Machine learning–derived major adverse event prediction of patients undergoing transvenous lead extraction: Using the ESC EHRA EORP European lead extraction ConTRolled ELECTRa registry

Vishal S. Mehta, MBBS,*† Hugh O’Brien, MSc,* Mark K. Elliott, MBBS,*† Nadeev Wijesuriya, MBBS,*† Angelo Auricchio, PhD,+ Salma Ayis, PhD,§ Carina Blomstrom-Lundqvist, PhD,‖ Maria Grazia Bongiorni, MD,¶ Christian Butter, PhD,¶ Jean-Claude Deharo, PhD,** Justin Gould, PhD,*† Charles Kennergren, PhD, FHRS,†† Karl-Heinz Kuck, PhD, FHRS,†† Andrzej Kutarski, PhD,†‖ Christophe Leclercq, PhD,‖ Aldo P. Maggioni, PhD,¶‖ Baldeep S. Sidhu, PhD,** Tom Wong, MD, FHRS,*** Steven Niederer, DPhil,* Christopher A. Rinaldi, MD, FHRS,††† on behalf of the ELECTRa Investigators Group

From the *School of Biomedical Engineering and Imaging Sciences, King’s College London, St Thomas’ Hospital, London, United Kingdom, †Cardiology Department, Guy’s and St Thomas’ Hospital, London, United Kingdom, ‡Division of Cardiology, Fondazione Cardiocentro Ticino, Lugano, Switzerland, §School of Population Health and Environmental Sciences, King’s College London, London, United Kingdom, ‖Department of Medical Science and Cardiology, Uppsala University, Uppsala, Sweden, ¶Cardiology Department, Direttore UO Cardiologia 2 SSN, Azienda Ospedaliero Universitaria Pisana, Pisa, Italy, ‡‡Department of Cardiology, Heart Center Brandenburg in Bernau/Berlin & Brandenburg Medical School, Bernau, Germany, **Department of Cardiology, CHU La Timone, Cardiologie, Service du prof Deharo, Marseille, France, ††Department of Cardiothoracic Surgery, Sahlgrenska University Hospital, Sahlgrenska/SU, Goeteborg, Sweden, ‡‡‡Department of Cardiology, Asklepios Klinik St. Georg, Hamburg, Germany, ‡‡Department of Cardiology, Medical University of Lublin, Lublin, Poland, ‡‡§Department Ordensklinikum Linz Elisabethinen, Linz, Austria, §§Maria Cecilia Hospital, GVM Care and Research, Cotignola, Italy, §§European Society of Cardiology, EORP, Biot, Sophia Antipolis Cedex, France, ***Royal Brompton and Harefield National Health Service Foundation Trust, London, United Kingdom, and †††National Heart and Lung Institute, Imperial College London, London, United Kingdom.

Funding Sources: The study was supported by Boston Scientific, Cook Medical, Medtronic, Spectranetics, and Zoll. The study was also supported by the Wellcome EPSRC Centre for Medical Engineering (WT203148/Z/16/Z). The research was supported by the National Institute for Health Research (NIHR) Biomedical Research Centre based at Guy’s and St Thomas’ NHS Foundation Trust and King’s College London. The views expressed are those of the authors and not necessarily those of the National Health Service, the NIHR, or the Department of Health. Disclosures: Drs Gould, Elliott, and Mehta have received fellowship funding from Abbott (outside the submitted work). Dr Sidhu is funded by the National Institute for Health Research and has received speaker fees from EBR Systems (outside the submitted work). Dr Gould has received project funding from Rosetrees Trust (outside the submitted work). Dr Kennergren has presented on behalf of, advised, and performed studies with Spectranetics/Philips (outside the submitted work). Dr Deharo has received minor honoraria from Philips for lectures and consulting (outside the submitted work). Dr Auricchio is a consultant to Boston Scientific, Backbeat, Biosense Webster, Cardiac, Corvia, Daiichi Sankyo, EBR Systems, Medtronic, Merit, MicroPort CRM, Philips, and V-WAVE; he has received speakers’ fee from Daiichi Sankyo, Boston Scientific, Biosense Webster, Medtronic, MicroPort CRM, and Philips; he has participated in clinical trials sponsored by Boston Scientific, EBR Systems, and Philips; he reports intellectual properties with Boston Scientific, Biosense Webster, and MicroPort CRM (outside the submitted work). Dr Kuck reports grants and personal fees from St. Jude Medical, Biosense Webster, and Medtronic (outside the submitted work). Dr Niederer has received research funding and/or consultation fees from Siemens, Abbott, and Pfizer (outside the submitted work). Dr Rinaldi has received research funding and/or consultation fees from Abbott, Medtronic, Boston Scientific, and MicroPort (outside the submitted work). Dr Blomstrom-Lundqvist reports personal fees from Medtronic, Boston Sci, Bayer, CathPrint, Sanofi, Johnson & Johnson (outside the submitted work). The rest of the authors report no conflicts of interest. Address reprint requests and correspondence: Dr Vishal S. Mehta, School of Biomedical Engineering & Imaging Sciences Rayne Institute, 4th Floor, Lambeth Wing, St. Thomas’ Hospital, Westminster Bridge Road, London SE1 7EH, United Kingdom. E-mail address: vishal.mehta@kcl.ac.uk.

