An Evidential Reasoning Rule Based Feature Selection for Improving Trauma Outcome Prediction

Fatima Almaghrabi, Dong-Ling Xu, Jian-Bo Yang
Alliance Manchester Business School, the University of Manchester,
Manchester, M15 6PB, United Kingdom
E-mail: fatima.almaghrabi@manchester.ac.uk

Abstract: Various demographic and medical factors can be linked to severe deterioration of patients suffering from traumatic injuries. Accurate identification of the most relevant variables is essential for building more accurate prediction models and making more rapid life-saving medical decision. The intention of this paper is to select a number of features that can be used to accurately predict patients' outcomes through three feature selection methods: random forest, ReliefF and the evidential reasoning (ER) rule. The impact of an outcome's class imbalance on feature selection is discussed, and synthetic minority over-sampling technique (SMOTE) is performed to show the differences in the selected features. The results show that length of stay in hospital, length of stay in intensive care unit, age and Glasgow Coma Scale (GCS) are the most selected features across different techniques. The prediction models based on the features selected by the ER rule show the highest prediction performance represented by the area under the receiver operating characteristic curve (AUC) values, which has a median of 0.895 for the model employed by the ten highest-weighted variables, while the median AUC values are 0.827 and 0.885 if the ten highest-weighted variables are selected by ReliefF and random forest respectively. The results also show that after the ten most important features, increasing the number of the less important features has only a slight increase in prediction accuracy.

Keywords: Feature selection; Trauma; Evidential reasoning rule; Random forest; ReliefF; Imbalance classes.

1. Introduction

Trauma is a major public health issue and a significant cause of mortality and disability worldwide. In 2010, there were 17,201 injury-related deaths in England and Wales. For every death following an injury, there were approximately 10 people who survived, potentially with serious permanent disabilities. In addition, trauma is the leading cause of mortality in the under-40 age group [1]. As such, many medical solutions are introduced to minimise the effect of traumatic injuries and therefore minimise trauma mortality rates. One solution involves increasing the ability to predict outcomes for a trauma patient with a high degree of prediction accuracy. Many medical decisions, such as choice of treatment location, can be based upon the recognition of the severity of patients' condition, and identifying patients with the highest severity can aid medical teams to prioritise treatments. Therefore, outcome prediction techniques and prognostic models have received considerable attention in trauma literature in recent years.

Deciding which variables are useful to predict negative effects on patient outcomes can guide medical trauma teams in their swift attempts to treat trauma patients [1]. However, many factors are linked to poor outcomes for patients suffering from traumatic injuries. Several demographic and medical factors, such as age and post-injury health conditions, are associated with the risk of deterioration after traumatic injuries. Nevertheless, identifying the most relevant ones can help develop better models, as emphasis can be placed on the best-known key features. Previous work includes comparison of different factors to identify the most influential ones in patients' outcomes and medical conditions. Some variables can be redundant as they serve the same purpose. The Charlson Comorbidity Index and the modified Charlson Comorbidity Index, for instance, serve the same purpose of providing a score for patients' existing conditions. Data reduction methods provide a solution for minimising inputs in prediction models. However, feature selection methods have the advantage of selecting the key features without altering the collected variable [2]. The choice of a sufficient number of variables that highly predict patients' outcomes can produce an applicable and efficient model

[3], reduce the possibilities of overfitting, and provide faster and more cost-effective models [2].

Recently, data mining in the fields of medicine and healthcare has been more widely applied by developers and academic researchers [4]. The feature selection and classification of imbalanced data sets are two of the most interesting machine learning (ML) challenges [5]. Feature extraction and selection techniques have provided various solutions to handle the increasing high dimensionality in data. For example, in the traumatic injuries field, recent work involving the automation of feature extraction have improved the selection of trauma features from data images [6], brain signals [7], and radiology examination reports [8]. Moreover, feature selection methods are used in several applications, including distinguishing between the healthy and injured arms of elbow trauma patients [9], identifying infectious complications following trauma laparotomy prediction [10], and improving the outcome prediction of patients with severe brain trauma [11]. Although feature selection methods have been applied in the medical domain [3,12], trauma literature includes only a few previous studies that have utilised feature selection techniques to rank and identify the most relevant features.

The evidential reasoning (ER) rule [13] has been introduced to advance the Dempster–Shafer evidence theory and the ER algorithm. The applications of the ER rule have been extended recently to include data classification [14]. The ER rule combines evidence from multiple features while calculating the weight and reliability of evidence to make a classification decision. When a non-linear optimisation of a performance measure is combined with the process, this method can optimise the weight value of each piece of evidence and therefore determine the importance of the respective features in the classification decision. The present work proposes an optimisation-based feature selection method using the ER rule and presents the process of applying the method using trauma as an example. Other feature selection methods, such as the random forest (RF) algorithm [15] and the ReliefF algorithm [16], have been applied to identify the variables that are consistently influential across these three techniques. It also considers the imbalanced data issue during the selection process and while applying the synthetic minority over-sampling technique (SMOTE) to balance the outcome.

This paper contributes to the growing area of feature selection research by exploring different techniques for the selection of the most influential features in the area of trauma. It

presents a process for feature selection through the ER rule classifier, which provides an explainable and probabilistic method for determining the importance of features. Additionally, the paper seeks to highlight feature selection techniques and their importance in building more accurate prognostic models. The techniques should be regarded as tools that assist practitioners in deciding on inputs and outputs for efficient evaluation models. This can help to facilitate medical decision-making processes by considering the most contributory factors. Based on the previous objectives, this paper sets out to answer the following questions:

1. What are the key features that can be used to predict the outcomes of traumatic injuries? The range of features included in this paper exceeds the number of features used in most of the existing trauma feature selection (FS) literature and contains patients' relevant demographics features, injury incident features and common medical variables.
2. How comparable is the performance of the ER rule to some of the previous FS techniques? The ER rule is compared to two common FS techniques: a filter FS technique, which is the ReliefF algorithm [16] and variables' importance weight extracted from the RF algorithm [15], which is an embedded FS technique.
3. Would the highly imbalanced classes in trauma data affect the selected features of each technique? Are the differences in the selected features statistically significant? When the set of the selected features varies after balancing the classes, this indicates that the feature selection method is sensitive to class imbalance and select the feature that are associated more with the majority class, which is the 'alive' outcome of trauma patients in this case.
4. What is the optimal set of features that has high prediction performance with the least number of variables? Which FS algorithm produce the highest prediction accuracy with the least number of variables? This paper shows the trade-off between prediction accuracy and the number of variables of each technique.

First, this paper reviews the feature selection methods previously used in trauma literature to identify the most important variables. Second, detailed descriptions for the feature selection processes of the ER rule, the RF algorithm and the ReliefF algorithm are provided in addition to an illustration of the feature selection procedure proposed in this paper. Next, experimental results are presented in the results section. Finally, a discussion of the results is provided and followed by conclusions.

2. Literature Review

Many different variables are extracted from trauma patients, such as their demographics, previous comorbidities, injury characteristics, and medical conditions. Choosing the most important variables can simplify a prediction model [17] and help to reduce model over-fitting in cases in which the number of variables is very high compared to the number of patients [18]. Furthermore, some variables degrade the performance of a prediction model [19]. One of the first experiments to compare changes in performance using different variables in trauma prediction models was carried out by Hunter et al. [20]. They compared the three variables (revised trauma score, age, and injury severity score (ISS)) prediction model with the full set of TRISS score variables model; unexpectedly, the model with three variables outperformed the model with the whole set of variables. A research study conducted by Jakaite and Schetinin [21] introduced Bayesian model averaging technique, which was integrated in decision trees for feature selection for trauma patient death risk evaluation based on 16 variables collected when in hospital and from screening tests. Subsequently, an ensemble of decision trees was run, and the weakest variable was excluded from the trees. The paper concluded that the proposed method which selected nine variables achieved similar prediction accuracy results to the whole set of variables. An application of ReliefF algorithm for feature selection was performed by Farago et al. [9] and found that it improved class separation while reducing model complexity. Gelbard et al. [10] executed a backward elimination feature selection to predict infectious complications that may occur after trauma laparotomy. However, due to the specific domain of the paper and the particular type of sepsis, the sample size and the number of variables were relatively small. A recent paper by Bennis et al. [11] combined leave-one-out cross-validation with forward feature selection to rank the prediction feature before applying logistic regression. The model was designed to predict favourable outcomes following traumatic brain injury. However, the ranking varied between patients due to some of the correlated variables.

Using optimisation algorithms to define the features that contribute the most to maximizing the prediction accuracy is considered as one of the techniques for feature selection. One of the previous examples in trauma was carried out by Wu et al. [18] . Their paper optimised the SVM algorithm twice for a higher area under the ROC curve (AUC) score and a higher f-measure using sequential feature selection. One of the main findings of the paper was

that fewer than ten features were actually performing better in the prediction's sensitivity and precision compared to the whole set of features. Another finding was that the AUC-optimised model classified the test set more accurately and consistently. Lee et al. [12] designed a feature selection model for a group of paediatric of patients and applied the particle swarm optimisation technique for feature selection coupled with an optimisation-based discriminant analysis model. The feature selection method was designed to find the variables associated with the return to the ED within 72 hours for the different acuity levels and with accuracy of more than 80%. All the selected variables achieved accuracy of 85% and above.

One issue that has been discussed in the feature selection domain is class imbalance. Most feature selection techniques find the relevant key features that predict majority classes. Therefore, previous work has used two main methods to consider this imbalance: balancing classes using imbalance class handling techniques either during the FS process or over the training data before applying FS techniques; choosing performance measures such as AUC score, F-measure and the geometric mean scores for most of the studies that are less sensitive to imbalance in classes.

Yin, Ge, Xiao, Wang, and Quan [22] introduced the class-decomposition and the Hellinger distance techniques for class imbalance feature selection. The Hellinger distance technique distribution is not based on class information, which makes it less sensitive to the high skew in the classes. The class-decomposition divided the majority classes into smaller pseudo-subclasses. The proposed techniques were combined with some traditional feature selection techniques such as correlation and mutual information. With the application in four medical datasets, the F-measure of major class, the F-measure of minor class, and AUC were implemented to find the technique with the most accurate prediction performance. The two techniques showed good results; however, a comparison should be made with other class imbalance FS techniques. SYMON is another technique that was created to combine the dimensionality reduction of features with a focus on the minority class without changing the structure of the data [23]. The technique uses symmetrical uncertainty to rank features based on their dependency on the class label and is not based on the loss function. The performance of the proposed method showed a similar, or in some datasets a better, geometric mean than other FS methods with respect to class imbalance techniques such as SMOTE-RLF and SVM-

BFE techniques and across the different datasets. However, this method consumes a large amount of computational time.

Maldonado et al. [5] generated backward elimination for feature selection using various misclassification costs with SVM algorithm for binary classification of five microarray datasets. An oversampling SMOTE was applied over the training data for the imbalance in the classes and the average of feature ranking over five cross-validation was chosen during the process. The geometric mean was measured as a performance measure. The results showed that the SMOTE technique notably affected the performance results of the most imbalanced datasets. This implies that the differences in FS performance will be affected by the SMOTE technique when the data is highly imbalanced.

