysis by using ML to inform our understanding of combining algorithms with health risk factors from, TILDA dataset to predict the likelihood of falling and mitigate the risk of hazards. The TILDA dataset was used for this study as it provided substantial information and collected resourceful data on risk factors that were originally taken from a previous project 'Risk Communication in Dementia'. The TILDA dataset is a cohort study of approximately 8,000 ageing individuals aged from 50 onwards including both males and females based in the Republic of Ireland (RoI). Data used in this study were taken from Wave One and Wave Two, which were collected over a period of time from 2009-2013. Wave One represented a random sample of people (753) who were part of the sampling framework for the study. Wave Two consisted of a pre-interview questionnaire and a face-to-face computer assisted personal interview with the same random sample of people as Wave One apart from those who had passed away. Wave One and Wave Two both consist of different aspects of the lives of older people in Ireland including their health and healthcare, pensions, housing and accommodation, mobility issues, education and their employment. The data were collected by carrying out face-to-face interviews conducted by trained professionals after consent was given by the participants.

The dataset was collected in three steps, firstly a face-to-face interview containing questions based on socio-demographic, wealth, lifestyle, social support and health. A self-declaration questionnaire was then completed and lastly a detailed health assessment was carried out by trained professionals. If participants could not leave their home to attend a self-assessment health centre, professionals attended their homes and carried out tests regarding cognitive ability, mobility, strength, bone and vision tests.

The TILDA dataset has previously been used to examine the impact of risk factors for coronary heart disease in older Irish adults. Analysis in this study was stratified by gender, age group and socio-economic position then weighted on the related disability and compared against another dataset in Northern Ireland [19]. However, to date, these data have not been used to predict falls, which is the focus of this paper.

All personal attributes in Wave One and Wave Two of the TILDA dataset are used, including falls and health-related attributes. Throughout Wave One and Wave Two there were qualitative and numerical attributes. The numerical attributes such as number of falls were determined using the question: "How many times have you fallen in the last year?" The falls attribute in the TILDA dataset was originally the exact number of times fallen which has been modified for two-class classification, falls and no falls. The six attributes used as input factors to the ML models are as follows: first was the overall health description, participants were asked to choose from excellent, very good, good, fair or poor; secondly, to define their emotional mental health using the same answers, excellent, very good, good, fair or poor; the third attribute was a yes or no answer derived from the question: 'Have you any long-term health issues?' The fourth attribute used was a yes or no answer to the question: 'Have you previously had a blackout or fainted?' and participants were also asked if they were afraid of falling as this could directly relate to an increased risk of falling; and lastly, participants were asked if they had any joint replacements which used a yes and no response. All of these responses were turned into numerical values for the ML models.

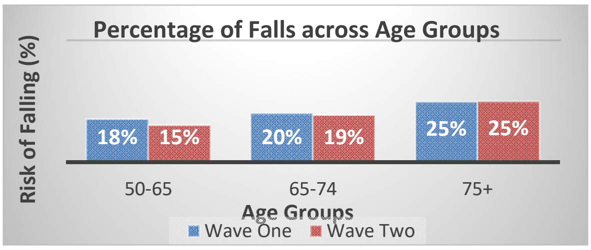

Fig. 2 Percentages of Falls (n=8000)

Ageing and falling are becoming a popular research topic, however the research does not specifically point to only one attribute as being the main cause of falls. Usually more than one attribute is responsible due to the complex nature of risk factors.

The data in Fig. 2 illustrate the percentage of falls across three different age groups and the slight difference in Wave One and Wave Two. It can be noticed that the percentages are increasing with the age group. Overall, those aged between 50-65 years represent 57% of the total dataset. The 65-74 year old category made up 26.5% and the 75+ year old group consisted of 16.5% of the total dataset, respectively.

World Academy of Science, Engineering and Technology
International Journal of Medical and Health Sciences
Vol:13, No:6, 2019

III. Methodology

A number of ML techniques were selected within the WEKA ML environment and used to train models that predict if an individual will fall within the next year. Inputs to the models included, Overall Health Description, Emotional Mental Health, Long-term Health Issues, Previous Blackout/Fainting, Fear of Falling, and Joint Replacement. The target model output is a binary classification, fall or no fall. The dataset was split into training and testing sets where the sets are composed of 90% training samples and 10% testing samples, respectively. The two classes, fall and no falls, each comprised of 1,621 values for both Wave One and Wave Two. Ten-fold cross validation was used to verify the model accuracy.