1547-5271/© 2022 Heart Rhythm Society. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/). https://doi.org/10.1016/j.hrthm.2021.12.036
BACKGROUND Transvenous lead extraction (TLE) remains a high-risk procedure.

OBJECTIVE The purpose of this study was to develop a machine learning (ML)-based risk stratification system to predict the risk of major adverse events (MAEs) after TLE. A MAE was defined as procedure-related major complication and procedure-related death.

METHODS We designed and evaluated an ML-based risk stratification system trained using the European Lead Extraction ConTRolled (ELECTRa) registry to predict the risk of MAEs in 3555 patients undergoing TLE and tested this on an independent registry of 1171 patients. ML models were developed, including a self-normalizing neural network (SNN), stepwise logistic regression model (“stepwise model”), support vector machines, and random forest model. These were compared with the ELECTRa Registry Outcome Score (EROS) for MAEs.

RESULTS There were 53 MAEs (1.7%) in the training cohort and 24 (2.4%) in the test cohort. Thirty-two clinically important features were used to train the models. ML techniques were similar to EROS by balanced accuracy (stepwise model: 0.74 vs EROS: 0.70) and superior by area under the curve (support vector machines: 0.764 vs EROS: 0.677). The SNN provided a finite risk for MAE and accurately identified MAE in 14 of 169 “high (>80%) risk” patients (8.3%) and no MAEs in all 198 “low (<20%) risk” patients (100%).

CONCLUSION ML models incrementally improved risk prediction for identifying those at risk of MAEs. The SNN has the additional advantage of providing a personalized finite risk assessment for patients. This may aid patient decision making and allow better preoperative risk assessment and resource allocation.

KEYWORDS Machine learning; Artificial intelligence; Transvenous lead extraction; Outcomes; Risk stratification; ELECTRa registry

Introduction
There is a progressive need for cardiac implantable electronic devices, which has been paralleled by a need for system revision and lead extraction, which are often complex and associated with inherent risk.1 Complications, including death, after transvenous lead extraction (TLE) have been explored in registry and observational analyses. In a previous comprehensive review of TLE including 18,433 patients, the incidence of major complications and procedure-related deaths was 1.6%.2 Overall, the procedural success rate remains high; however, with increasingly complex procedures and patients, the risk of severe complications needs to be considered.

Recognition of this has prompted efforts to risk stratify patients with cardiac implantable electronic devices on the basis of preprocedural assessments. However, the currently available risk scores have their own shortcomings including lack of generalizability and omitting routinely available and potentially powerful predictors of outcomes, while also being variable in their methodology and nature.3–5 In addition, few risk prediction tools use independent test sets and instead use the same data set through bootstrap techniques, which can lead to inherent bias.6

The European Society of Cardiology EURObservational Research Programme (EORP) European Lead Extraction ConTRolled (ELECTRa) registry has provided contemporary outcomes of TLE in European centers, reporting a 1.7% rate of procedure-related major complications including deaths and 0.5% rate of procedure-related deaths.7 This registry is the largest multicenter prospective registry undertaken to date. The recently published ELECTRa Registry Outcome Score (EROS) by Sidhu et al8 used current risk prediction tools and expert consensus to create a risk stratification tool for major adverse event (MAE) as defined by a composite of major complication and procedure-related death by using the ELECTRa database. Click or tap here to enter text.