Previous work by González- Robledo, Martín-González, Sánchez- Barba, Sánchez-Hernández, and Moreno- García [24] highlighted the class imbalance issue with feature selection specifically for trauma data. An ensemble-based classifier was introduced to deal with feature selection to improve accuracy. The paper focused on predicting the neurological outcome of severely injured trauma patients. Two types of feature selection techniques were implemented: correlation-based FS, which highlights the variables that are mostly correlated with the outcome and the redundancy between the variables; and calculating the information gain of features. To deal with the class imbalance issue, the ensemble of the classifiers was used to minimise the issues with oversampling and under-sampling techniques. However, the paper showed that the feature selection techniques did not change the accuracy or solve the class imbalance issue directly in the FS process, and there were only 497 patients in the study. A summary of the previous literature in trauma is presented in Table 1. The table shows examples of various types of feature selection methods and their trauma application.

Additionally, a number of previous studies have examined various ML algorithms to predict different outcomes following traumatic injuries. In a review of previous implementations of ML in trauma, Liu and Salinas [25] reported a common conclusion regarding the benefits of ML in improving outcome prediction and found that around half of the included studies used an artificial neural network algorithm. Other ML applications include algorithms such as RF [26], support vector machine [27,28], belief rule-based inference [29], interpretable machine learning (IML) [30], gradient boosting [28], and naïve Bayes [28].

Moreover, establishing the most discriminative ML algorithm has been implemented for various types of traumatic injuries. For example, the application of machine learning methods to predict burn injuries outcomes has been provided in Stylianou, Akbarov, Kontopantelis, Buchan, and Dunn [27]. This paper found that all the implemented ML algorithms performed well in predicting the outcome. However, artificial neural network performed the best, according to the AUC, and random forest was the best choice after optimising for equal sensitivity and specificity. It concluded that there is no significant relationship between ML algorithm model complexity and performance.”.

Table 1. Feature selection methods in trauma research

Feature selection method type	Feature selection technique	Trauma application	Reference
Filters	ReliefF	Trauma	[9]
	correlation-based FS and information gain	severe traumatic injuries	[24]
Wrappers	backward elimination feature selection	trauma laparotomy	[10]
	leave-one-out cross-validation with forward feature selection	traumatic brain injury	[11]
Embedded	Bayesian model averaging over decision trees	Trauma	[21]
	SVM algorithm optimisation	American Football Head Impacts	[18]
	particle swarm optimisation technique coupled with an optimisation-based discriminant analysis model	Paediatric patients	[12]

To summarise, feature selection techniques have been implemented widely in the medical domain. However, to date, few studies have been carried out to apply feature selection methods to trauma and highlighted the key predictive features. Previous studies highlighted the

relationship between the imbalance in the classes and the feature selection process. Most of the imbalance class feature selection methods presented in the existing literature are for high-dimensional data with a large number of features and fewer cases. Moreover, the imbalanced classes feature selection techniques have not been significantly highlighted in trauma data considering the highly skewed classes of injuries' outcomes. Therefore, this paper implements different feature selection techniques to rank a variety of demographic and clinical features collected from trauma patients. AUC optimisation integrated in the ER rule classifier process is implemented for feature selection and the prediction accuracy using the AUC statistic of the technique's selected features is measured. The paper also tests the effect of the SMOTE technique, which handles the class imbalance issue, in the feature selection results.

3. Methods

The process of feature selection followed in this work starts with normalisation of features to remove the effects of different scales across features. Afterwards, three times 10-fold cross-validation is applied to determine different training and testing folds. Based on the previous folds, the three feature selection techniques are applied. The choice of feature selection algorithm is one of the challenges of feature selection since each algorithm may perform differently depending on the dataset, as well as having its own advantages and restrictions [31]. Moreover, knowledge of existing feature selection algorithms in the application domain helps in reaching an informed decision. Here, the proposed feature selection method based on the ER rule is employed, in addition to two methods previously applied to trauma data that have provided valid results, namely the RF and ReliefF algorithms [9,28]. After obtaining the importance of each variable, the SMOTE technique is implemented to the training data of the 30 folds, creating synthetic examples on the training sets. Following that, the same procedure of feature selection is applied, and different sets of important features are created. The prediction performances of the models based on the features ranked most important according to the imbalanced class feature selection results are compared by considering five different models 10, 20, 30, 40 features and the whole dataset based on the ranked variables. Then, 30-fold cross-validation is applied to split the data before applying the support vector machine (SVM) algorithm, and the AUC is used as a performance measure with which to compare the

results. This process is summarised in Fig. 1. All of the present analysis is conducted using MATLAB 2017.

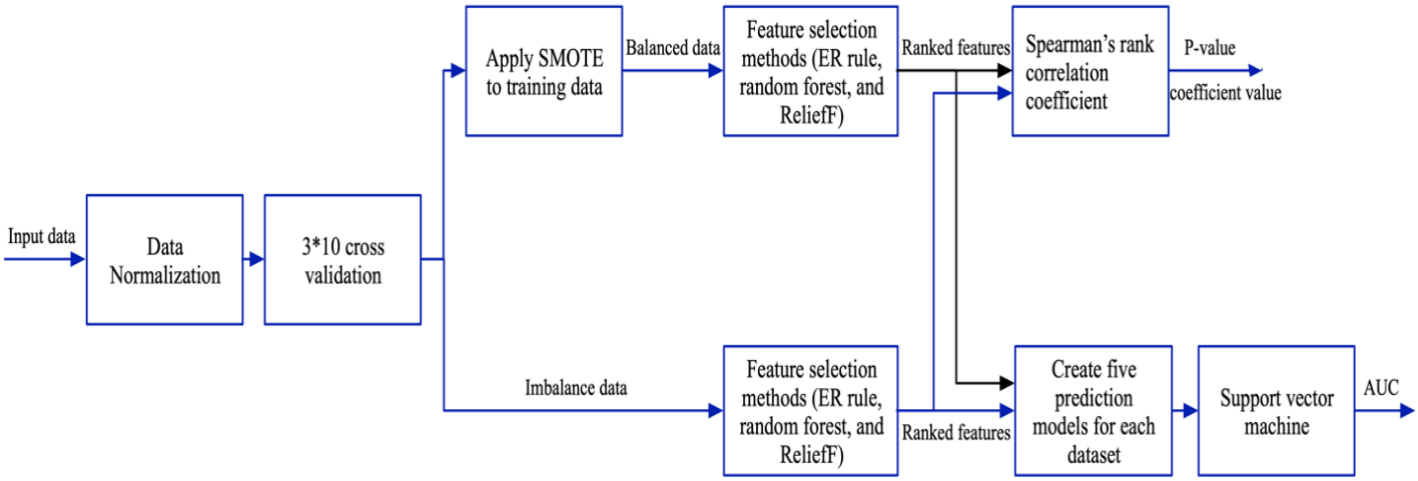


Fig.1. Feature selection process

3.1. Feature selection and normalisation

Feature selection methods allow for analysing different medical and demographic variables and choosing those with the highest prediction ability. Different feature selection methods are constructed differently and help to build different prediction models. New features are then tested and compared with whole feature sets to obtain the best combination for predicting trauma outcomes. There are three widely categorised feature selection techniques: filters, wrappers, and embedded [32]; each has its advantages and disadvantages.

Filter methods rank features based on their relationships with outcomes, but they are not model-oriented. Ranking is calculated either for each individual feature, which refers to univariate filters, or based on the correlation between variables and outcomes or for the whole set of feature correlation (called multivariate filters). Krakovska, et al. [33] categorised filter methods as non-linear feature selection methods, since filters do not assume a linear functional relationship between features and output variables. One of the main drawbacks of filters is that they ignore the interaction between variables and classifiers since the search is carried out in the feature space, not the hypothesis space [2].

Using learning techniques, wrapper techniques compare different combinations of the proposed features and return the most accurate one. Wrappers select features based on their predictive power before starting a predictive model. For example, one of the most common wrapper methods is sequential backward elimination (SBE). It starts with all features and tests their statistical significance in the classification; the least significant feature is excluded for each iteration [5]. This process can be implemented in both linear and non-linear feature selection methods [33]. All wrappers are multivariate [3] and can be categorised to randomised and deterministic wrappers. One of their disadvantages is that they are prone to overfitting and randomised wrappers may consume computational power [2].

Finally, embedded methods, such as the least absolute shrinkage and selection operator (Lasso), select features by learning a model, such as Lasso. They are classifier-dependent selection methods; therefore, the selection is built based on the classifiers' chosen hypotheses [32]. Embedded methods consider the interaction between prediction variables when generating the prediction model [33]. Examples of techniques under each category are presented in Fig. 2

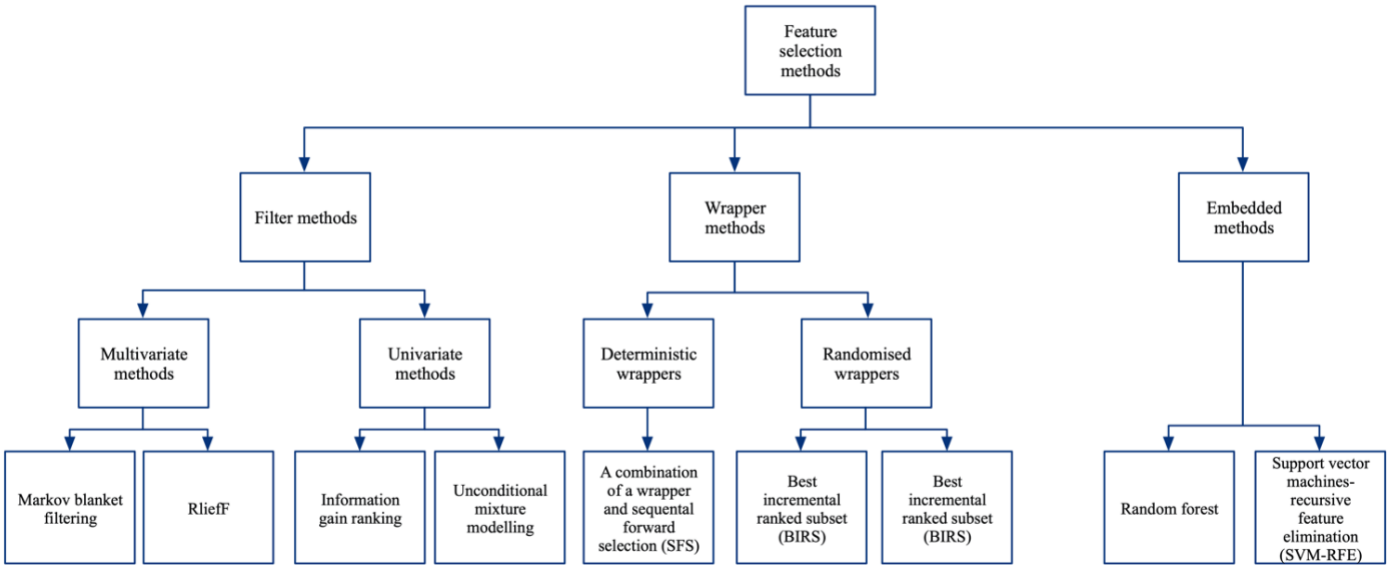


Fig. 2. Feature selection techniques [34].

