CrossMark

Ensemble-Based Risk Scoring with Extreme Learning Machine for Prediction of Adverse Cardiac Events

Nan Liu[1,2,3] · Jeffrey Tadashi Sakamoto[4] · Jiuwen Cao[5] · Zhi Xiong Koh[3] ·
Andrew Fu Wah Ho[6] · Zhiping Lin[7] · Marcus Eng Hock Ong[2,3]

Abstract Accurate prediction of adverse cardiac events for the emergency department (ED) chest pain patients is essential in risk stratification due to the current ambiguity in diagnosing acute coronary syndrome. While most current practices rely on human decision by measuring clinical vital signs, computerized solutions are gaining popularity. We have previously proposed an ensemble-based scoring system (ESS). In this paper, we aim to extend the ESS system using extreme learning machine (ELM), a fast learning algorithm for neural networks. We recruited patients from the ED of Singapore General Hospital, and extracted features such as heart rate variability, 12-lead ECG parameters, and vital signs. We also proposed a novel algorithm called ESS-ELM to predict adverse cardiac events. Different from the original ESS algorithm, ESS-ELM uses the under-sampling technique only in model training. Our proposed method was compared to the original ESS algorithm and several clinical scores in predicting patient outcome. With a cohort of 797 recruited patients, we demonstrated that ESS-ELM outperformed the original ESS algorithm and three established clinical scores, namely HEART, TIMI, and GRACE, in terms of receiver operating characteristic analysis. Furthermore, we have investigated the impact of hidden node number and ensemble size on the predictive performance. ELM has demonstrated the flexibility in its integration with the ESS algorithm. Experiments showed the value of ESS-ELM in prediction of adverse cardiac events. Future works may include the use of new ELM-based learning methods and further validation with a new cohort of patients.

✉ Nan Liu
liu.nan@singhealth.com.sg

1   Health Services Research Centre, Singapore Health Services, Singapore, Singapore

2   Duke-NUS Medical School, National University of Singapore, Singapore, Singapore

3   Department of Emergency Medicine, Singapore General Hospital, Singapore, Singapore

4   School of Medicine, Duke University, Durham, NC, USA

5   Institute of Information and Control, Hangzhou Dianzi University, Zhejiang, China

6   SingHealth Emergency Medicine Residency Program, Singapore, Singapore

7   School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, Singapore

Introduction

Risk stratification is an important element of emergency department (ED) care as it impacts patient acuity, resource allocation, and clinical decision-making. Chest pain patients present unique challenges for risk stratification because of the current ambiguity in diagnosing acute coronary syndrome (ACS). A high volume of ED visits each year are due to chest pain [1], but only approximately 9% are due to ACS [2]. Missing the diagnosis of ACS may have catastrophic clinical consequences, but over-treating chest pain patients may also lead to ED and hospital inefficiency and

adverse clinical outcomes. Therefore, accurately risk stratifying chest pain patients is important for ensuring optimal patient care.

Physicians typically use subjective history and objective data such as vital signs, electrocardiogram (ECG) changes, and troponin in the context of their clinical experience to risk stratify patients. Physiologic measures such as heart rate variability (HRV) have also been employed to risk stratify chest pain patients based on the hypothesis that HRV can measure cardiac stress through dynamic analysis of the autonomic nervous system [3], [4]. With all this available information, it is becoming important to explore clinically meaningful and validated methods to combine and assess patient data beyond clinical gestalt. Risk scores, such as the HEART score, combine key aspects of patient information (for example age, ECG changes, medical history, and troponin readings) through simple, but effective scoring methods [5]. The HEART score is an evidence-based method that has been proposed [5] and validated to risk stratify chest pain patients [6, 7]. Other clinical scores that have been validated on ED chest pain patients include TIMI score [8] and GRACE score [9]. Additionally, studies have employed logistic regression to identify prognostic factors for adverse cardiac outcomes in chest pain patients [3, 4]. However, these techniques are stationary and unadaptable to different patient populations and clinical settings.

In contrast, machine learning (ML) provides a more dynamic process to combine data for the risk stratification of chest pain patients. Liu et al. showed that an ML-based scoring system employing the variables of HRV, vital signs, and ECG changes can accurately predict major adverse cardiac events in chest pain patients [10–12]. In the ensemble-based scoring system (ESS) [10], support vector machine (SVM) was adopted as the individual classifier, and a hybrid sampling (under-sampling + over-sampling) technique was proposed to handle data imbalance and output combination in the decision ensemble. The over-sampling component uses a method named SMOTE [13] to generate artificial data samples. However, this process increases the computational burden. In this paper, we adopt a fast neural network training method, extreme learning machine (ELM) [14] as the individual classifier, and aim to improve predictive performance of the ESS framework while reducing the computational complexity.

