

Using Machine Learning to Predict if a Profiled Lay Rescuer can Successfully Deliver a Shock using a Public Access Automated External Defibrillator?

Raymond R. Bond¹, Hannah Torney², Peter O'Hare², Laura Davis², Bruno Delafont³, Hannah McReynolds², Anna McLister², Ben McCartney², Rebecca Di Maio², Dewar D. Finlay¹, Daniel Guldenring¹, James McLaughlin¹, David McEneaney⁴

¹Ulster University, Jordanstown, Northern Ireland, United Kingdom

²HeartSine Technologies, Belfast, Northern Ireland, United Kingdom

³Exploristics Ltd., Belfast, Northern Ireland, United Kingdom

⁴Craigavon Area Hospital, Portadown, Northern Ireland, United Kingdom

Abstract

A public access automated external defibrillator (AED) is a device that is intended to be used by lay rescuers in an event where a member of the public experiences a sudden cardiac arrest due to a severe ventricular arrhythmia. Therefore, it is imperative that the human-machine interface of an AED is optimized in terms of its usability and intuitive design. This study involved the recruitment of 362 subjects (lay people) in a shopping mall to undertake the task of using an AED in a simulated environment as facilitated by a 'sensorised' manikin and an AED that was developed by HeartSine Technologies. We found that a large proportion (91.44%) of lay people can successfully use an AED in a simulated emergency scenario to deliver a successful shock. We also found that CPR training did not provide greater likelihood for shock success whilst those with AED training did. Exploratory data analysis and machine learning were used to determine if demographics and other variables are potential predictors for delivering a successful shock using an AED. We found that user demographics and educational attainment were not predictive for AED 'usage' success, which is reassuring since the objective of the medical industry is to develop AEDs that are intuitive to any member of the public.

1. Introduction

Each year, cardiac arrest kills 60,000 people in the United Kingdom alone and many of these events take place outside the clinical setting. [1-2]. The use of an automated external defibrillator (AED) in the first few minutes can increase the probability of survival from less than 5% to over 75%. However, there is a challenge to build AEDs that are 'usable' by all members of the public. The

human-machine interface (or membrane) of a public accessible AED needs to have a high degree of 'usability' and intuitive design since, (1) an AED device should be user-friendly to any lay person regardless of their demographics and educational attainment, and (2) time-to-successful-shock can be crucial and is subject to the usability of the AED. Moreover, given sub-optimal usability is avoidable, a counter-intuitive design should not be a bottle-neck in a life threatening scenario. This study investigated how members of the public would use an AED in a simulated emergency scenario. The study measures the proportion of lay people that can deliver a successful shock using an AED and whether their demographic influences their shock success. In addition, we investigated if a machine learning model could predict whether a profiled bystander is likely to succeed in delivering a shock using an AED. This is important as AEDs are intended to be used by members of the public independent of their demographic or educational status. Therefore, we hypothesized that we would not be able to predict successful 'usage' of AEDs by lay people. We also investigated if other features derived from user interaction with the AED (such as time-to-apply-electrode-pads) could be used to improve the predictive ability of a machine learning model.

2. Methods

We recruited members of the general public to use a public access AED device manufactured by HeartSine Technologies Ltd. Subjects were randomly recruited at a shopping mall to take part in the study. Each subject provided consent and completed a pre-test survey in order to collect their demographics, educational attainment and to determine whether they had prior Cardio Pulmonary Resuscitation (CPR) or AED training. Subsequently, each

subject was exposed to a simulated scenario that included a manikin and an AED device positioned on the floor. The subject was advised to treat this as an emergency situation but were instructed not to call the emergency services. Each subject was simply asked to operate the AED, to provide CPR and to deliver a shock to the simulated patient (manikin). Figure 1 provides an illustration of the simulated scenario.

Data analysis

After data collection, the first stage of the exploratory data analysis involved the use of descriptive statistics and logistic regression to identify those predictive variables that had the potential to contribute to the likelihood for a successful shock delivery. In this study, a successful shock delivery was defined as a scenario in which the user accurately placed the pads in a manner that would ensure that the defibrillating energy passed through the heart, and that they pressed the shock button (or a shock was automatically delivered). Statistics involved the use of Spearman correlation for testing associations between variables, Chi-square testing for determining differences in proportions between demographic sub-groups and Wilcoxon signed rank test for assessing the difference between time based metrics (e.g. time-to-place pads) for those who did and did not successfully deliver a shock. Statistical significance was defined as the alpha level of 5% (or 0.05). A 95% Confidence Interval (CI) was also used for a number of statistics. All data analysis was carried out using the R programming language and R studio.