3.1.1. Evidential reasoning rule

The Bayesian theorem sets the basis for a wide range of different applications and theories that have served many purposes, including examples such as hypotheses testing and model

comparisons using Bayes factors [35], probabilistic reasoning using the Bayesian inference method [36], and reasoning with uncertainty using the Dempster-Shafer (D-S) theory [37]. The D-S theory has been implemented for combining evidence from different sources through the orthogonal sum technique. To overcome the well-known counterintuitive problems inherent in the D-S theory, the ER theory has been introduced. The ER theory can be used to aggregate highly conflicting evidence [38]. The ER theory has been developed over the past 30 years [39,40]. That development covers the establishment of the ER rule which takes the ER algorithm as a special case in which the reliability of evidence is equal to its weight, and the weight of all evidence is normalised [13]. In the joint evidence-state space, evidence reliability refers to the conditional probability for a system state to be true given that the evidence points to the state [41].

The ER rule is capable of handling uncertainty in data that arises due to randomness, ambiguity, and inaccuracy [41]. The first application of the ER rule as a data driven probabilistic inference to solve a classification problem is by Xu, Zheng, Yang, Xu, and Chen [14] and their results proves that it can be simply applied in different domains. The ER rule is also proved to be able to handle missing data in medical data better than common ML algorithms such as logistic regression and artificial neural network [42].

Assuming that with a set of mutually exclusive and collectively exhaustive hypotheses $\Theta = \{\theta_1, \theta_{2,}, \ldots, \theta_N\}$, where Θ represents the frame of discernment, all of the subsets of Θ represented in the power set of Θ are denoted by $P(\Theta)$ or Θ^2as follows:

$$P(\Theta) = 2^{\Theta} = \begin{Bmatrix} \phi, \theta_1, .., \theta_N, \{\theta_1, \theta_2\}, \ldots, \{\theta_1, \theta_N\}, \ldots, \\ \{\theta_1, \theta_{N-1}\}, \ldots, \Theta \end{Bmatrix} \tag{1}$$

A piece of evidence e_j can be shown in a random set format. The element of evidence e_j that points to proposition θ, which can be any subset of Θ, with a probability or belief degree $P_{\theta,j}$ can be presented as follows:

$$e_j = \{(\theta, P_{\theta,j}), \forall\theta \subseteq \Theta, \textstyle\sum_{\theta\subseteq\Theta} P_{\theta,j} = 1\} \tag{2}$$

where $(\theta, P_{\theta,j})$ is the focal element of e_j, if $P_{\theta,j} > 0$. A reliability, detonated by r_j, and a weight detonated by w_j, can be associated with evidence e_j. Evidence weight represents the relative importance of evidence in comparison to other evidence when they need to be combined, while evidence reliability r_j indicates the ability of the evidence to provide the correct assessment for a classification problem [14]. The reliability can be equal to the weight of evidence if pieces of evidence are gathered from the same source [13]. The ER rule combines a piece of evidence with other evidence by combining each evidence's components: belief distribution, weight, and reliability to 'weighted belief distribution with reliability' as follows [13]:

$$m_j = \{(\theta, \widetilde{m}_{\theta,j}), \forall\theta \subseteq \Theta; (P(\Theta)\widetilde{m}_{P(\Theta),j})\} \tag{3}$$

where $\widetilde{m}_{\theta,j}$ indicates the degree of belief for proposition θ acquired from evidence e_j with its weight and reliability, called the weighted belief distribution with reliability (WBDR). $\widetilde{m}_{\theta,j}$ can be defined as follows:

$$\widetilde{m}_{\theta,j} = \begin{cases} 0 & \theta = \phi \\ c_{rw,j} m_{\theta,j} & \theta \subseteq \Theta\theta \neq \phi \\ c_{rw,j}\,(1 - r_j) & \theta = P(\Theta) \end{cases} \tag{4}$$

where $m_{\theta,j} = w_j P_{\theta,j}$ and $c_{rw,j}$ is a normalization factor equal to:

$$c_{rw,j} = \frac{1}{1 + w_j - r_j} \tag{5}$$

Furthermore, the ER rule uses the orthogonal sum to combine two pieces of independent evidence, such as e_0 and e_1, jointly supporting proposition θ, which is denoted by $P_{\theta,e(2)}$. Two pieces of evidence are considered independent when information gathered from the first piece of evidence does not change by knowing the other evidence [13]. The $P_{\theta,e(2)}$ for two pieces of evidence generated from their WBDR is presented in equation (6):

$$P_{\theta,e(2)} = \begin{cases} 0 & \theta = \emptyset \\ \dfrac{\widehat{m}_{\theta,e(2)}}{\sum_{D\subseteq\Theta} \widehat{m}_{D,e(2)}} & \theta \subseteq \Theta, \theta \neq \emptyset \end{cases}$$

$$\widehat{m}_{\theta,e(2)} = \left[(1-r_1)m_{\theta,0} + (1-r_0)\, m_{\theta,1}\right] + \sum_{B\cap C=\theta} m_{B,0} m_{C,1} \;\; \forall\theta \subseteq \Theta \tag{6}$$

The previous recursive combining formula can be used to combine multiple pieces of evidence in any order to obtain the final-result. The weight of each piece of evidence w_j can be represented as importance for that variable in supporting the different proposition θ. An optimisation model can be constructed, and the weight of each variable w_j can be trained to maximise the AUC objective function.

Most of embedded methods for feature selection are based on optimization which is combined with a classification algorithm. A classification algorithm measures features' importance during the classification procedure. Then, a single optimisation model is trained to modify the weights in order to find the best set of weights that result in better prediction accuracy based on one of the classification performance's measures as an objective for optimisation. The MATLAB function 'fmincon' implements the interior-point method for nonlinear optimisation that maximises the AUC function.

The following pseudocode shows the procedure to conduct the ER rule for feature selection:

Input: each data instance, an outcome class value and a vector of features values

Output: a vector W of attributes' weight

1. **for** all training data instance **do**
2. **for** each variable x_j **do**
3. **for** each value i **do**
4. **for** each outcome class θ_i **do**
5. Calculate the likelihood C_{ij}, the likelihood of observing $x_{i,j}$ given that the outcome of the instance is class θ_i;
6. Calculate the degree of belief $P_{\theta ij}$ for the jth piece of evidence by normalising the likelihood over j;
7. **for** each instance **do**
8. **for** each variable **do**

9. Assign the weighted belief with reliability $\widetilde{m}_{\theta,j}$ value for the valriable's value $x_{i,j}$;
10. Combine the weighted belief with reliability from all the variables;
11. Find the probability for each outcome;
12. Calculate the AUC score based on all the instances;

Optimisation procedure

13. Initialise the weight;
14. set the sum of weights W ≔ 1.0;
15. Optimalweight :=W and
16. **While** maximum number of iteration not reached
17. Repeat
18. for each instance do
19. for each variable do
20. Change the weight W in $\widetilde{m}_{\theta,j}$;
21. Combine the weighted belief with reliability from all the variables;
22. Find the probability for each outcome;
23. Calculate the AUC score based on all the instances R;
24. If the AUC is less than oldAUC, then oldAUC=AUC and Optimalweight=W
25. end While;
26. end;

Here i is the ith value of the variable x_j.

3.1.2. ReliefF algorithm

The Relief algorithm is a filter method which was introduced by Kira and Rendell [43] to calculate variables' importance based on how well their values distinguish between instances that are near to each other. Subsequently, ReliefF was introduced as an extension to Relief that can handle multi-class problems, by splitting the problem to multiple 2-class problems, and help to reduce the noise in data [44]. The following pseudocode for the ReliefF algorithm [16]:

Algorithm RefliefF

Input: for each data instance an outcome class value and a vector of features values

Output: a vector W of attributes qualities estimations

1. set all weights $W[A] \coloneqq 0.0$;

2. for i := 1 to n do

3. randomly select an instance R_i ;

4. find k nearest hits H_j;

5. **for** each class $C \neq$ class(R_i) **do**

6. from class C find k nearest misses $\mathrm{M}_j(C)$;

7. **for** A := 1 **to** a **do**

8. $$W[A] := W[A] - \sum_{j=1}^{k} \frac{diff(A, R_i, H_j)}{n \cdot k} + \sum_{C \neq \mathrm{class}(R_i)} \frac{\left[\frac{P(C)}{1 - P(class(R_i))} \sum_{j=1}^{k} diff\left(A\ R_i, M_j(C)\right)\right]}{n \cdot k};$$

9. **end**;

where n represents the number of instances chosen to search for k nearest neighbours; R_i referred to an instance; j refers to the different classes.

For each instance, the original Relief algorithm searches for the two nearest neighbours: one called the nearest hit (H_j), which points to the same class, and the other called the nearest miss (M_j), which results in a different class. The extension in the ReliefF algorithm looks for the k nearest misses $\mathrm{M}_j(C), i = 1 \ldots . k$ for every class C and takes the average to update the estimated $W[A]$ [45].

The weight of each variable can be described as follows: the highest difference between the instances with different classes and the closest instances that have similar outcome classes, the better the variable is at distinguishing between classes and predicting the outcome. The features' importance provided by the algorithm ranges from -1 to 1, with large positive weights assigned to important variables. The negative values in the weight usually associate with irrelevant variables to the outcomes [46].

3.1.3. Random forest algorithm

The RF algorithm builds multiple trees during the classification process based on bootstrapped sample of training data [47]. RF allows for non-linear interaction between variables through using the gini index for data splitting [48]. The RF algorithm is one of the embedded techniques which uses a subset of observations through bootstrapping techniques in each random binary tree.

It monitors the error rate for observations left out of the bootstrap for each tree grown, which is called an out of bag (OOB) error rate [15]. The mean decrease in accuracy is determined during the calculation of the OOB error and indicates the importance of each predictor in the resulting RF model [49]. This assists in the ranking of independent variables. Variable importance is implemented by permuting the values of each variable and for every observation and measuring the increase in the mean squared error following the permutation. Afterwards, for each feature, the increase in the mean square error is averaged over all trees and divided by the standard deviation that is taken over by the trees. The higher the resulted value, the more important the variable is. Although a variable's importance may change for each run, it is stable across multiple runs [50]. The importance range starts from zero (which means that a variable is not important) and positive values for each variable.

3.2. Data normalisation and classes balancing

Each original feature is normalised by removing its mean and dividing its standard deviation to neutralise the effect of different scales across features [51].

The ranking of features' importance in the traditional feature selection process for imbalanced data is affected by the percentage of the imbalance in the classes, so the selection of features is usually not impacted by minor classes [22]. Thus, it is important to consider the imbalance issue when selecting features and measure the impact of the imbalance between the classes on FS. In ML research, different ways of dealing with the imbalanced class issue has been proposed, such as increasing the cost of minority class's classification error [52], resampling data using oversampling or under-sampling, or learning one class [5,53].

Chawla et al. [53] developed a technique for resampling data called SMOTE. This technique over-samples the minority class by taking each minority class sample and creating synthetic examples near that sample, with the number of neighbours depending on the amount of over-sampling. The SMOTE criteria have been adopted to resample the data and make comparisons with the established imbalanced data. However, Blagus and Lusa [54] argued that SMOTE techniques could create correlation between the cases, which would affect the FS results since some FS techniques assume that samples are independent. The results in that paper showed that the technique is effective to handle imbalance classes in low dimensional datasets, which is the case in the current work.

Since most medical data and trauma data are imbalanced, the current work applies the SMOTE technique to balance the classes, and the Spearman ranking test is implemented to measure the difference between the selected features.