In recent years, ELM has attracted numerous attentions and many ELM variants and extensions have been developed. For example, there are pruned ELM algorithms [15, 16], evolutionary ELM [17], sparse Bayesian ELM [18], and ensemble based ELM [19, 20]. ELM has also been extended to handle imbalanced data [21–23]. Furthermore, there are studies combining fuzzy logic with ELM for performance improvement [24, 25]. Other than algorithmic extensions, ELM methods have been widely implemented

in applications such as medical decision making [26–28], image processing [29, 30], bioinformatics [31, 32], and industrial applications [33, 34].

Given the simplicity and flexibility in extension that ELM has, we aim to investigate the feasibility of combining ESS and ELM in a sophisticated process to facilitate predictive modeling and decision making. The reminder of this paper is organized as follows. The "Clinical Setting and Predictive Variables" section presents the study design, data collection, and selected predictive variables for model building. The "Prediction of Adverse Cardiac Events" section describes a novel ensemble-based framework with extreme learning machine for the prediction of major adverse cardiac events. The "Results" section presents the performance evaluation and "Conclusion" section draws the conclusions.

Clinical Setting and Predictive Variables

Study Design

Adult patients with chest pain suggestive of cardiac etiology were prospectively and conveniently recruited from September 2010 to July 2015 at the Emergency Department (ED) of Singapore General Hospital (SGH). We excluded patients who had an obvious non-cardiac etiology of chest pain (e.g., trauma, pneumothorax) as assessed by the primary ED physician. We also excluded patients if their one-lead ECG showed non-sinus rhythm (e.g., arrhythmia, ectopic beats >5%) or artifacts precluding adequate HRV analysis. Singapore Health Services (SingHealth) Centralized Institutional Review Board approved the study and patient consent was waived.

For each patient, demographic information and medical history were retrospectively acquired from hospital's electronic health records (EHR). The first vital signs acquired in the ED or in triage were also obtained from the EHR. The first 12-lead ECG acquired in the ED was read by the primary ED physician as well as a second, independent reviewer for signs of ischemia (e.g., ST segment changes) and other significant abnormalities (e.g., bundle branch block, left ventricular hypertrophy, QTc prolongation, etc.). The first troponin lab taken in the ED was recorded. At SGH, troponin-T is used with an abnormal value defined as >99th percentile of assay (0.03 ng/mL).

Five- to six-minute one-lead ECG signals were recorded with X-Series Monitor (ZOLL Medical Corporation, Chelmsford, MA). ECG signals were then loaded into Kubios version 2.2 (Kuopio, Finland) for HRV analysis. Within the program, software automatically identified QRS complexes. ECG signals were also manually screened to ensure QRS detection accuracy. If necessary, QRS detection could be manually adjusted. The R-R interval time-series

was then created for each ECG and screened for irregular rhythm, ectopic beats, or artifacts. Time domain, frequency domain, and non-linear HRV variables were computed using internal software of the Kubios program described by Niskanen et al. [35].

In this study, we define the primary outcome as a composite outcome of adverse cardiac events within 72-h of presentation to the ED, including revascularization, cardiac arrest, cardiogenic shock, lethal arrhythmia, and mortality. We obtained the outcomes from patient EHR review.

Heart Rate Variability

We used three types of HRV parameters, including time domain, frequency domain, and non-linear variables. Time domain variables are computed using traditional statistical and geometric methods. Average R-R interval (aRR), standard deviation of the R-R time series (sdRR), and the square root of the mean squared differences between R-R intervals (RMSDD) are derived from statistical computations and are meant to depict the overall variability of the R-R time series. RMSDD is most sensitive to vagal influences, but may also be easily skewed by ectopic beats or irregular rhythms. Other time domain measures such as the number of times that the absolute difference between two successive R-R intervals exceeds 50 ms (NN50) and NN50 divided by the total number of R-R intervals (pNN50) can convey beat-to-beat variation as well as the variability of the total R-R time series. The triangular baseline width of a triangle fit into the R-R interval histogram using least squares (TINN) is a geometric based variable again used to convey overall variability.