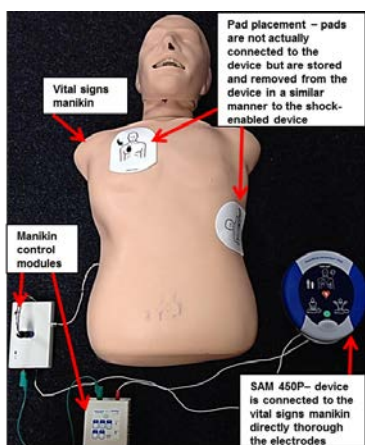


Figure 1. The simulated scenario and experimental setup involving the ‘sensorised’ manikin and the AED.

Data mining

Logistic regression as described in (1) was also used to predict shock success.

$$\text{logit}(p) = \beta_0 + \sum_{i=0}^n \beta_i X_i \quad (1)$$

Where β_0 is the intercept, β is a vector of coefficients (log odds) and X is a vector of values from each independent variable (i.e., age, gender, education, prior CPR Training, prior AED training, acceptable electrode placement, time to place electrodes and time to first shock).

In addition, machine learning was used in this experiment to develop a model that also predicted shock success. This included an optimized C5.0 decision tree, which is a rule induction algorithm that produces a decision tree (using the Carot package for R [3]). C5.0 [4] is an algorithm that produces a decision tree that is human readable as opposed to a black-box. A decision tree can be described as a series of hierarchal rule-based decisions that recursively dichotomise the feature space which eventually leads to a classification (conceptually known as a leaf). However, the hierarchy of these decisions that are based on each of the features is determined using Information-Gain (IG). IG measures how effective each feature is in splitting the data towards its respective classifications (in this case, Class 1: successful shock, Class 2: unsuccessful shock). IG is calculated by subtracting the entropy (a measure of class heterogeneity) in the data before the split from the entropy in the data calculated after the split. IG is defined in (2).

$$\text{InfoGain}(F) = \text{Entropy}(S_1) - \text{Entropy}(S_2) \quad (2)$$

Where F is a given feature (i.e. gender, age, education etc.), $\text{Entropy}(S_1)$ is the entropy before the split and $\text{Entropy}(S_2)$ is the entropy after the split. Entropy is a measure from 0 to 1 (where 0=homogeneity [all data is of the same class] and 1=disorder [or heterogeneous classes]). The greater the IG, the better that feature is for splitting the data into homogenous classes. Entropy is defined in (3).

$$\text{Entropy}(S) = \sum_{i=1}^e -p_i \log_2(p_i) \quad (3)$$

Where S is a segment of data, e is the number of classes and p_i is the percentage of values that are classified into class i .

After the dataset was randomly stratified to avoid order bias, both logistic regression and the decision tree were trained and tested on separate datasets (70% of the data was used for training and 30% for testing). Models were evaluated using typical metrics such as Receiver Operator Characteristic-Area Under the Curve (ROC-AUC), kappa, sensitivity and specificity. Accuracy of the model was also tested against the no-information-rate (alpha=0.05).

3. Results

A total of 362 subjects were recruited (190 females; 172 males; mean age=41.70 [SD=18.87]). A large proportion of subjects (45%) had prior CPR training and a small proportion (9%) had prior AED training. A total of 91.44% (331/362) of subjects used the AED successfully to deliver a shock. However, almost 10% of subjects did not properly position the pads. A total of 8.58% (14/163) of those with prior CPR training failed to deliver a successful shock, whereas 8.54% (17/199) of those without CPR training failed to deliver a shock. Thus, surprisingly there was no significance between these groups ($p=0.98$) indicating that CPR training was not a factor in delivering a successful shock. However, 100% of subjects (33) with prior AED training successfully delivered a shock.

Table 1 shows shows no evidence of association between education level attained by the subjects and their ability to deliver a successful shock. The proportion of unsuccessful shocks amongst those who did not graduate from college level education (9.52%) is not statistically significant ($p=0.32$) from the proportion of unsuccessful shocks from those with college/postgraduate education (6.36%), although the difference in rate is 3%.