3.3. Classification

Classification algorithm

ML offers a range of methods to deal with different modelling issues such as the nonlinear relationships between variables that most medical data encounter. Therefore, they have been used in many prediction models in different fields, including remote sensing-based estimation [55], stock price direction [56], and health and medical diagnosis [27].

The classification process is unified among all the feature selection models using the support vector machine algorithm (SVM). The SVM algorithm have been incorporated widely in the FS literature to test the performance of the techniques [23] and have been implemented in this paper to train different prediction models.

Cross-validation

The cross-validation procedure is a sample partitioning strategy used with ML models. It prevents the over-fitting problem [57] and improves prediction according to bias and/or variance [58]. Ten-fold cross-validation is one of the most common cross-validation methods. It has been used in this paper to train the model and to test and estimate all ML algorithms.

3.4. Spearman rank correlation

Spearman rank correlation test is implemented to measure the association between the means of two groups of features' importance to compare the importance of the imbalanced classes' features and the balanced classes group calculated by the three different feature selection techniques: RF, ReliefF and ER rule. The Spearman correlation coefficient can be computed using the following formula in the case of distinct ranking in each group:

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} \tag{7}$$

where d is the difference between the two ranks of each observation, and n is the number of samples.

3.5. Trauma data

Data is obtained from the trauma audit and research network (TARN) database, which is the largest trauma database in Europe and includes the data of trauma patients from all of the hospitals in England and Wales [1]. All trauma patients, irrespective of age, that arrived alive at any hospital in England or Wales from 1 January 2012 to 31 December 2015 suffering from a traumatic injury and their final outcomes were recorded were included for analysis. This amounts to 177,014 patient records.

The number in the outcome data category of deceased patients was lower than the number in the category of patients who remained alive, which would affect prediction accuracy results. Approximately 92.72% of the trauma patients' outcomes were recorded as alive, while 7.28% were died. Patients' outcomes at 30 days or discharge, whichever came first, were considered as their final outcomes. Fifty variables are included in the feature selection process, comprising a variety of demographic variables and medical tests (Table 2). Missing data are handled using encoding where they were encoded as another value to the existing categories in the variables [59].

3.6. Model features

3.6.1. Demographic features

Patients' ages are categorised into eight groups (0-5.9, 6-10.9, 11-15.9, 16-44.9, 45-54.9, 55-64.9, 65-75.9, 76 and above); another added demographic feature is gender.

3.6.2. Incident features

This category includes features related to the injury circumstances, such as the intention behind the incident, including sports injuries. The mechanism of an injury is another important aspect; most of the included cases are caused by falls and vehicle incidents, which accounts to 89.74% and 88.29% from the survival and non-survival groups, respectively. Vehicle-related variables include a patient's position in vehicle, the protection in vehicle, and whatever they were trapped in vehicle. Variables associated with transfer process measures the cases of readmission, the occurrence of pre-alert to the hospital before patients' arrival, the mode of

arrival, e.g. ambulance or helicopter, the in-reason, the out-reason, such as the availability of critical care bed, and the need for further specialist care, transfer type, year of incident, site ID, and site type.

3.6.1.Medical features

Medical features comprise of medical scores and in-hospital collected data, including pre-hospital and emergency department values for the following vital signs: systolic blood pressure (SBP), respiratory rate (RR), pulse rate (PR), and pupil reactivity in right and left eye. Injury-related variables consist of the use of intubation/ventilation, penetrating or blunt injury, severity for head, severity for face, severity for thorax, severity for abdomen, severity for spine, severity for pelvis, severity for limbs, severity for other parts of body, body region with maximum severity, Charlson comorbidity index, Glasgow coma scale (GCS), and injury severity score (ISS).

In-hospital variables cover the time to computerised scan x-rays (TCT) (hrs), time to operation (hrs), number of operations, total length of stay, total length of stay in intensive care unit (ICU), visiting ED, and the type of wards in ward 1, ward 2, and ward 3.

Table 2. Model features

Feature type		**Feature name**
Demographics		Age, Gender
Incident features		Incident intention, Mechanism of an injury, Patient's position in vehicle, Vehicle protection, Trapped in vehicle, Pre-alert Mode of arrival, in-reason, out-reason, Need for further specialist care, Transfer type, Year of incident, Treatment site ID, Site type.
Medical features	**Medical scores**	Pre-hospital and ED: Systolic blood pressure (SBP), Respiratory rate (RR), Pulse rate (PR), Pupil reactivity in right and left eye; Use of intubation/ventilation, Penetrating or Blunt injury, Severity for head, Severity for face, Severity for thorax, Severity for abdomen, Severity for spine, Severity for pelvis, Severity for limbs, Other severity, Body region with maximum severity, Charlson comorbidity index, GCS, ISS

		Time to TCT, Time to operation, Number of operations, Total length of stay, Total length of stay in ICU, ED admission, Ward 1, Ward 2, Ward 3

4. Results

4.1. Feature importance values

After applying the ER rule, the results show that the optimised weight of the length of stay (LOS) in hospital, which is 0.1368, is the highest of all the analysed variables, as shown in Fig. 3. The GCS (0.0783) and age (0.0753) are the variables with the next highest weights. The body region with the greatest severity, limb injuries, and the number of operations had the least optimised weights at 0.0085, 0.0095 and 0.0096, respectively. After balancing the outcome, age variable becomes the most important variable. Following age, the LOS in hospital and spine injuries are key predictors of outcome.

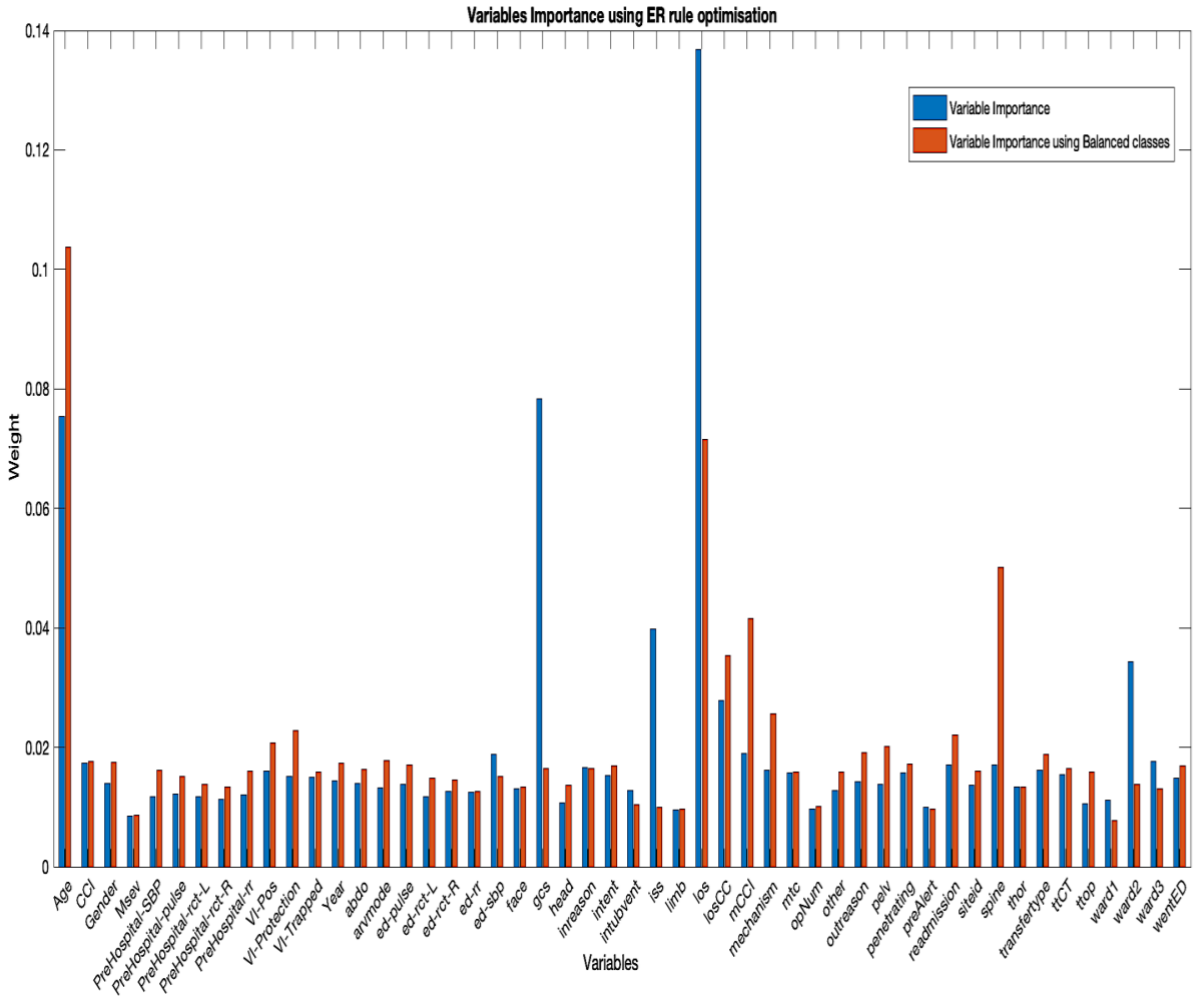


Fig. 3. Variables' importance using the evidential reasoning rule.

Blue: Variables' Importance; Red: Variables' Importance using balanced classes dataset.

Fig. 4 shows the mean importance of each variable generated by the ReliefF algorithm for 30 runs. The GCS, LOS in ICU and the modified Charlson Comorbidity Index place the highest by importance, at 0.042, 0.035 and 0.031, respectively. The least weights are set to the following variables: time to operation (-0.0162), injury intent (-0.0154) and vehicle protection (-0.0136). Over-sampling the minority class affected the importance of the variables. The LOS (0.0576), the speciality of the second ward (0.0310) and the GCS score (0.0407) are the most important variables on average, while vehicle protection, vehicle position and time to operation are the least at -0.0060, -0.0044, and -0.0026, respectively.