Frequency domain variables are computed by transforming the R-R time series into the frequency domain using power spectral analyses such as fast Fourier transform (FFT). HRV arises from different systems, such as respiration, baroreceptors, circadian rhythms, and the central nervous system providing feedback to the sinoatrial node through neural circuits. Such negative feedback systems tend to oscillate producing HRV in healthy adults [36]. The autonomic nervous system provides the majority of this input and therefore HRV frequency parameters give dynamic insight into the balance between parasympathetic and sympathetic tone. At the sinoatrial node, parasympathetic nervous system inputs do not require a secondary messenger and therefore have a higher frequency oscillation. Therefore, high-frequency (HF) components (0.15 to 0.40 Hz) represent parasympathetic input, while low-frequency (LF) components (0.04 to 0.15 Hz) represent sympathetic inputs as well as a component of parasympathetic input. Very low frequency (VLF) components (0 to 0.04 Hz) may not be appropriate for analysis of short R-R interval time series [37]. Low-frequency and high-frequency

components are also normalized over total power minus very low frequency power. The ratio of normalized LF to normalized HF can furthermore depict autonomic balance.

Non-linear variables used to analyze HRV include Poincare plot, sample and approximate entropy, and detrended fluctuation analysis (DFA). The Poincare plot is created by plotting an interval against the subsequent interval for a series of R-R intervals. Multiple methods have been described to characterize HRV using this plot [13]. The technique used in our analysis fits an ellipse to the plot shape with the axis positioned along the line of identity (LOI). The standard deviation of points perpendicular to the LOI measure short-term variability (SD1), while the standard deviation of points parallel to the LOI measure long-term variability (SD2). However, these two variables may be more representative of linear characteristics of the R-R interval rather than non-linear [38].

The short length and noise of the R-R time interval series used in our study present challenges for measuring entropy. Two methods have been previously described to estimate the degree of regularity in the R-R time series including approximate entropy (ApEn) and sample entropy (SampEn). Approximate entropy has been described previously [39] and searches for similar epochs within the time-series. However, this measure has been shown to have bias with respect to length of the R-R interval time series as well as difficulty with relative consistency. Therefore, a new measure called sample entropy was created and shown to be more consistent and less dependent on the length of the R-R interval time series [40].

Detrended fluctuation analysis measures the long-range correlation in noisy signals making it ideal for evaluating R-R interval time series. It has been studied in different contexts among cardiovascular physiology and pathology [41–43]. The method, described previously by Penzel et al. [44], finds correlations over different time scales. In our study, correlations were divided into short term (range 4–16 beats) and long term (range 16–64 beats) quantified by the slope of log-log plot and represented by the variables $\alpha 1$ and $\alpha 2$ respectively. Due to the length of R-R time series, $\alpha 1$ may be the more appropriate measurement for our data.

12-Lead ECG and Vital Signs

In this study, we measured 12-lead ECG using Philips PageWriter TC Series device. Some ECG parameters were automatically computed and shown in the device, and the rest parameters were manually calculated by a trained medical practitioner using a continuous hardcopy paper printout of the electrical signals. Within an ECG cardiac cycle, there are P wave, T wave, U wave, and QRS complex. We used the following 12 parameters as candidate variables in predictive modeling: ST segment changes (ST elevation and ST

depression), T wave inversion, Q wave, QRS axis, QT interval correction (QTc), left bundle branch block (BBB), right BBB, intraventricular conduction delay (IVCD), left atrial abnormality (LAA), left ventricular hypertrophy (VH), and right VH. A brief description of these parameters is given as follows.

The ST segment represents the connection between QRS complex and T wave. Typically, ST segment is isoelectric and matched with the baseline. An ST elevation may be defined when the ST segment is abnormally high above the isolectric baseline. The ST elevation is obtained by measuring the vertical elevation between the ECG trace of the ST segment and the baseline, and may correspond to damage or pathological change to the cardiac muscle. A QRS axis can be determined from the QRS complex, where the QRS axis is the net vector of ventricular depolarization. T wave occurs after the QRS complex and represents the repolarization (or recovery) of the ventricles. T wave inversion may be a sign of coronary ischemia or central nervous system disorder, etc. The QT interval is measured from the initial negative deflection of Q wave, to the end of T wave. A prolonged QT interval may indicate ventricular tachyarrhythmias and sudden death. The QT interval varies with heart rate, and for clinical relevance requires a correction for this, giving the QT interval correction (QTc).

Presence of a left bundle branch block (LBBB) is a cardiac conduction abnormality. In a LBBB, left ventricle contraction is later than the right ventricle due to the delay of activation of the left ventricle. In a RBBB, the right ventricle is not directly activated by impulses traveling through the right bundle branch. IVCD could be determined from a QRS duration widening, where by a process of elimination, the QRS duration widening is caused by an IVCD if the manifestation is not caused by a LBBB or a RBBB. IVCD may correspond to a myocardial infarction, a cardiomyopathy with ventricular fibrosis, or a chamber enlargement. Atrial abnormalities or atrial enlargements, atrial dilatations or atrial hypertrophy may also be detected in an ECG.