Figure 2. illustrates that there is an association between the amount of time taken to apply the pads and delivering a successful shock ($\rho = -0.25$, $p < 0.001$). It can also be seen that there is an association between the time to first shock and delivering a successful shock ($\rho = -0.23$, $p < 0.001$). Hypothesis testing indicated that these time based metrics are significantly different between those who were and were not successful in delivering a shock ($p < 0.001$). As expected, there was a strong correlation between placing pads appropriately and shock success ($\rho = 0.85$, $p < 0.001$).

Whilst descriptive statistics appear to show a number of associations between variables and success, logistic regression using the entire dataset did not provide any statistically significant odds ratios for any of the variables. Logistic regression was also a very poor predictive model (accuracy = 86.58% [CI 79.39%, 92.80%], $p=0.87$). However, as shown in Table 2, the best performing C5.0 decision tree achieved a reasonable accuracy level when using all of the demographics and user interaction variables (i.e. time to place electrode pads) as features (96.33%, CI:90.87, 98.99). Whilst this accuracy is statistically greater than the no-information-rate (Chi-square, $p=0.011$), the confusion matrix (Table 3) indicates that the model misclassified three subjects as successful when they were actually unsuccessful. Nevertheless, as hypothesized, the model performed extremely poorly without the user interaction based features (accuracy=89.91%, CI: 0.8266, 0.9485 [no better than the no-information-rate). Conversely, the model retained a statistically significant model without the

demographic features (accuracy=96.33%, CI: 0.9087, 0.9899).

Table 1. Educational attainment of all subjects.

Education level	Proportion	% of Unsuccessful Shocks
Did not complete high school	8% (30)	6.6% (2)
Some high school	2% (9)	0% (0)
High school	27% (99)	8.1% (8)
High school/some college	31% (114)	12.3% (14)
College	22% (78)	6.4% (5)
Postgraduate	9% (32)	6.2% (2)

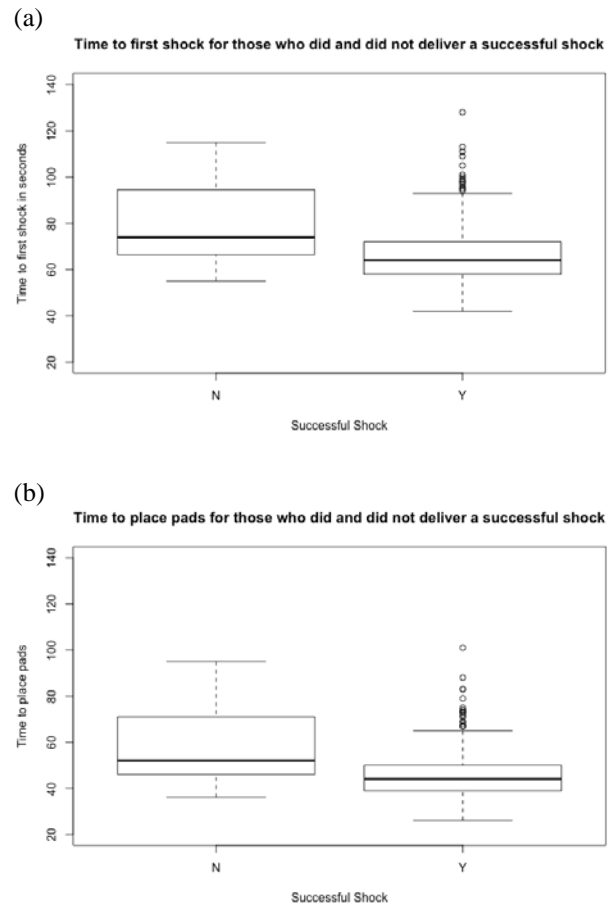


Figure 2. (a) the time to first shock for those who were and were not successful in delivering a shock and (b) the time to place electrode pads for those who were and were not successful in delivering a shock. (Y= Yes and N=No).

Table 2. Evaluation metrics of an optimised C5.0 decision tree for predicting shock success.