When training each of the models, each input risk factor was added incrementally one by one to establish whether or not the risk factor had an effect on falls. This took a very simplistic approach to distinguish which risk factors were of more importance and build upon each risk factor.

The results from the various models for Wave One and Wave Two are presented in Table I and Table II, respectively. There are no significant differences in the model performance, however the models with the highest classification accuracy when predicting falls are highlighted in green. In Wave One, the best model was the Classification via Regression classifier which predicted with an accuracy of 62%. In Wave Two, there were three models with identical performance, Decision Tree, Simple Logistic and Classification via Regression all with an accuracy of 69%.

TABLE I
WEKA Results for Wave One

WEKA Classifier	Correctly Classified %
Naïve Bayes	0.61
SMO	0.60
PART	0.60
Random Forest	0.57
Decision Tree	0.59
Bayes Net	0.61
Logistic	0.60
Multilayer Perceptron	0.56
SGD	0.59
Simple Logistic	0.60
Classification via Regression	**0.62**

Fig. 3 presents a Receiver Operating Characteristic (ROC) which is generated by plotting the true positive rate against the false positive rate at different likelihoods for the Classification via Regression method using Wave Two. The Area Under the ROC Curve (AUROC) is commonly used to provide a measure of the performance where optimal performance will have a value of 1.0. For the Classification via Regression method the AUROC is 0.696.

Fig. 4 presents a ROC curve, for the Decision Tree method using Wave Two with an AUROC value of 0.579. Similarly, Fig. 5 presents a ROC curve for the Logistic method using Wave Two with an AUROC value of 0.886.

TABLE II
WEKA Results for Wave Two

WEKA Classifier	Correctly Classified %
Naïve Bayes	0.63
SMO	0.64
PART	0.67
Random Forest	0.68
Decision Tree	**0.69**
Bayes Net	0.63
Logistic	0.67
Multilayer Perceptron	0.67
SGD	0.66
Simple Logistic	**0.69**
Classification via Regression	**0.69**

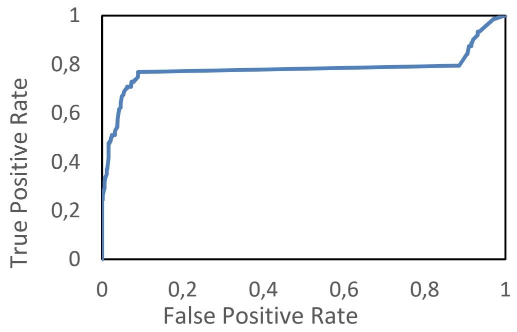

Fig. 3 ROC for Classification via Regression

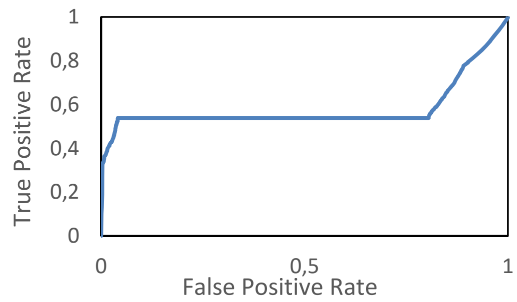

Fig. 4 ROC for Decision Tree

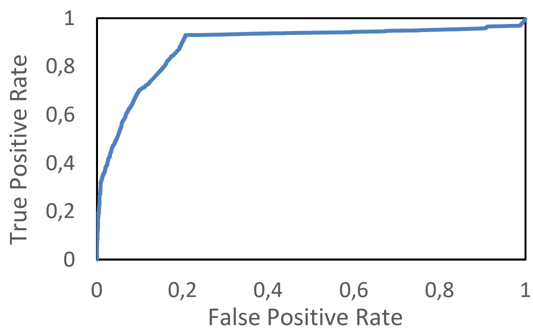

Fig. 5 ROC Curve Plotted for a Logistic Model

IV. Conclusion

This study has explored the relationship between the health and social care factors by utilising different ML algorithms such as Decision Trees. The health and social care factors explored were Overall Health Description, Emotional Mental Health, Long-term Health Issues, Previous Blackout/Fainting, Fear of Falling, and Joint Replacements and the target output was the risk of falling in older adults. The objective was to explore the models produced by different ML algorithms from the TILDA dataset, with the goal of identifying the most accurate algorithm which predicted falls in older adults within the context of the dataset.