In addition to HRV and 12-lead ECG parameters, we chose eight clinical vital signs in predictive modeling. We used Propaq CS (Welch Allyn, Skaneateles Falls, NY, USA) Vital Signs Monitor to measure heart rate, systolic, and diastolic blood pressure. While the patients were presented at the ED, we recorded respiratory rate, GCS, and temperature. Additional vital signs included pain score and oxygen saturation (SpO_2). Furthermore, we collected patients' medical history and relevant information.

Prediction of Adverse Cardiac Events

The original ensemble-based scoring system (ESS) [10] was proposed to risk stratify chest pain patients in the emergency department, where support vector machine (SVM) was implemented as the individual classifier. There are rooms for further improvement: First, a combined use of both under-sampling and over-sampling techniques increases computational load; second, as reported in the literature, ELM-based algorithms outperform SVM in various applications [45].

In this section, we present a brief introduction to the basic ELM algorithm, and then show the detailed descriptions of the proposed ESS-ELM algorithm. ESS-ELM is more than a simple replacement of SVM with ELM, but an integration of unique ELM features into the ESS architecture. We will elaborate the proposed algorithm in "Ensemble-Based Scoring System with ELM (ESS-ELM)" section.

Extreme Learning Machine

ELM [46] was proposed as a learning algorithm for single-layer feed-forward network (SLFN) where the weights and biases for hidden nodes were randomly selected and the output weights are determined with least square solution. In a training set L with N samples,

$$L = \{(\mathbf{x}_j, \mathbf{t}_j) | \mathbf{x}_j \in \mathbf{R}^p, \mathbf{t}_j \in \mathbf{R}^q, j = 1, 2, ..., N\} \quad (1)$$

$\mathbf{x}_j$ is the input feature vector with p components and $\mathbf{t}_j$ is a q-dimensional target vector. Assume $g(x)$ is the activation function for hidden nodes and $\mathbf{w}_i$ is the weight vector that connects input neurons and the ith hidden node, we can define an SLFN with $\tilde{N}$ hidden nodes as follows,

$$f_{\tilde{N}}(\mathbf{x}_j) = \sum_{i=1}^{\tilde{N}} \beta_i g(\mathbf{w}_i \cdot \mathbf{x}_j + b_i) = \mathbf{t}_j, \quad j = 1, 2, ..., N \quad (2)$$

Equations 2 can be further written as

$$\mathbf{H}\hat{\beta} = \mathbf{T} \quad (3)$$

where $\mathbf{H}(\mathbf{w}_1, ..., \mathbf{w}_{\tilde{N}}, b_1, ..., b_{\tilde{N}}, \mathbf{x}_1, ..., \mathbf{x}_N)$ is the output matrix, and $h_{ji} = g(\mathbf{w}_i \cdot \mathbf{x}_j + b_i)$ is the output of ith hidden neuron with respect to $\mathbf{x}_j$. Furthermore, the output weight matrix is $\hat{\beta} = [\beta_1, ..., \beta_{\tilde{N}}]^T$ and the target matrix is $\mathbf{T} = [\mathbf{t}_1, ..., \mathbf{t}_N]^T$. As proposed in Huang et al. [46], parameters $\mathbf{w}_i$ and b_i are randomly assigned so that the output weights can be estimated as $\beta = \mathbf{H}^\dagger \mathbf{T}$. In ELM algorithm, Moore-Penrose generalized inverse [47] is implemented to convert $\mathbf{H}$ into $\mathbf{H}^\dagger$.

Ensemble-Based Scoring System with ELM (ESS-ELM)

As described in [10], ESS was developed for risk stratification of ED chest pain patients, which was specifically designed to handle imbalanced data where samples (i.e., patients) with normal outcome are much more than those samples with abnormal outcome (in our study, a