Recent improvements in machine learning (ML), a field of artificial intelligence (AI), whereby techniques using nonlinear, highly interactive combinations of predictors to uncover novel patterns that may produce predictive performance are a way of developing a personalized assessment of risk, are increasingly being used.7 It has been used in the field of cardiovascular medicine, in areas such as prediction of mortality and morbidity.10 Nonlinear ML is not exclusively superior to linear statistical techniques, and Loring et al11 have demonstrated the noninferiority of linear ML methods such as stepwise logistic regression in comparison to more complex ML techniques when evaluating outcomes of patients with atrial fibrillation in the Global Anticoagulant Registry in the Field-Atrial Fibrillation (GARFIELD-AF) and Feasibility study of an objective randomized blinded investigation of therapeutic ablation versus cardioversion for persistent Atrial Fibrillation (ORBITA-AF) registries.

The EROS tool provided a statistically significant ability to discriminate between those considered at high, medium, and low risk of MAEs; only a limited number of variables were used to develop the risk score from a previously developed risk score by Kancharla et al.4 We hypothesized that ML would predict MAEs more accurately than does the EROS tool. We designed and evaluated an ML-based risk stratification system using the ELECTRa registry to predict the risk of MAEs in patients undergoing TLE and tested this on an independent prospectively collected observational registry of patients undergoing TLE. We compared the performance of linear and nonlinear ML algorithms to the EROS risk stratification tool.

Methods
This study complies with the Declaration of Helsinki and locally appointed ethics committees have approved the research.
Two separate independent data sets were collected. The ELECTRa registry data set was used for training all ML models and was defined as the training data set. The design and results of the ELECTRa study have been previously published. The executive committee, in collaboration with the EURObservational Research Programme, provided the study design, protocol, and scientific leadership of the registry under the responsibility of the European Heart Rhythm Association Scientific Initiatives Committee. Overall, 3510 patients underwent TLE in 73 European centers.

An additional prospectively collected database of all consecutive patients undergoing TLE between October 2000 and November 2019 recorded on a computer database at Guys’ and St Thomas’ NHS Foundation Trust, a high-volume UK TLE center, served as the test data set. A total of 1171 patients underwent TLE in this period. Multiple preprocedural parameters were recorded. The extraction procedure undertaken at this center has been described in detail elsewhere. All potentially duplicated subjects from the training and test data set based on the site of enrollment were excluded from the final analysis. TLE was defined as per the European Heart Rhythm Association and Heart Rhythm Society guidelines in both data sets.

For conventional statistical methods, see “Conventional Statistics” in the Online Supplement.

Feature selection and model development
The intended outcome from the ML algorithms was preprocedural identification of patients at high risk of negative outcomes. The outcome of MAEs, defined as a composite of procedure-related major complication and procedure-related death, was selected as the ground truth to capture this goal. These were defined in both data sets by the consensus statement in the ELECTRa data set, with the same definition used in both the training and test data sets (Online Supplemental Table 1). Each MAE was counted only once per patient (ie, if a major complication and death were recorded for the same patient, this was only counted as 1 MAE). To standardize renal function across data sets, the estimated glomerular filtration rate was estimated using the Modification of Diet in Renal Disease 4-variable equation.

Features were selected on the basis of whether they were available in both data sets without significant gaps. Binary classifiers demonstrating missing data, such as sex, would result in the case being removed from the data set, as no reliable replacement could be computed. Continuous or categorical variables with acceptable levels of missing data were replaced by accepted normal values for the respective database, or a default of false for binary classes where a reliable replacement was feasible. Features such as operator and length of the procedure could correlate with outcomes but would not be a useful required input for a preprocedural tool and so were not included in developing the model.

The sequential backward selection (SBS) algorithm was used to obtain the optimal classifiers using the test data set as there were a large number of available features. The primary objective of SBS is to sequentially remove features from the given feature list, and at each stage of removal, the feature that causes the least performance loss gets removed. SBS aims to reduce the dimensionality of the initial feature subspace from the maximum number of available features 1 at a time, with a minimum reduction in the model performance to improve on computational efficiency and reduce generalization error. A graphical representation of the SBS method is shown in Figures 1A and 1B.