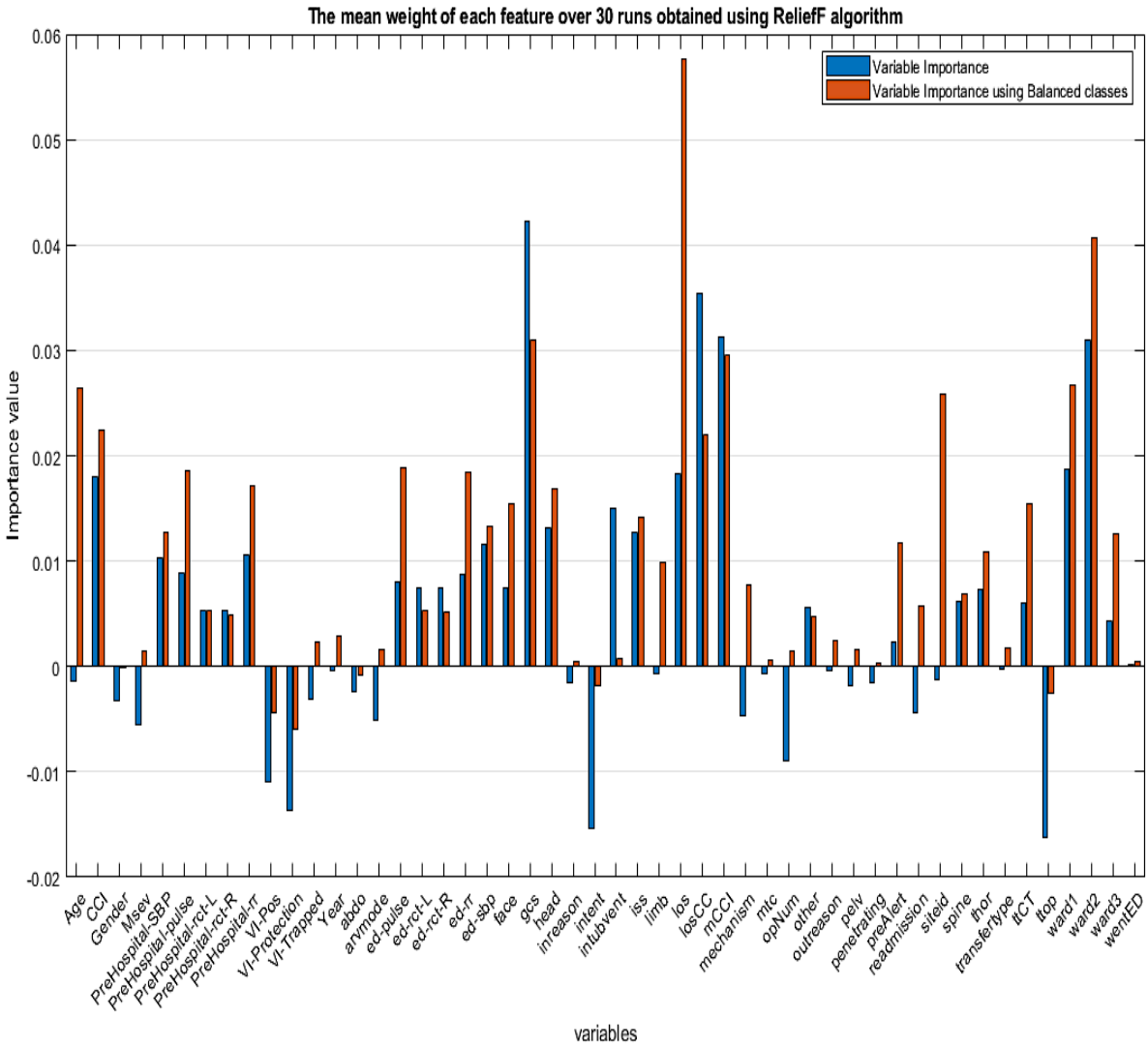


Fig. 4. Variables' importance provided by the ReliefF algorithm.

Blue: Variables' Importance; Red: Variables' Importance using balanced classes dataset.

In this paper, the number of bagged trees is set to 50 trees in the RF algorithm. The three most important variables resulting from the RF algorithm are total length of stay (LOS) at hospital (8.3409), age of the patient (4.5490) and total LOS in ICU (3.6183). Fig. 5 displays the importance of each variable.

On the other hand, the least important variables are the Charlson Comorbidity Index, penetration, and the year of the incident, with values of 0.2572, 0.5918, and 0.6070, respectively. After balancing the classes by applying the SMOTE technique to the training

data, the LOS in hospital remained the most important feature for prediction (5.8667), while the importance of the site of treatment increased from 1.3709 to 5.4661, making it the second most important. The age of patient is third, with a similar value to the RF variables' importance of 4.4799. By contrast, vehicle position, vehicle protection and penetration contributed least to prediction, with values of 0.6549, 0.6995, and 0.9871, respectively.

Add ER rule

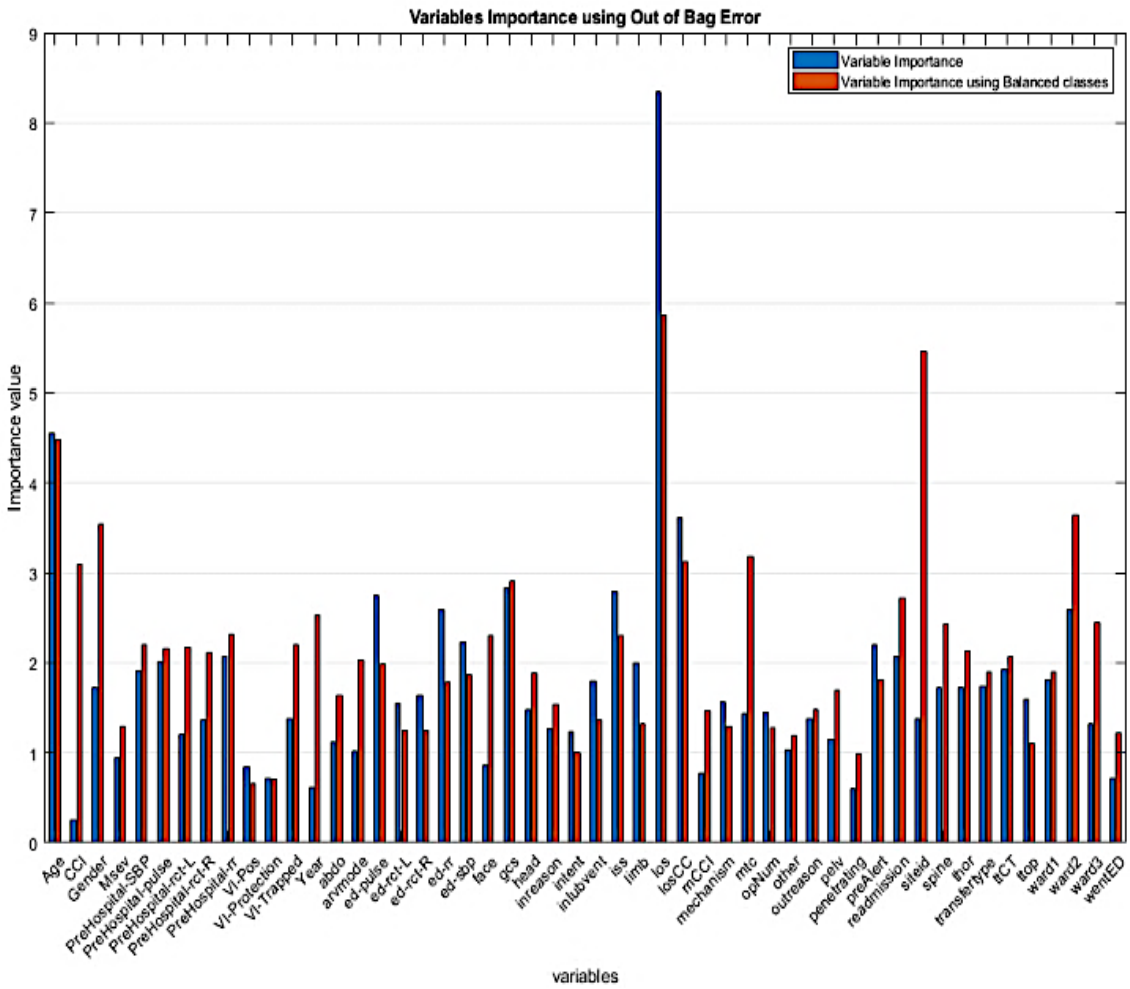


Fig. 5. Variables' importance provided by the RF classifier.

Blue: Variables' Importance; Red: Variables' Importance using balanced classes dataset.

Spearman rank correlation for calculating the impact of imbalanced classes is measured first on the RF feature selection results. The P-value is 0.0024, and the correlation coefficient is 0.4228. These results indicate a moderate correlation and a difference in variables' importance when balancing classes. By contrast, the correlation coefficient of ER feature selection is 0.7991, which indicates a significant correlation between the variables' importance.

Similarly, the differences among the variable importance groups obtained by the ReliefF algorithm are not significant, and the correlation between the two groups is high at 0.7699.

Table 3. The most selected 10 features across the three techniques and their importance values

No.	ER rule		ReliefF		Random forest	
1	**Length of stay**	0.137	**Glasgow coma scale**	0.042	**Length of stay**	8.34
2	**Glasgow coma scale**	0.078	**Length of stay in ICU**	0.035	Age	4.55
3	Age	0.075	mCCI	0.031	**Length of stay in ICU**	3.62
4	**Injury severity score**	0.040	**Ward2**	0.031	**Glasgow coma scale**	2.83
5	**Ward2**	0.034	Ward1	0.019	**Injury severity score**	2.80
6	**Length of stay in ICU**	0.028	**Length of stay**	0.018	Pulse rate in ED	2.74
7	mCCI	0.019	CCI	0.018	**Ward2**	2.59
8	SBP in ED	0.019	intubation/ ventilation	0.015	RR in ED	2.59
9	Ward3	0.018	The severity of the head	0.013	SBP in ED	2.23
10	CCI	0.017	**Injury severity score**	0.013	Pre alert	2.19

Table 3 summarises the key features based on the different FS techniques. Some variables frequently appeared that needed to be highlighted, which are the length of stay, length of stay in ICU, GCS, the type of the second attended ward, and the ISS score. Data shows that patients are more prone to deterioration within the first few days of entering to hospital, with the higher probability during the first day of arrival. Similarly, the length of stay in ICU would indicate a higher probability of survival with the extension of stay in ICU.

4.2. Performances of SVM models built using different number of highest weighted features

4.2.1. Performances of SVM models when features are selected from imbalanced dataset

Next, the classification of the data based on different numbers of highest weighted variables (obtained using imbalanced dataset), either 10, 20, 30, 40 or 50 (all the features available in the dataset), is examined using the SVM algorithm and three times ten-fold cross-validation.

The results for the AUC based on the variables ranked by the weights generated by the ER rule, as presented in Fig. 6, have the least variation between the different models, ranging from 0.895 to 0.919. Each edges of the box indicate the 25th and 75th percentiles, and the middle line indicates the median of the model. The AUC value is 0.895 for the model employing the 10 highest-weighted variables and increases to 0.9 upon adding 10 more variables. Another significant increase occurs between that and the 30-variable model. Finally, the AUC score of the 40-variable model is close to the score for that including all the features.

The performance of the 40 most important variables is similar to that achieved by incorporating the whole set of variables based on the ReliefF algorithm. Fig. 6 shows that when 30 variables are included, the prediction AUC decreased significantly, and it continued to decrease when more variables were removed.

The results show that the highest AUC value is reached by using the 50 most important variables obtained using the OOB error and is equal to 0.918, as shown in Fig. 6. The models, including 20, 30 and 40 variables, have similar AUC values, ranging from 0.9 to 0.905. The least AUC is reached by using the 10 most important variables only.

4.2.2. Performances of SVM models when features are selected from balanced dataset

Fig. 7 illustrates the prediction results of the models ranked based on the balanced dataset. The mean of the AUC values of the models selected by ER rule and random forest has generally decreased after balancing the dataset. The AUC obtained from the dataset ranked using the ER rule optimisation shows a large difference in the range of the AUC value of the model based on the 10 most relevant variables. The mean of the AUC of the prediction model based on the highest 10 variables changed from 0.89 to 0.72. However, the other four models based on the 20, 30, 40, and 50 most relevant variables show less variation. Random forest resulted in a similar high change in the first model's AUC values. Moreover, the other four models' AUC

values ranged from 0.88 to 0.92 and are less than the imbalanced dataset models (0.9 to 0.92). The ReliefF algorithm selected the key features better after balancing the dataset. The AUC values of the first three models (10, 20, and 30 models) are higher than those of the imbalanced data. For example, the mean AUC of the model based on the highest-ranked 20 variables from imbalanced data is 0.84 and it increased to 0.90, after balancing the classes.