Fig. 1 The architecture of the proposed ESS-ELM algorithm
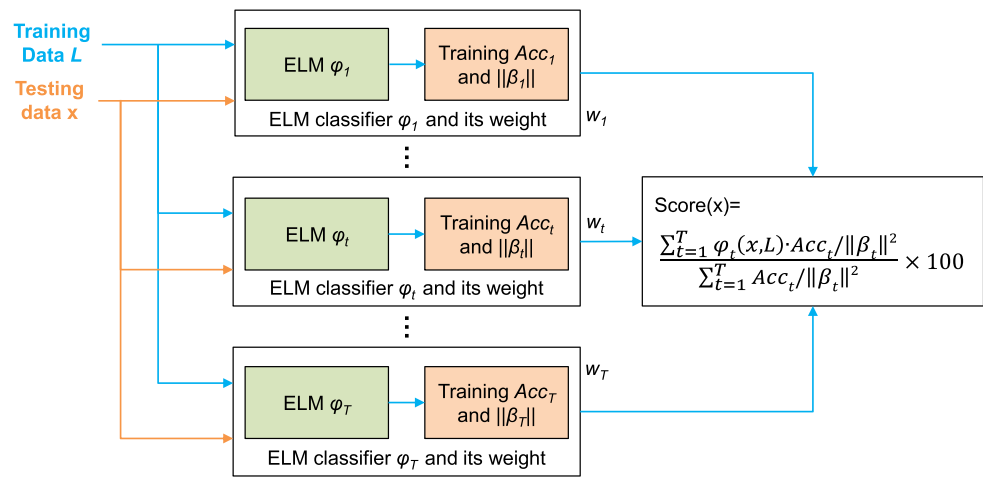
composite adverse cardiac events within 72-h). Since conventional machine learning algorithms cannot handle imbalanced data well [48], ESS algorithm [10] was created by adopting a hybrid sampling technique (under-sampling + over-sampling). The main mechanism behind the ESS algorithm is using ensemble learning to combine individual classifiers for reliable decision making. In this paper, we will integrate ELM into the ESS framework while removing the over-sampling component to reduce algorithmic complexity.

The structure of the proposed ESS-ELM algorithm is depicted in Fig. 1. In this ensemble architecture, each ensemble classifier is defined as φ_t and there are a total of T individual classifiers. A weight is assigned to each classifier φ_t to indicate its significance in the decision ensemble.

As described in the original ESS algorithm, the weight w_t is determined from over-sampled data that are generated using the SMOTE technique [13]. In this new ESS-ELM algorithm, we propose to derive the weights directly from the ELM learning process, which will avoid using over-sampling during model training. Due to the nature of data imbalance in our dataset (146 out of 797 patients met the outcome), we adopted the under-sampling structure in creating the decision ensemble.

Given a training dataset $L = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_K]$ where $\mathbf{x}$ is a feature vector (HRV parameters, 12-lead ECG parameters, and vital signs) representing a patient, we start the analysis with normalizing [49] the original values into the interval of [-1, 1]. We assume that in the training set there are one minority set P (patients with positive

---

Algorithm: ESS-ELM

Inputs: Training set L consisting of a set of minority class samples P and a set of majority class samples N, the number of ensemble classifiers T

Output: Predicted risk score Score($\mathbf{x}$) for the testing sample $\mathbf{x}$.

Normalization: Transform both training data L and testing data $\mathbf{x}$ into interval $[-1, 1]$ with min-max normalization.

For $t = 1, 2, ..., T$

1. Create a balanced dataset S_t by combining P and N_t that is randomly sampled from N where P and N_t have the same number of samples.
2. Build a classification model φ_t with S_t, generate $\|\beta_t\|^2$ and training accuracy Acc_t, and assign $\text{Acc}_t/\|\beta_t\|^2$ to w_t.
3. Build a classification model with S_t and apply the trained model to the testing sample $\mathbf{x}$ to receive prediction output $\varphi_t(\mathbf{x}, S_t)$ that is either 0 or 1.

End For

Prediction: Calculate risk score for $\mathbf{x}$ with

- $\text{Score}(\mathbf{x}) = \left(\sum_{t=1}^{T} \varphi_t(\mathbf{x}, S_t) \cdot w_t \Big/ \sum_{t=1}^{T} w_t \right) \times 100$

where the weight w_t is defined as $\text{Acc}_t/\|\beta_t\|^2$

Fig. 2 Descriptions of the proposed ESS-ELM algorithm

Table 1 Vital signs, heart rate variability (HRV) variables, electrocardiogram (ECG) variables, and troponin variable of patients in the study population