Because of the high class imbalance in both data sets, that is, there were a disproportionately low number of observations classified as MAEs in each data set, a second output of the algorithms was considered appropriate. Using the probabilistic outcomes from the best performing methods, a third “medium risk” class was predicted and made of cases where there was not high certainty in the class definitions by the algorithm. Softmax is a mathematical function that converts a vector of numbers into a vector of probabilities, where the probabilities of each value are proportional to the relative scale of each value in the vector. This function was used to normalize the output of the self-normalizing neural network (SNN) to a probability distribution, ensuring all the predictions sum up to 1. After passing through the Softmax function, a different threshold of under 0.2 was considered a medium-risk case (Figure 2).

Model selection
For model selection, we compared the ML algorithms stepwise logistic regression model ("stepwise model"), support
vector machines (SVMs), and random forest model, although the random forest model produced poor results and was removed from the analysis. The implementations used for these are all available in Scikit-learn 0.22.1. The best of these was then taken and compared with an SNN. This network architecture allows for standard feed-forward networks of increased depth by incorporating normalization into the activation layers to improve stability during training. This was implemented using PyTorch 0.4.1. Training was done using the Adam optimization algorithm. The SNN had 7 layers, each fully connected with scaled exponential linear unit activation functions. Cross-entropy loss was used as the error function, with class weighting used to adjust the class imbalance (0.5 and 40 for low and high risk, respectively). All methods are also compared against the EROS score. The model’s overall performance was compared using (1) balanced accuracy, (2) receiver operating characteristic curve analysis with area under the curve (AUC) calculation, and (3) odds ratio (OR) of having an MAE. Notably, balanced accuracy, rather than total accuracy, was used as the primary performance measure in view of the skewed distribution of outcomes with no MAEs vs those with MAEs: that is, there were a few cases involving MAEs in both data sets. Using accuracy alone would overestimate the ability of the risk stratification method in appropriately identifying the patient risk (Figure 3). Balanced accuracy is the mean accuracy for each class, which can be defined as follows:

$$\text{True positive rate + True negative rate} \over 2$$

## Results

### Baseline characteristics

The final training cohort included 3555 patients, of whom 3122 (81.8%) had the required features without significant missingness with 53 MAEs (1.7%). The test data set included 1171 patients, of whom 998 (85.2%) had the required features without significant missingness with 24 MAEs (2.4%). The baseline characteristics of the total training database of 3555 patients have been described elsewhere. The baseline characteristics of the final patients included in the training and test data sets are described in the Online Supplement (Online Supplemental Table 2 and “Description of Baseline Demographics”).

![Experimental flow diagram for training each classifier and determining risk classes from the classifier output. Both feature selection and training are done on the ELECTRa dataset only. ELECTRa = European Lead Extraction ContRolled; GSTT = Guys’ and St Thomas’ NHS Foundation Trust.](image1.png)

![Explanation of balanced accuracy. If accuracy alone was used as the metric to evaluate the effectiveness of the machine learning algorithm, then even if no MAEs were predicted (left), there would be an accuracy of 76%, which would be the same even if 2 MAEs were correctly predicted (right). MAE = major adverse event.](image2.png)
Leading predictors of MAEs are presented in Figure 4, and the full list of features considered for inclusion and included in the final ML and stepwise models is provided in Online Supplemental Table 3.

ML models and comparison with alternative risk scores

The results on the test data set for the SVM, SNN, and logistic regression are presented in Table 1. Balanced accuracy, as described previously, is displayed as the average accuracy between the 2 classes. The best performing ML model by balanced accuracy was from the logistic regression but with small differences between the 3 methods (logistic regression: 0.74 vs SNN: 0.73 vs SVM: 0.72). The probabilistic output from the SNN that allowed constructing a third class of medium risk from the output of the network as described in Methods is provided in Table 2 and compared against EROS. In this table, we can see that most major complication cases are placed in the medium- or high-risk categories while most of the noncomplication cases end up in the low- or medium-risk categories in both the EROS and ML risk scores. The SNN was similar in appropriately identifying the risk of MAE compared with EROS (high risk vs EROS 3: 8.28% vs 8.38%; medium risk vs EROS 2: 1.63% vs 0.98%). In the low-risk category, the SNN was superior in appropriately identifying those with low MAE risk compared with EROS 1 (190 of 190 cases [100%] vs 186 of 190 cases [97.9%]). Receiver operating characteristic curves demonstrated highest AUCs for SVM (0.764; 95% confidence interval [CI] 0.67–0.84), followed by SNN (0.763; 95% CI 0.70–0.86), logistic regression (0.760; 95% CI 0.67–0.83), and EROS 3 (0.677; 95% CI 0.56–0.78). The SVM was the best performing of the ML algorithms on the basis of AUC results but with nearly similar results between the 3 ML classifiers compared (see Figure 5 and Table 1).