Metric	Result
Accuracy	96.33% (95% CI:90.87, 98.99)
Kappa	0.8129
Sensitivity	90.90%
Specificity	96.93%
Pos. Pred. Value	76.92%
Neg. Pred. Value	98.96%
Detection Rate	0.917%
Balanced Accuracy	93.92%
ROC AUC	0.939

Table 3. Confusion matrix of decision tree results.

		Ground truth	
		Unsuccessful	Successful
Prediction	Unsuccessful	10	3
	Successful	1	95

4. Discussion

This paper adds to the body of usability engineering research applied to the design of AEDs [5-6]. Other studies have defined similar research questions but include much smaller sample sizes. For example, Yang et al. [7] recruited a small sample of subjects (n=36) and theorised that trained and untrained users are equally as successful when using an AED. In relation to this paper, we provided evidence that users who are CPR trained do not necessarily perform any better than those who are not trained. However, our research shows that 100% of those with AED training do deliver a successful shock.

This paper also shows that demographical features of a lay rescuer coupled with a decision tree cannot be used to accurately predict if a person is likely to deliver a successful shock. This could however be due to class-imbalance since only 8.6% (31/362) of cases were unsuccessful, which is the main limitation in this study. We could solve this problem by collecting more data or by simulating synthetic data. Nevertheless, since demographics and education attainment in this dataset cannot predict success, this work provides evidence that the AED is user-friendly independent of the profile of the lay rescuer. However, if more data is collected and a predictive model is viable, then such a model can be used to automatically profile ‘near-by’ members of the public and notify the best candidates to access and apply an AED in an emergency scenario. This study does present features with predictive ability that were recorded during user-interaction with the AED (such as time-to-place pads etc.), hence there is an opportunity to investigate the viability of real-time intervention strategies that could be used if an unsuccessful shock is anticipated by the machine.

5. Conclusion

We discovered that there is a small 8.56% (CI: 5.98%, 12.05%) chance that a member of the public will not be able to successfully use an AED. However, we found no evidence to suggest that successful use of AEDs is subject to a person’s demographic or educational attainment. This is very encouraging since public access AEDs are intended to be user-friendly and usable to the general public. This finding is partly confirmed by the fact that a demographic based machine learning model had no predictive ability greater than the no-information-rate. However, it was found that user interaction features such as time-to-first shock and time-to-pad placement etc. do have modest predictive power.

In addition to these findings, we have identified that more work needs to be carried out to improve pad placement and affirmed the lack of retention of CPR training or possibly, the lack of influence of CPR training on the successful usage of AEDs.

References

- [1] Atwood C, Eisenberg MS, Herlitz J, Rea TD. Incidence of EMS-treated out-of-hospital cardiac arrest in Europe. *Resuscitation*. 2005 Oct;67(1):75-80.
- [2] Nolan JP, Soar J, Zideman DA, Biarent D, Bossaert LL, Deakin C, Koster RW, Wyllie J, Böttiger B; European Resuscitation Council Guidelines for Resuscitation 2010 Section 1. Executive summary. *Resuscitation*. 2010 Oct;81(10):1219-76
- [3] Kuhn M. Caret package. *Journal of Statistical Software*. 2008 Feb 29;28(5).
- [4] Lantz B. Machine learning with R. Packt Publishing Ltd; 2013 Oct 25.
- [5] Bond R, O’Hare P, Di Maio R. Usability testing of a novel automated external defibrillator user interface. In proceedings of the International Conference on Bioinformatics and Biomedicine, 2015 (pp 1486-1488). IEEE.
- [6] Torney H, O’Hare P, Davis L, Delafont B, Bond R, McReynolds H, McLister A, McCartney B, Di Maio R, McEneaney D. A Usability Study of a Critical Man-Machine Interface: Can Layperson Responders Perform Optimal Compression Rates When Using a Public Access Defibrillator with Automated Real-Time Feedback During Cardiopulmonary Resuscitation?. *IEEE Transactions on Human-Machine Systems*. (published online and in press)
- [7] Yang Z, Wang J, Wu X, Tang Z, Tang W. 259: Comparison Of Performance Of Aed Between Trained And Untrained Rescuers In A Manikin Study. *Critical Care Medicine*. 2014 Dec 1;42(12):A1423.

Address for correspondence:

Dr Raymond R. Bond, University of Ulster (UUI), Shore Road, Newtownabbey, Co. Antrim, BT370QB, rb.bond@ulster.ac.uk