	Patients without outcomes (n = 651)	Patients with outcomes (n = 146)	p value
HRV predictors, mean (SD)			
- aRR (s)	831.4 (171.9)	826.3 (161.9)	0.743
- sdRR (s)	32.3 (23.1)	32.8 (21.6)	0.805
- RMSSD (s)	22.2 (21.9)	19.5 (14.8)	0.163
- NN50 (count)	18.5 (36.7)	13.6 (25.4)	0.057
- pNN50 (%)	5.7 (11.8)	4.3 (8.7)	0.111
- TINN	117.6 (77.9)	118.0 (78.0)	0.950
- LF power (ms^2)	332.0 (791.9)	365.1 (697.5)	0.641
- HF power (ms^2)	256.1 (953.2)	157.8 (293.0)	0.218
- LF power norm (n.u.)	60.8 (20.3)	66.3 (19.9)	0.003
- HF power norm (n.u.)	39.0 (20.1)	33.5 (19.7)	0.003
- LF/HF	2.6 (2.8)	3.3 (3.3)	0.017
Nonlinear HRV predictors, mean (SD)			
- Poincare plot SD1 (ms)	15.7 (15.5)	13.8 (10.5)	0.163
- Poincare plot SD2 (ms)	42.3 (29.6)	43.8 (29.6)	0.586
- Approximate entropy	1.08 (0.15)	1.05 (0.15)	0.005
- Sample Entropy	1.38 (0.35)	1.28 (0.36)	0.002
- DFA, $\alpha1$	1.08 (0.34)	1.18 (0.33)	0.001
- DFA, $\alpha2$	1.03 (0.22)	1.07 (0.21)	0.094
12-lead ECG predictors, n (%)			
- ST elevation	30 (4.6)	36 (24.7)	0.001
- ST depression	52 (8.0)	40 (27.4)	0.001
- T wave inversion	164 (25.2)	45 (30.8)	0.162
- Q wave	57 (8.8)	29 (19.9)	0.001
Vital sign predictors, mean (SD)			
- Pulse rate (bpm)	78.4 (17.4)	78.6 (15.6)	0.918
- Respiratory rate (bpm)	17.8 (2.8)	17.6 (1.5)	0.357
- Systolic BP (mmHg)	142.4 (28.6)	143.2 (29.1)	0.761
- Diastolic BP (mmHg)	76.7 (14.6)	79.2 (14.5)	0.058
- Pain score (0–10)	2.4 (2.6)	3.3 (3.2)	0.002

outcome) and one majority set N (patients with negative outcome). We apply the under-sampling technique to randomly select a subset N_t from N. The number of samples in subset N_t is the same as that in dataset P. In such a way, we are able to create T balanced dataset S_t $(P + N_t)$ for building a decision ensemble.

In the training dataset L, we have K samples $(\mathbf{x}_k, y_k)$ where y_k is C_0 or C_1, with C_1 indicating patient $\mathbf{x}_k$ met the outcome, i.e., a composite outcome of serious adverse cardiac events within 72 h. Assume that we have a testing sample $\mathbf{x}$, we aim to predict its label y using an ensemble of single classifiers $\varphi(\mathbf{x}, L)$ with training set L. As depicted in Fig. 1, we will build total T independent classifiers and combine them to make a final decision. Also, as previously mentioned, the weights are derived from ELM learning process and they should represent the significance of corresponding individual ELM classifiers. Referring to several

ELM literature [19, 46], $\|\beta\|$, the norm of output matrix weights, is closely associated with ELM generalization performance. That is, smaller $\|\beta\|$ leads to better generalization ability. This property is an indicator on how important an ELM classifier is, so we consider $\|\beta\|$ as one key component in creating the decision ensemble. To achieve a trade-off between training accuracy and generalization performance, we define the weight w_t for the t-th ELM classifier as $\text{Acc}_t / \|\beta_t\|^2$ where Acc_t is the training accuracy. The predicted outcome y_t of classifier $\varphi_t(\mathbf{x}, L)$ is either 0 or 1. A risk score for testing sample $\mathbf{x}$ is derived using the following equation,