Discussion

The present study incorporating 2 large and independent data sets with >4000 patients demonstrated that ML models provided improved risk stratification capability compared with EROS alone. The main findings are as follows:

### Table 1 Different metrics evaluating the effectiveness of ML models to identify the risk of MAEs

<table>
<thead>
<tr>
<th>Metric*</th>
<th>SVM</th>
<th>SNN</th>
<th>Logistic regression</th>
<th>EROS 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>0.70</td>
<td>0.75</td>
<td>0.83</td>
<td>–</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.73</td>
<td>0.72</td>
<td>0.66</td>
<td>–</td>
</tr>
<tr>
<td>Balanced accuracy</td>
<td>0.72</td>
<td>0.73</td>
<td>0.74</td>
<td>0.70</td>
</tr>
<tr>
<td>AUC (95% CI)</td>
<td>0.764 (0.67–0.84)</td>
<td>0.763 (0.70–0.86)</td>
<td>0.760 (0.67–0.83)</td>
<td>0.677 (0.56–0.78)</td>
</tr>
<tr>
<td>Odds ratio (95% CI)</td>
<td>7.06 (3.36–14.84)</td>
<td>7.54 (3.45–16.54)</td>
<td>7.22 (3.15–16.55)</td>
<td>6.71 (3.35–13.46)</td>
</tr>
</tbody>
</table>

*Sensitivity, specificity, balanced accuracy, area under the curve (AUC) calculations, and odds ratio for identifying major adverse events (MAEs), with 95% confidence intervals (CIs) of machine learning (ML) models (support vector machine [SVM] and self-normalizing neural network [SNN]), logistic regression, and ELECTRa Registry Outcome Score (EROS) III scores.
Table 2  SNN-derived 3-stage risk score compared with the 3 EROS score levels

<table>
<thead>
<tr>
<th>Number of observed events</th>
<th>EROS</th>
<th>EROS 1</th>
<th>EROS 2</th>
<th>EROS 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>No complications</td>
<td>190</td>
<td>611</td>
<td>167</td>
<td></td>
</tr>
<tr>
<td>Major adverse event</td>
<td>4</td>
<td>6</td>
<td>14</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number of observed events</th>
<th>SNN-derived score</th>
<th>Low risk</th>
<th>Medium risk</th>
<th>High risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>No complications</td>
<td>190</td>
<td>615</td>
<td>169</td>
<td></td>
</tr>
<tr>
<td>Major adverse event</td>
<td>0</td>
<td>10</td>
<td>14</td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations as in Table 1.

1. ML models demonstrated incrementally improved effectiveness in identifying those at risk of MAE.
2. The SNN had the added advantage of being able to provide a finite risk of MAE and was effective at identifying those at low (<20%) and high (>80%) risk of MAE.
3. Traditional risk factors for major complications after TLE were similarly identified by the ML models, suggesting that clinical decision making remains paramount.
4. ML was limited by the reliance on binary classifiers, imbalanced data sets, and the depth of the data sets. Larger and richer registry data sets could improve risk assessment.

**Importance of risk stratification**

Risk stratification is essential when planning appropriate settings and resources for planned TLE procedures and appropriate resource allocation. Several risk scores have been proposed for patients undergoing TLE, primarily derived retrospectively from observational studies using a combination of expert consensus and logistic analysis. Bontempi et al developed the “SAFeTY TLE score,” which identified that sum of lead dwell times, anemia, female gender, number of procedures preceding TLE, and age younger than 30 years had a significant influence on the occurrence of major complications during TLE.3 Jache et al developed the “SAFeTY TLE score,” which identified that sum of lead dwell times, anemia, female gender, number of procedures preceding TLE, and age younger than 30 years had a significant influence on the occurrence of major complications during TLE.4 Kancharla et al derived a risk score derived from established data identifying factors associated with an increased risk of major complications, high-risk patients were deemed to have an implantable cardioverter-defibrillator lead >5 years old and a pacemaker lead >10 years old. This risk score was prospectively validated on whether patients should have TLE performed in an operating room vs device laboratory.5

The present study demonstrated that lead dwell time was a primary feature predicting MAE in the SNN and SVM models (Figure 3). This supports current clinical guidance and suggests that clinical judgment remains paramount in appropriate patient risk stratification, and risk scores should not supersede expert consensus.6 In comparison to prior studies, the present study is the first to develop a risk score on a multicenter international registry and validate it on an independent data set that limits selection and treatment bias.