This study has successfully trained models with varying performance accuracy. The Classification via Regression method produced the best result using both waves of the data although the accuracy did not vary greatly between the models compared.

Building on the success of these pilot risk analyses, further work on this study will explore more of the health risk factors associated with older adults and in particular those with dementia to determine which factors have the most significant impact on fall prediction. The English Longitudinal Study of Ageing (ELSA), will also be analysed alongside TILDA to analyse further risk factors, so as to better inform hazard mitigation for professionals and family members caring for someone in older age

To conclude, this study presented a ML approach to predict falls in older adults using health risk factors, whereby results have reflected that these risk factors are somewhat contributing to falls in older adults.

REFERENCES

[1] Health Service Executive (2018) Conditions & Treatments - Falls. Available at: https://www.hse.ie/eng/health/az/ (Accessed: 04/12/2018).

[2] Costa, B., Rutjes, A., Mendy, A., Freund-Heritage, R. and Vieira, E. (July 17, 2012) Can Falls Risk Prediction Tools Correctly Identify Fall-Prone Elderly Rehabilitation Inpatients? A Systematic Review and Meta-Analysis, doi: doi.org/10.1371/journal.pone.0041061.

[3] Stevenson, M., McDowell, M. and Taylor, B. (2017) Concepts for communication about risk in dementia care: A review of the literature. Doi: 10.1177/1471301216647542.

[4] Allan, LM., Ballard, CG., Rowan, EN., Kenny, RA. (2009) Incidence and Prediction of Falls in Dementia: A Prospective Study in Older People. doi: 10.1371/journal.pone.0005521.

[5] Oyebode, J. R., P. Bradley, and J. L. Allen. (2013) Relatives' Experiences of Frontal-Variant Frontotemporal Dementia, pp. 156–166. doi: /10.1177/1049732312466294.

[6] While, C., Duane, F., Beanland, C., Koch. (2013) Medication management: The perspectives of people with dementia and family carers. Doi.org/101177/1471301212444056

[7] Alzheimer's Society (2007) Home from home: *A report highlighting opportunities for improving standards of dementia care in care homes*

[8] Godolphin, W. (2009) Shared Decision Making. Healthcare Quarterly, 12, pp. e186 - e190. doi: http://www.longwoods.com/content/20947.

[9] Thornton, H. (2003) Patients' understanding of risk Enabling understanding must not lead to manipulation. British Medical Journal, 327, pp. 693 - 694. doi: doi.org/10.1136/bmj.327.7417.693.

[10] Stevenson M & Taylor BJ (2016) Risk communication in dementia care: family perspectives. *Journal of Risk Research, 18(1-2),* 1-20. doi:10.1080/13669877.2016.1235604.

[11] Wang, K. (2003) *Intelligent Condition Monitoring and Diagnosis Systems: A Computational Intelligence Approach.* Amsterdam: IOS Press.

[12] Parodi, P. (2009) Computational intelligence techniques for general insurance, pp. 1 - 167.

[13] Brownlee, J. (March 16 2016) *Understand Machine Learning Algorithms-Supervised and Unsupervised Machine Learning Algorithms.* Available at: https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/.

[14] Meyer, David. (2001). Support Vector Machines The Interface to libsvm in package e1071. R News. 1.

[15] Suresh, K. and Dillibabu, R. (2018) Designing a Machine Learning Based Software Risk Assessment Model Using Naïve Bayes Algorithm.

[16] Bal, R. and Sharma, S. (May 2016) Review on Meta Classification Algorithms using WEKA.

[17] Rafiq, M., McGovern, A., Jones, S., Harris, K., Tomson, C., Gallagher, H. and de Lusignan, S. (2014) *Falls in the elderly were predicted opportunistically using a decision tree and systematically using a database-driven screening tool.*

[18] Zheng, B., Yoon, S.W. and Lam, S.S. (2014) *Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms.*

[19] Cruise, S.M., Hughes, J., Bennett, K., Kouvonen, A. and Kee, F. (2017) *The impact of risk factors for coronary heart disease on related disability in older Irish adults.* Doi: doi.org/10.1177/0898264317726242.