$$\text{Score}(\mathbf{x}) = \frac{\sum_{t=1}^{T} \varphi_t(\mathbf{x}, L) \cdot w_t}{\sum_{t=1}^{T} w_t} \times 100$$
$$= \frac{\sum_{t=1}^{T} \varphi_t(\mathbf{x}, L) \cdot \text{Acc}_t / \|\beta_t\|^2}{\sum_{t=1}^{T} \text{Acc}_t / \|\beta_t\|^2} \times 100 \qquad (4)$$
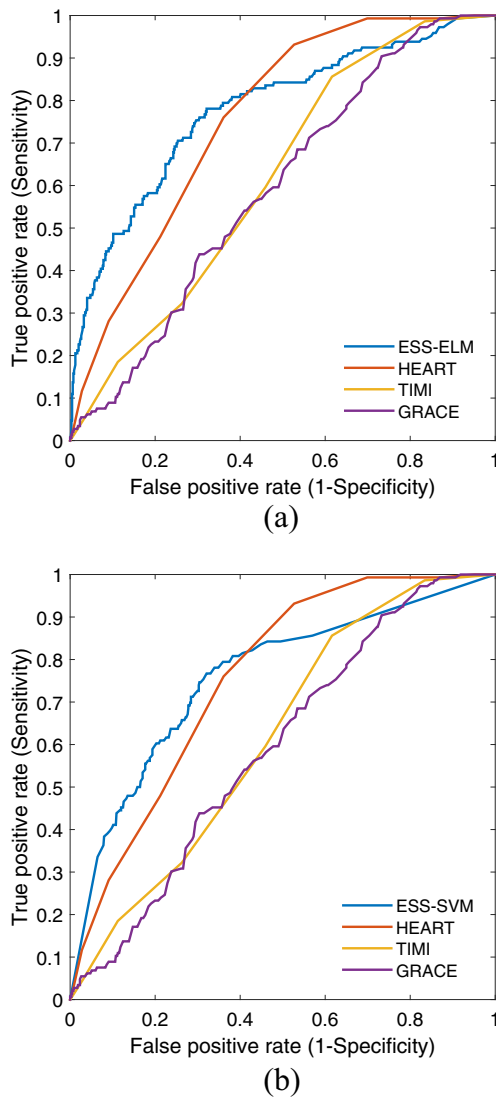
Fig. 3 ROC curves by ESS-ELM algorithm, ESS-SVM algorithm, HEART score, TIMI score, and GRACE score

Compared with the original ESS algorithm, we use ELM-derived parameters (i.e., Acc_t and $\|\beta_t\|^2$) instead of the over-sampled data for weight calculation. By using this weight, the computationally expensive over-sampling component in the original ESS is replaced with an efficient ELM algorithm. Moreover, we assign a randomly selected number of hidden nodes to each individual classifier so that the diversity within the decision ensemble is increased. In ensemble learning, it is known that a diversified decision ensemble has better generalization ability [50]. The detailed description of the ESS-ELM algorithm is presented in Fig. 2.

Results

As observed in our dataset, the average age of all patients was 60 years old and majority of the cohort were male (68%) and Chinese (62%). Table 1 shows predictive variables including HRV, ECG, and vital signs. We found out that in linear HRV predictors, only time domain parameters (normalized LF power, normalized HF power, and LF/HF) were statistically significant ($p < 0.05$). Moreover, approximate entropy, sample entropy, and DFA $\alpha1$ of the non-linear HRV variables were also significant. In 12-lead ECG predictors, the proportions of patients with ST-elevation, ST-depression, and Q-waves were much higher in the group of patients who met the outcome. Among vital sign predictors, we only observed that pain score was statistically significant.

Figure 3 illustrates the ROC curves generated by ESS-ELM algorithm, ESS-SVM algorithm, HEART score, TIMI score, and GRACE score. In performance evaluation, we use the leave-one-out cross-validation (LOOCV) framework. Figure 3a shows the comparisons between ESS-ELM and the clinical scores and Fig. 3b shows the comparisons between ESS-SVM and the clinical scores. Obviously, both

Table 2 Comparison of the predictive results. The range of number of hidden nodes for ELM algorithms was [10,100] and ensemble size was 100 for both ESS-ELM and ESS-SVM algorithms

Method	AUC (95% CI)	Cutoff	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)
ESS-ELM	0.778 (0.732–0.825)	34.9	75.3 (68.4–82.3)	70.4 (66.8–73.9)	36.3 (30.9–41.7)	92.7 (90.4–95.0)
ESS-SVM	0.759 (0.711–0.807)	19	74.7 (67.6–81.7)	69.6 (66.1–73.1)	35.5 (30.2–40.9)	92.4 (90.1–94.8)
HEART	0.758 (0.710–0.806)	5	76.0 (69.1–83.0)	63.9 (60.2–67.6)	32.1 (27.2–37.0)	92.2 (89.8–94.7)
TIMI	0.618 (0.566–0.671)	2	59.6 (51.6–67.5)	54.1 (50.2–57.9)	22.5 (18.4–26.7)	85.6 (82.3–89.0)
GRACE	0.593 (0.540–0.646)	108	56.2 (48.1–64.2)	56.8 (53.0–60.6)	22.6 (18.3–26.9)	85.3 (81.9–88.6)

ESS-ELM and ESS-SVM algorithms significantly outperform TIMI and GRACE scores. As shown in Table 2, ESS-SVM and HEART achieve similar AUC values while ESS-SVM gives better specificity and PPV. ESS-ELM is the top performer among all scores. Both ESS-based algorithms use 100 as the ensemble size and ESS-ELM adopts [10,100] as the range of number of hidden nodes for its individual ELM classifiers. The results presented in Table 2 are derived from optimal cutoff score for each method; optimal cutoff point on the ROC curve is defined as the point nearest to the upper-left corner.