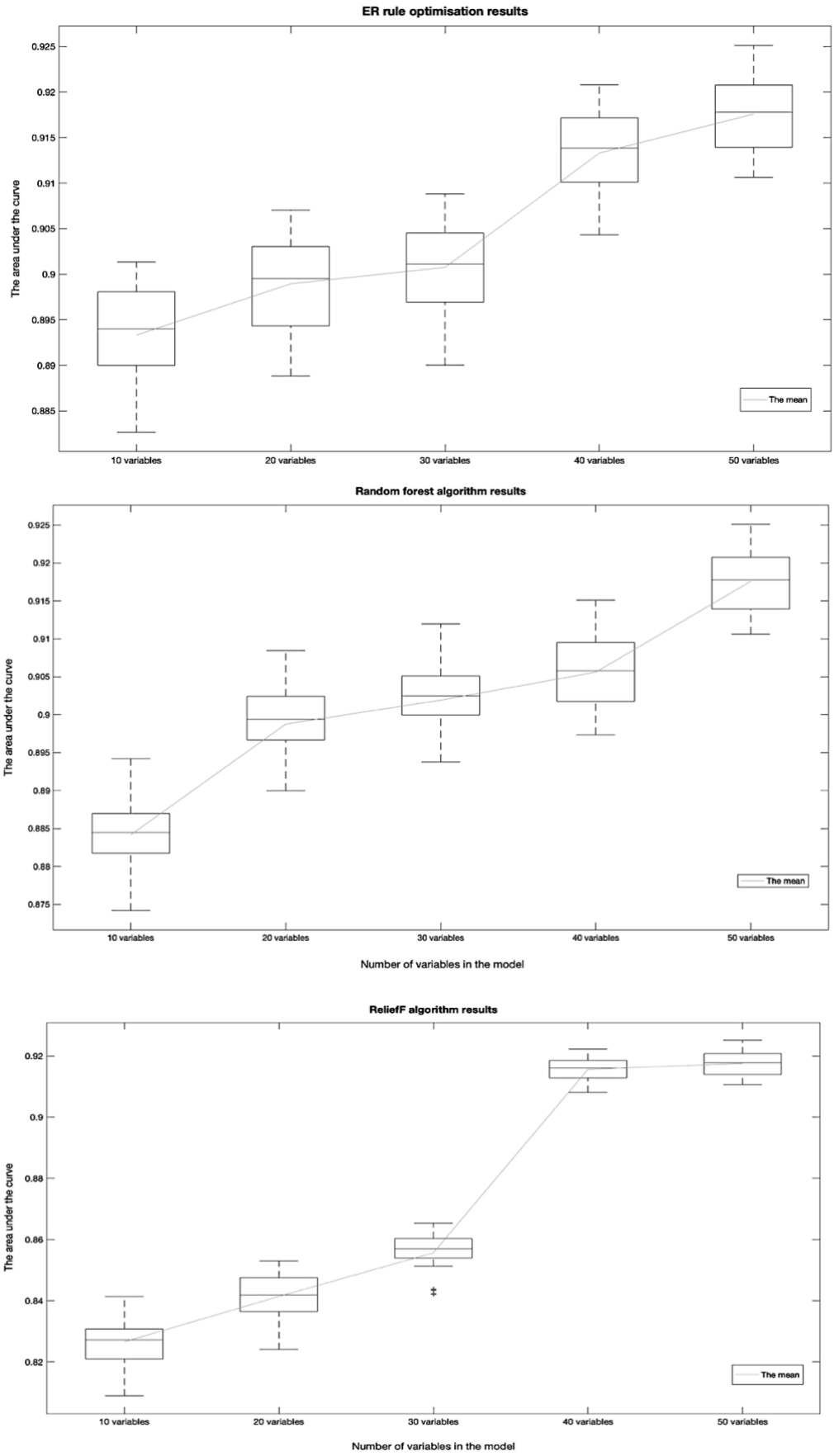
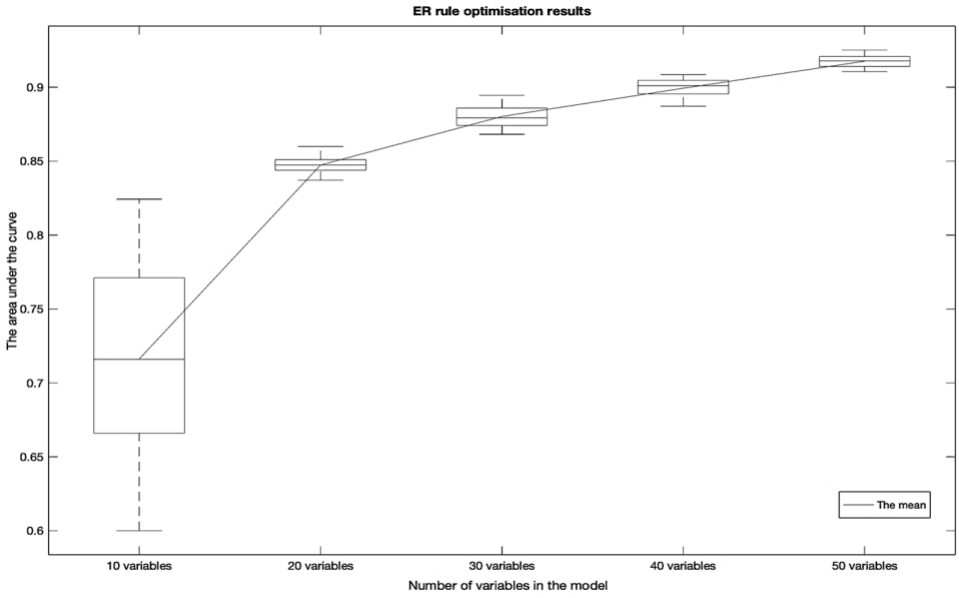
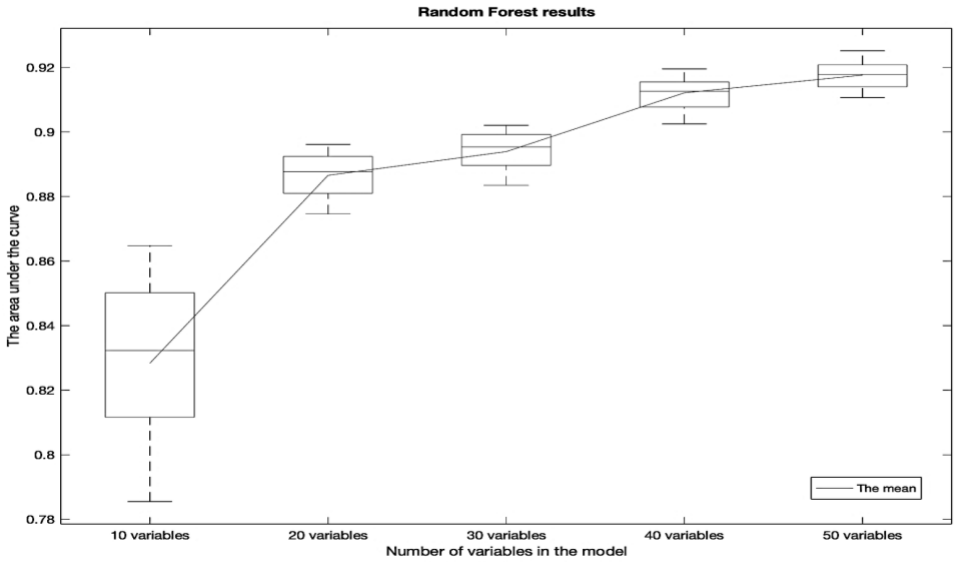


Fig. 6. The mean AUC over 3*10-fold cross-validation on the test folds.


ER rule optimisation results
The area under the curve
0.9
0.85
0.8
0.75
0.7
0.65
0.6
The mean
10 variables
20 variables
30 variables
40 variables
50 variables
Number of variables in the model


Random Forest results
The area under the curve
0.92
0.9
0.88
0.86
0.84
0.82
0.8
0.78
The mean
10 variables
20 variables
30 variables
40 variables
50 variables
Number of variables in the model

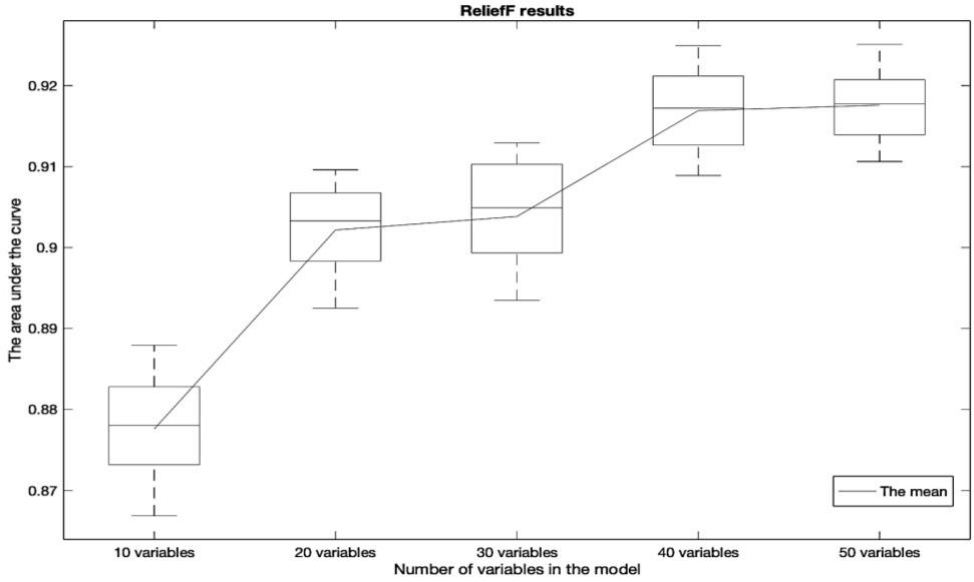

ReliefF results
The area under the curve
0.92
0.91
0.9
0.89
0.88
0.87
The mean
10 variables
20 variables
30 variables
40 variables
50 variables
Number of variables in the model

Fig. 7. The mean AUC over 3*10-fold cross-validation on the test folds after applying SMOTE.

5. Discussion

The present paper is designed to investigate different predictors of outcomes following traumatic injuries and estimate the importance of each variable in a prediction model. It aims to extend the application of the ER rule for classification [14] to find the most important features within the classification process and compare its performance to common feature selection techniques, such as ReliefF and RF algorithms. The ER rule is capable of addressing the uncertainty in data that arises due to randomness, ambiguity, and inaccuracy [41] and it is less sensitive to missing data comparing to other common ML techniques, such as logistic regression [42]. It investigates the impact of the class imbalance and tests the different combinations of features that highly predict the outcome.

Some of the features were selected to be the most relevant for classification by the three FS techniques. The features, including length of stay in hospital and length of stay in ICU, and the medical scores, e.g. ISS and GCS, were ranked consistently as the most predictive features among the features studied in the analysis. The current data showed that patients who stayed longer in the hospital developed more stable medical conditions and the first 48 hours were considered critical for their situations. The length of stay in ICU showed similar effects in the probability of survival for most of the duration, except that most of the patients who spent one day in the ICU has a higher probability of survival. The GCS and the ISS scores are commonly collected from trauma patients and have been implemented widely to measure the severity of patients' injury. They are very common factors in identifying those patients with severe injury who require transfer to a major trauma centre. The higher the value of GCS, the more gradually the survival rate increases. On the contrary, for the ISS, which is the square of the severity score of the most injured part of the body, the higher the score, the lower the probability of survival.