Sidhu et al retrospectively applied a similarly designed risk score to that published by Kancharla et al to the ELECTRa registry in the form of EROS—low (EROS 1), intermediate (EROS 2), and high (EROS 3) risk—with which we have compared the present study. Patients who were considered high (EROS 3) risk for MAE were classified as having a pacemaker lead >15 years and an implantable cardioverter-defibrillator lead >10 years from implantation. Patients classified in EROS 3 were independently associated with procedure-related major complications including deaths in the original validation study (OR 3.333; 95% CI 1.879–5.914; P < .001).7 When used in our test cohort, patients in the EROS 3 risk category demonstrated an OR for MAE (OR 6.71; 95% CI 3.35–13.46) that was inferior to that of the best performing ML method used on the test cohort (SVM: OR 7.54; 95% CI 3.45–16.54), albeit the large CIs suggest that this difference may be limited (Table 1). EROS 3 was inferior in identifying the number of patients who went on to have MAEs compared with high-risk patients identified by the SNN model (EROS 3: 14 of 181 [7.8%] vs SNN: 16 of 217 [7.4%]). All patients were correctly
identified in the low-risk group as having no MAE compared in the ML tool when compared to EROS 1 (low risk) (SNN: 190 of 190 [100%] vs EROS 1: 186 of 190 [97.9%]).

At-risk patients should be considered to have TLE performed in a high-risk environment. Gould et al\textsuperscript{29} identified the high cost of TLE in health care systems, with a mean reimbursement cost per admission being £17,399.09 ± £13,966.49. Click or tap here to enter text. for the National Health Service. A more discriminative risk stratification system has the potential to reduce financial burden on institutions on the basis of need for an operating theater, consumables, surgeon, and perfusionist, which is often associated with TLE procedures undertaken in high-risk settings or when complications arise.\textsuperscript{30}

**Advantages of ML models**

ML has a significant potential to affect outcomes for patients at risk of adverse outcomes, as it allows a vast number of clinical variables associated with adverse events to be considered in increasingly complex networks. Several recent commentaries have identified the “state-of-the-art” potential of ML and AI to identify patterns and potentially solve complex issues in cardiovascular care with a marked rise in cardiac AI/ML studies.\textsuperscript{31} The outputs of ML algorithms can be presented in different ways that may allow the clinician to interpret the results in the context of the patient wishes and expectations to facilitate shared decision making. The potential influence of complex concealed interactions between multiple weaker interactions may be identified. In our study, we identified that supervised learning using nonlinear ML techniques, whereby we provided data along with a corresponding label for what the algorithm is meant to learn, that is, likelihood of MAE after TLE, improved on established clinical risk scores and logistic regression. Our ML model identified some similar characteristics that increased the likelihood of MAE as compared with established risk factors. However, as ML models capture higher-dimensional, nonlinear interactions among features, the independent influence of each variable on the risk of MAE is easier to establish. Our evaluation of ML algorithms was rigorous, including attempting numerous different classifiers within a hyperparameteric space. In this study, we felt that the SNN provided the best overall tool for identifying risk, as it was close in effectiveness in establishing the risk of MAE with respect to balanced accuracy and AUC score while also being able to stratify patients into low, medium, and high risk of MAE by use of the Softmax function. Identifying a neural network as the optimal ML predictor of cardiovascular end points is consistent with previous studies.\textsuperscript{32}