Systolic blood pressure can be measured multiple times during the treatment process of trauma patients. One of the most common stages is the prehospital stage, which is when the medical team first diagnoses a patient, and also when the patient first arrives at the emergency department. Both variables are found to be relevant in all of the implemented feature selection techniques and the SBP variable collected in ED is found to be more significant.

One of the highlighted features by the current results is the type of wards and its relation to patients' outcomes. This empathizes that extended medical care should be given at specific wards such as level 3 ward. That would affect the decision of distributing nursing staff within wards and evaluating the performance of each ward within hospitals. Moreover, this highlights the relationship between the type of ward that patients are transferred to following treatment in emergency department and their outcomes. The type of ward chosen for a patient is either based on the most injured part of the patient, e.g. the coronary care unit, a specific ward for geriatric or paediatric patients, or on the type of surgery required, e.g. plastic surgery. In current trauma data, features included the type of the first three wards the patients admitted to after arriving to the hospital. Approximately 24% of patients who survived had entered orthopaedic ward as the second ward, and around 20% of the non-survival group had been admitted to level 3 ward, which includes patients requiring multi-organ support. A general classification of the type of wards for specific traumatic injuries treatment was presented in a previous report by NICE [60] which divided wards into multidisciplinary, general, multidisciplinary with trauma and specialist wards. The report suggested that a multidisciplinary ward, in addition to a trauma ward, had the best clinical outcomes for traumatic treatment.

Similarly, previous work by Jakaite and Schetinin [21] applied Bayesian inference to a decision tree model to predict survival of traumatic injuries. This manifests that the most important features were age, gender, the severity of head and chest injury, and SBP. By contrast, external severity, injury type (blunt or penetration), heart rate, respiration rate, and the severity of limb contributed the least. A similar conclusion was derived from Hunter [20], where age was shown to be one of the most relevant factors, compared to other variables in the TRISS score, while gender, eye-movement and RR decreased prediction accuracy.

Conversely, vehicle position, vehicle protection, which are associated with traffic accidents, and penetration are commonly found among the least relevant variables to outcomes. This could be due to the small percentage of patients who had traumatic injuries following traffic accidents. The recorded accident injuries in the current data are amount to 21% approximately. Moreover, an approximately similar percentages (3%) of the two classes had penetrating injuries, which explains the reason for that feature to be one of the least selected in most models. Most of the mechanisms of the injury groups in the current dataset are not associated with penetrating injuries, except for two types: injuries caused by shooting and stabbing.

Although, most outcome prediction models are trained on imbalance class data, feature ranking after balancing classes highlights the features that are more associated with the minority outcome which is the risk of mortality in this paper. The results summarised that the RF algorithm is sensitive to the high skew of the outcome variables in selecting key features, in contrast to the ReliefF algorithm. The RF algorithm illustrated a significant difference in the resulted features' weights after balancing classes using the SMOTE technique.

Feature selection algorithms evaluate features by using either an individual evaluation or a subset evaluation [31]. A subset evaluation identifies the best set of features through an iterative evaluation of all possible subsets of features. This paper focuses on feature selection algorithms that use individual selection and rank the features based on their relevance to the outcome. To determine the best set of features, many solutions can be considered. One of these involves finding a threshold that retains a defined percentage of the features [61]. Previous work attempted to apply automated thresholds, such as the Fisher discriminant ratio, or multiple percentages of the total number of features, such as the highest 10% of the ranked features [31]. Four thresholds are defined in the current research, which are the top 20%, 40%, 60%, and 80% of the features based on the features ranking results, which accounts for the top 10, 20, 30, and 40 features, respectively.

SVM algorithm was implemented to calculate features' importance, and the most important first 10, 20, 30, and 40 variables-based models were compared to the whole set of variables on a new dataset. The ReliefF-ranked features showed that there is no significant difference between the 40 and the 50 feature models. However, prediction

AUC decreased significantly when the number of variables was 30 or less. The RF selected models resulted in higher AUC for 10 features and a slight rise after increasing the number of variables to 20, 30, and 40. Lastly, in this current paper, the ER rule for classification was combined with non-linear optimisation to rank the analysed features. The ER rule provided good results for classification and handling missing data [42]. The results illustrated that the ER rule prediction performance of feature ranking has the highest AUC among the selected techniques. This implies that the ER rule weight optimisation can be applied for feature selection, and with specific application in the trauma domain. The findings suggested that there is no huge difference in the AUC score between choosing the 40 highest-weighted variables and the whole set of features. This suggests that, when using fewer variables, similar prediction accuracy could be achieved.

6. Conclusion and Future Work

In this paper, multiple feature selection methods were considered to identify the most important features affected traumatic injury outcomes. Three techniques (ER rule, random forest, and ReliefF) were applied to rank these features and recommend different important variables. Length of stay, LOS in ICU, ISS, GCS and the second ward appeared to be the most important features across the different methods. Moreover, the imbalance in classes in the current datasets affected the variables' importance in two of the implemented techniques, which are the ER rule and ReliefF. It can be concluded that the feature selection results obtained by applying the ER rule construct a more accurate model and improved performance with fewer important variables.

This paper complements the previous research of variable selection in the trauma field by employing empirical cases. The different techniques shared some of the key prediction features, and there is some consistency regarding the least influential factors. That would help in feature reduction in a field that features a continuously increased number of features and cases. Moreover, it could aid practitioners in making better decisions by eliminating the least important features and building more accurate prediction models.

Future research should customise prediction models for each severity level of patients and for each age group, e.g. paediatric patients. Based on the current results, the

the mechanism of injury can have an impact on the variables included in the feature selection process. For example, Lee et al. [12] applied feature selection to identify the most discriminatory factors for paediatric patients and suggested that the selected features varied for different acuity level groups, such that level 1 patients' predictors are shown to be mostly related to the following features: type or number of treatment providers, time until ED bed allocation, and patient complaint. Moreover, future research would include experts' subjective opinions to provide quantitative and qualitative judgements on the selected features. Although previous studies have proven that cross-validation can outperform external validation in cases where the testing data exhibit different features from the training data[62], a limitation of this paper is that the current results are not validated on external datasets. Since medical scores are regarded as essential for patients' outcome classification, future work would calculate and include more trauma scores in the input data, such as the trauma and injury severity score (TRISS) and the revised trauma score (RTS).

Credit authorship contribution statement:

All the authors contributed equally in the paper.

Acknowledgments: The first author is grateful for the research scholarship awarded to her by the Saudi Arabian Government. The other authors are supported by the EU project code named ReMESH under the grant No.: GA No. 823759, and the NSFC-Zhejiang Joint Fund for the Integration of Industrialization and Informatization (U1709215).

References

[1] O. Bouamra, R. Jacques, A. Edwards, D.W. Yates, T. Lawrence, T. Jenks, M. Woodford, F. Lecky, Prediction modelling for trauma using comorbidity and "true" 30-day outcome, Emerg. Med. J. 32 (2015) 933–938. doi:10.1136/emermed-2015-205176.

[2] Y. Saeys, I. Inza, P. Larra??aga, A review of feature selection techniques in bioinformatics, Bioinformatics. 23 (2007) 2507–2517. doi:10.1093/bioinformatics/btm344.

[3] F. Bagherzadeh-Khiabani, A. Ramezankhani, F. Azizi, F. Hadaegh, E.W.

Steyerberg, D. Khalili, A tutorial on variable selection for clinical prediction models: feature selection methods in data mining could improve the results, J. Clin. Epidemiol. 71 (2016) 76–85. doi:10.1016/j.jclinepi.2015.10.002.

[4] N. Esfandiari, M.R. Babavalian, A.M.E. Moghadam, V.K. Tabar, Knowledge discovery in medicine: Current issue and future trend, Expert Syst. Appl. 41 (2014) 4434–4463. doi:10.1016/j.eswa.2014.01.011.

[5] S. Maldonado, R. Weber, F. Famili, Feature selection for high-dimensional class-imbalanced data sets using Support Vector Machines, Inf. Sci. (Ny). 286 (2014) 228–246. doi:10.1016/j.ins.2014.07.015.

[6] H. Asaturyan, E.L. Thomas, J.D. Bell, B. Villarini, A Framework for Automatic Morphological Feature Extraction and Analysis of Abdominal Organs in MRI Volumes, J. Med. Syst. (2019). doi:10.1007/s10916-019-1474-3.

[7] M. Vishwanath, S. Jafarlou, I. Shin, M.M. Lim, N. Dutt, A.M. Rahmani, H. Cao, Investigation of machine learning approaches for traumatic brain injury classification via eeg assessment in mice, Sensors (Switzerland). (2020). doi:10.3390/s20072027.

[8] G. Trivedi, C. Hong, E.R. Dadashzadeh, R.M. Handzel, H. Hochheiser, S. Visweswaran, Identifying incidental findings from radiology reports of trauma patients: An evaluation of automated feature representation methods, Int. J. Med. Inform. (2019). doi:10.1016/j.ijmedinf.2019.05.021.

[9] E. Farago, S. Chinchalkar, D.J. Lizotte, A.L. Trejos, Development of an EMG-based muscle health model for elbow trauma patients, Sensors. (2019). doi:10.3390/s19153309.

[10] R.B. Gelbard, H. Hensman, S. Schobel, V. Khatri, B.M. Tracy, C.J. Dente, T. Buchman, A. Kirk, E. Elster, Random forest modeling can predict infectious complications following trauma laparotomy, in: J. Trauma Acute Care Surg., 2019. doi:10.1097/TA.0000000000002486.

[11] F.C. Bennis, B. Teeuwen, F.A. Zeiler, J.W. Elting, J. van der Naalt, P. Bonizzi, T. Delhaas, M.J. Aries, Improving Prediction of Favourable Outcome After 6 Months in Patients with Severe Traumatic Brain Injury Using Physiological Cerebral Parameters in a Multivariable Logistic Regression Model, Neurocrit. Care. (2020). doi:10.1007/s12028-020-00930-6.

[12] E.K. Lee, F. Yuan, D.A. Hirsh, M.D. Mallory, H.K. Simon, A Clinical Decision Tool for Predicting Patient Care Characteristics: Patients returning within 72

Hours in the Emergency Department, AMIA ... Annu. Symp. Proceedings. AMIA Symp. (2012) 495–504. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3540516/pdf/amia_2012_symp_0495.pdf.

[13] J.B. Yang, D.L. Xu, Evidential reasoning rule for evidence combination, Artif. Intell. 205 (2013) 1–29. doi:10.1016/j.artint.2013.09.003.

[14] X. Xu, J. Zheng, J. bo Yang, D. ling Xu, Y. Chen, Data classification using evidence reasoning rule, Knowledge-Based Syst. 116 (2017) 144–151. doi:10.1016/j.knosys.2016.11.001.

[15] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32. doi:10.1023/A:1010933404324.

[16] M. Robnik-Šikonja, I. Kononenko, Theoretical and Empirical Analysis of ReliefF and RReliefF, Mach. Learn. (2003). doi:10.1023/A:1025667309714.