**Future perspective**

With the increasing use of electronic health records, there is an increased opportunity to integrate a risk score that automatically populates a risk calculator that gives a risk score—this could be tiered or given an absolute percentage risk. By using SBS we were able to develop a risk score using a relatively small number of features, which may make the risk model more practicable (Online Supplemental Table 3). It may also allow better health care planning for TLE procedures to minimize the risk of MAE and identify the number of patients who may have an adverse outcome at the explanting center. Precision risk tools can also add nuance to the decision-making process of the heart team regarding whether to proceed with TLE. It may provide an invaluable tool to the patient in their perception of procedural risk and to the clinician in the consent process. This ML algorithm is particularly sensitive to predict low-risk cases, and validation of this model in a prospective study, similar to Kancharla et al,\textsuperscript{4} may be warranted. Click or tap here to enter text. A potential limitation of this would be that the large sample size required to ensure a significant reduction in MAEs is detected. Similar ML techniques can also be used on longer outcome data to predict the risk of mortality.\textsuperscript{33}

We foresee that the role of ML-based risk scores will become increasingly relevant and structured, dense databases combined with state-of-the-art analytics will pave the way for precision cardiovascular medicine. Ultimately, the ability of ML to evaluate latent associations in this study was limited by the depth of the data set (ie, the data sets primarily used binary classifiers and continuous data). To increase the ability of the ML algorithms to detect a rare event such as MAE after TLE would require larger data sets or more data points, such as imaging data.\textsuperscript{34} The depth of data should be considered when future registry studies are designed, and the inclusion of high-resolution imaging should be considered if ML or AI is relevant to the end points being explored.

**Limitations**

While the risk score was applied to an independent data set from an unrelated explanting center, it was applied retrospectively, which may introduce bias. There are inherent differences in the training and test data sets that were not accounted for in the ML model, namely, that the training data set was based on a multicenter multinational registry, in centers of differing volumes of TLE procedures and extraction techniques, whereas the test data set was based on a single, high-volume, prospectively collected data set. There were differences in baseline demographics, lead information, and procedural data recorded in the 2 data sets, resulting in certain features not being used if there was not a reliable match in both data sets. Data mismatch between the test and training data sets can reduce the efficacy of the ML algorithms used.\textsuperscript{35} Certain low-frequency but traditional high-risk features of patients undergoing TLE, such as vegetation size and lead calcification, were not included in the analysis as they were not available in both data sets. There was a difference in rates of MAEs, which resulted in a class imbalance between the 2 data sets; to mitigate this, we used balanced accuracy to ensure that this did not result in bias when reporting the results; however, this can lead to the reduced overall effectiveness of ML techniques. Although our ML algorithm aids risk stratification, the tools and
environment required for the intervention will largely be affected by the center availability and may be affected by additional factors such as familiarity with different tools. While missing data were dealt with in a standardized fashion, the proportion of missing data varied between data sets and occasionally resulted in some data points being removed from the cohorts. To improve the accuracy of the ML models for the purpose of improving discriminative power to detect those at the highest risk of MAE, the data sets available need to be expanded to include more cases or individual patient data set with more data points.

Conclusion
Identification of patients at risk of MAE is essential in ensuring that procedures are performed in the appropriate environment and may aid shared decision making between clinicians and patients in deciding whether to proceed with TLE. Using commonly available clinical variables, we created a neural network–based risk stratification system to predict patients at an increased risk of MAE after TLE and is the first TLE–related risk stratification tool to validate its results on an independent cohort of patients. The neural network–based risk stratification system was comparable to stepwise logistic regression and had marginally improved power to identify those at high and low risk of MAE as compared with the currently available EROS risk tool. The integration of this ML tool into the heart team’s decision-making process may allow better resource allocation, improve outcomes, and aid patient choice. Our study has demonstrated that ML is a feasible method of developing a risk prediction tool in the field of TLE. ML can therefore be used to answer other questions in the field of TLE risk prediction, such as identifying whether certain physiological factors can make seemingly low-risk patients at higher risk of using a powered extraction tool use. These are important considerations when forming the basis of future investigation and registry design.

Acknowledgments
We thank the EURObservational Research Programme (EORP) Oversight Committee and the Registry Executive Committee of the EORP. Data collection was conducted by the EORP department from the European Society of Cardiology by Myriam Glemot as Project Officer, Maryna Andarala as Data Manager. Overall activities were coordinated and supervised by Aldo P. Maggioni as EORP Scientific Coordinator. The TITAN Xp used for this research was donated by the NVIDIA Corporation.

Appendix
Supplementary data
Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.hrtmm.2021.12.036.

References