[17] Z. Zador, M. Sperrin, A.T. King, Predictors of outcome in traumatic brain injury: New insight using receiver operating curve indices and Bayesian network analysis, PLoS One. (2016). doi:10.1371/journal.pone.0158762.

[18] L.C. Wu, C. Kuo, J. Loza, M. Kurt, K. Laksari, L.Z. Yanez, D. Senif, S.C. Anderson, L.E. Miller, J.E. Urban, J.D. Stitzel, D.B. Camarillo, Detection of American Football Head Impacts Using Biomechanical Features and Support Vector Machine Classification, Sci. Rep. (2018). doi:10.1038/s41598-017-17864-3.

[19] K. Chen, F.-Y. Zhou, X.-F. Yuan, Hybrid particle swarm optimization with spiral-shaped mechanism for feature selection, Expert Syst. Appl. 128 (2019) 140–156. doi:https://doi.org/10.1016/j.eswa.2019.03.039.

[20] A. Hunter, L. Kennedy, J. Henry, I. Ferguson, Application of neural networks and sensitivity analysis to improved prediction of trauma survival, Comput. Methods Programs Biomed. (2000). doi:10.1016/S0169-2607(99)00046-2.

[21] L. Jakaite, V. Schetinin, Feature selection for Bayesian evaluation of trauma death risk, in: IFMBE Proc., 2008: pp. 123–126. doi:10.1007/978-3-540-69367-3-33.

[22] L. Yin, Y. Ge, K. Xiao, X. Wang, X. Quan, Feature selection for high-dimensional imbalanced data, Neurocomputing. (2013). doi:10.1016/j.neucom.2012.04.039.

[23] A. Moayedikia, K.-L. Ong, Y.L. Boo, W.G. Yeoh, R. Jensen, Feature selection

for high dimensional imbalanced class data using harmony search, Eng. Appl. Artif. Intell. 57 (2017) 38–49. doi:10.1016/j.engappai.2016.10.008.

[24] J. González-Robledo, F. Martín-González, M. Sánchez-Barba, F. Sánchez-Hernández, M.N. Moreno-García, Multiclassifier Systems for Predicting Neurological Outcome of Patients with Severe Trauma and Polytrauma in Intensive Care Units, J. Med. Syst. (2017). doi:10.1007/s10916-017-0789-1.

[25] N.T. Liu, J. Salinas, Machine Learning for Predicting Outcomes in Trauma, Shock. (2017). doi:10.1097/SHK.0000000000000898.

[26] K. Li, H. Wu, F. Pan, L. Chen, C. Feng, Y. Liu, H. Hui, X. Cai, H. Che, Y. Ma, T. Li, A Machine Learning–Based Model to Predict Acute Traumatic Coagulopathy in Trauma Patients Upon Emergency Hospitalization, Clin. Appl. Thromb. 26 (2020) 1076029619897827. doi:10.1177/1076029619897827.

[27] N. Stylianou, A. Akbarov, E. Kontopantelis, I. Buchan, K.W. Dunn, Mortality risk prediction in burn injury: Comparison of logistic regression with machine learning approaches, Burns. 41 (2015) 925–934. doi:10.1016/j.burns.2015.03.016.

[28] K. Matsuo, H. Aihara, T. Nakai, A. Morishita, Y. Tohma, E. Kohmura, Machine learning to predict in-hospital morbidity and mortality after traumatic brain injury, J. Neurotrauma. (2020). doi:10.1089/neu.2018.6276.

[29] G. Kong, D.L. Xu, J.B. Yang, X. Yin, T. Wang, B. Jiang, Y. Hu, Belief rule-based inference for predicting trauma outcome, Knowledge-Based Syst. 95 (2016) 35–44. doi:10.1016/j.knosys.2015.12.002.

[30] F. Almaghrabi, D.L. Xu, J. Yang, A new machine learning technique for predicting traumatic injuries outcomes based on the vital signs, in: 25th Int. Conf. Autom. Comput., Lancaster, UK, 2019. doi:10.23919/IConAC.2019.8895012.

[31] B. Seijo-Pardo, I. Porto-Díaz, V. Bolón-Canedo, A. Alonso-Betanzos, Ensemble feature selection: Homogeneous and heterogeneous approaches, Knowledge-Based Syst. 118 (2017) 124–139. doi:10.1016/j.knosys.2016.11.017.

[32] Z.M. Hira, D.F. Gillies, A review of feature selection and feature extraction methods applied on microarray data, Adv. Bioinformatics. (2015). doi:10.1155/2015/198363.

[33] O. Krakovska, G. Christie, A. Sixsmith, M. Ester, S. Moreno, Performance comparison of linear and nonlinear feature selection methods for the analysis of large survey datasets, PLoS One. (2019). doi:10.1371/journal.pone.0213584.

[34] F. Almaghrabi, Machine Learning Methods for Predicting Traumatic Injuries Outcomes, The University of Manchester, 2020.

[35] L.J. Norman, A. Tokarev, Spatial attention does not modulate holistic face processing, even when multiple faces are present, Perception. (2014). doi:10.1068/p7848.

[36] J.B. Yang, D.L. Xu, A Study on Generalising Bayesian Inference to Evidential Reasoning, in: Int. Conf. Belief Funct., Springer, Cham., 2014: pp. 180–189.

[37] X. Guo, R. Wang, P. Gao, Y.O. Cui, Knowledge based evidential reasoning analysis of imaging spectrometer data, in: 22nd Asian Conf. Remote Sens., 2001.

[38] J.B. Yang, M.G. Singh, An evidential reasoning approach for multiple-attribute decision\nmaking with uncertainty, IEEE Trans. Syst. Man. Cybern. 24 (1994). doi:10.1109/21.259681.

[39] D.L. Xu, An introduction and survey of the evidential reasoning approach for multiple criteria decision analysis, Ann Oper Res. 195 (2012) 163–187. doi:10.1007/s10479-011-0945-9.

[40] J.B. Yang, P. Sen, General multi-level evaluation process for hybrid MADM with uncertainty, IEEE Trans. Syst. Man Cybern. 24 (1994) 1458–1473. doi:10.1109/21.310529.

[41] J.B. Yang, D.L. Xu, Inferential Modelling and Decision Making with Data, in: 23rd Int. Conf. Autom. Comput., IEEE, 2017: pp. 1–6.

[42] H. Zhu, J.B. Yang, D.L. Xu, C. Xu, Application of Evidential Reasoning rules to identification of asthma control steps in children, in: 2016 22nd Int. Conf. Autom. Comput. ICAC 2016 Tackling New Challenges Autom. Comput., 2016. doi:10.1109/IConAC.2016.7604960.

[43] K. Kira, L.A. Rendell, A practical approach to feature selection, 1992. doi:10.1016/S0031-3203(01)00046-2.

[44] M. Hall, Correlation-based feature selection for machine learning, Diss. Univ. Waikato. (1999). doi:10.1.1.149.3848.

[45] I. Kononenko, E. Šimec, M. Robnik-Šikonja, Overcoming the Myopia of Inductive Learning Algorithms with RELIEFF, Appl. Intell. (1997). doi:10.1023/A:1008280620621.

[46] J. Demšar, B. Zupan, N. Aoki, M.J. Wall, T.H. Granchi, J. Robert Beck, Feature mining and predictive model construction from severe trauma patient's data, Int. J. Med. Inform. (2001). doi:10.1016/S1386-5056(01)00170-8.

[47] S. Adelabu, O. Mutanga, E. Adam, R. Sebego, Spectral discrimination of insect defoliation levels in mopane woodland using hyperspectral data, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. (2014). doi:10.1109/JSTARS.2013.2258329.

[48] A.T. Hudak, E.K. Strand, L.A. Vierling, J.C. Byrne, J.U.H. Eitel, S. Martinuzzi, M.J. Falkowski, Quantifying aboveground forest carbon pools and fluxes from repeat LiDAR surveys, Remote Sens. Environ. (2012). doi:10.1016/j.rse.2012.02.023.

[49] H.R. Pourghasemi, N. Kerle, Random forests and evidential belief function-based landslide susceptibility assessment in Western Mazandaran Province, Iran, Environ. Earth Sci. 75 (2016) 1–17. doi:10.1007/s12665-015-4950-1.

[50] A. Liaw, M. Wiener, Classification and Regression by randomForest, R News. (2002). doi:10.1177/154405910408300516.

[51] Y.R. Shiue, Data-mining-based dynamic dispatching rule selection mechanism for shop floor control systems using a support vector machine approach, Int. J. Prod. Res. (2009). doi:10.1080/00207540701846236.

[52] M. Liu, C. Xu, Y. Luo, C. Xu, Y. Wen, D. Tao, Cost-sensitive feature selection by optimizing f-measures, IEEE Trans. Image Process. (2018). doi:10.1109/TIP.2017.2781298.

[53] N. V. Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, SMOTE: Synthetic minority over-sampling technique, J. Artif. Intell. Res. 16 (2002) 321–357. doi:10.1613/jair.953.

[54] R. Blagus, L. Lusa, SMOTE for high-dimensional class-imbalanced data, BMC Bioinformatics. (2013). doi:10.1186/1471-2105-14-106.

[55] F.E. Fassnacht, ∗ F.H., H. Latifi, C. Berger, J. Hernández, P. Corvalán, B. Koch, Importance of sample size, data type and prediction method for remote sensing-based estimations of aboveground forest biomass, Remote Sens. Environ. 154 (2014) 102–114.

[56] M. Ballings, D. Van Den Poel, N. Hespeels, R. Gryp, Evaluating multiple classifiers for stock price direction prediction, Expert Syst. Appl. 42 (2015) 7046–7056. doi:10.1016/j.eswa.2015.05.013.

[57] and C.-J.L. Chih-Wei Hsu, Chih-Chung Chang, Chih-Wei Hsu, Chih-Chung Chang, C.-J. Lin, and C.-J.L. Chih-Wei Hsu, Chih-Chung Chang, Chih-Wei Hsu, Chih-Chung Chang, C.-J. Lin, A Practical Guide to Support Vector

Classification, BJU Int. (2008). doi:10.1177/02632760022050997.

[58] G. Santafe, I. Inza, J.A. Lozano, Dealing with the evaluation of supervised classification algorithms, Artif. Intell. Rev. 44 (2015) 467–508. doi:10.1007/s10462-015-9433-y.

[59] A.B. Pedersen, E.M. Mikkelsen, D. Cronin-Fenton, N.R. Kristensen, T.M. Pham, L. Pedersen, I. Petersen, Missing data and multiple imputation in clinical epidemiological research, Clin. Epidemiol. (2017). doi:10.2147/CLEP.S129785.

[60] NICE, Major trauma: service delivery (ng40), 2016.

[61] V. Bolón-Canedo, N. Sánchez-Maroño, A. Alonso-Betanzos, Recent advances and emerging challenges of feature selection in the context of big data, Knowledge-Based Syst. (2015). doi:10.1016/j.knosys.2015.05.014.

[62] I.M. Scott, W. Lin, M. Liakata, J.E. Wood, C.P. Vermeer, D. Allaway, J.L. Ward, J. Draper, M.H. Beale, D.I. Corol, J.M. Baker, R.D. King, Merits of random forests emerge in evaluation of chemometric classifiers by external validation, Anal. Chim. Acta. (2013). doi:10.1016/j.aca.2013.09.027.