We note several factors that determine algorithm complexity and predictive performance, for example, the number of hidden nodes and the ensemble size. As seen from the algorithm architecture (Fig. 1), the complexity of ESS-ELM has a linear correlation with the ensemble size. Compared with the original ESS algorithm [10], ESS-ELM reduces computations by removing the over-sampling component. Figure 4a demonstrates the impacts on AUC using different number of hidden nodes. As mentioned in the previous section, we use different number of hidden nodes for each individual ELM classifiers, so the number indicated in the figure is the range of number of hidden nodes, i.e., [10,25], [10,50], [10,100], [10,150], and [10,200]. The best prediction performance was obtained when the range of number of hidden nodes was [10,100]. Figure 4b depicts different
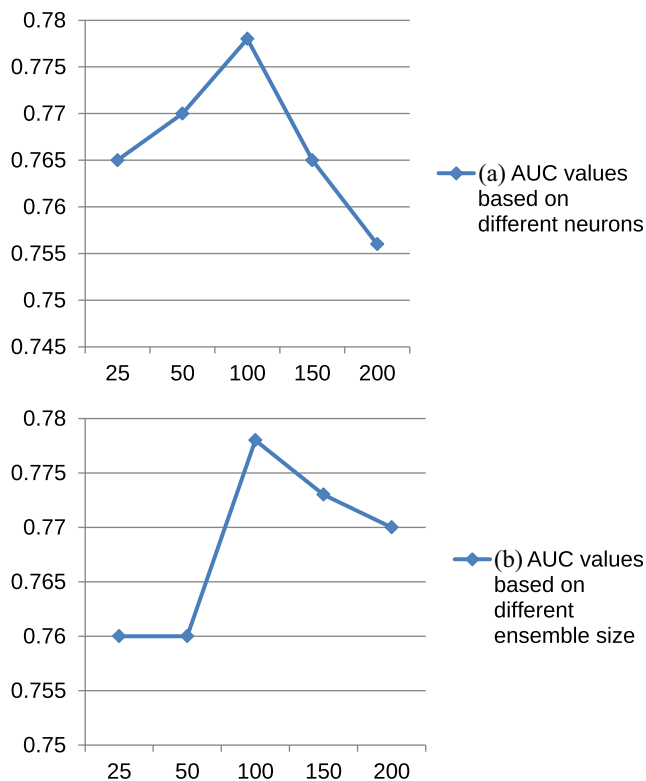
ensemble sizes and their corresponding AUC values while keeping the range of hidden nodes as [10,100]. We observe that the highest AUC value was achieved when ensemble size was 100. In general, larger ensemble size corresponded to better prediction performance. In all of our experiments on the ESS-ELM algorithm, we chose ensemble size as 100 and the hidden node range as [10,100] to produce the best trade-off between predictive performance and complexity.

Conclusion

In this paper, we presented an ensemble-based risk scoring and conducted an observational study of 797 ED chest pain patients. We proposed a novel ELM-based ensemble scoring method named as ESS-ELM, and demonstrated that the new algorithm outperformed the original ESS-SVM algorithm and three established clinical scores, namely HEART, TIMI, and GRACE. Moreover, we have investigated the effects of parameter changes in terms of ROC analysis. AUC values, sensitivity, specificity, PPV, and NPV, were used as the performance indicators. ELM has shown the flexibility on its integration with the ESS framework.

Fig. 4 The effects of the number of hidden nodes and ensemble size on area under the curve (AUC) values

References

1. Bhuiya FA, Pitts SR, McCaig LF. Emergency department visits for chest pain and abdominal pain: United States, 1999-2008, NCHS Data Brief. 2010.
2. Lindsell CJ, Anantharaman V, Diercks D, Han J, Hoekstra JW, Hollander JE, Kirk JD, Lim SH, Peacock WF, Tiffany B, Wilke EK, Gibler WB, Jr, CVP. The internet tracking registry of acute coronary syndromes (i*trACS): a multicenter registry of patients with suspicion of acute coronary syndromes reported using the standardized reporting guidelines for emergency department chest pain studies. Ann. Emerg. Med. 2006;48:666–677.
3. Ong MEH, Goh K, Fook-Chong S, Haaland B, Wai KL, Koh ZX, Shahidah N, Lin Z. Heart rate variability risk score for prediction of acute cardiac complications in ED patients with chest pain. Am. J. Emerg. Med. 2013;31:1201–1207.
4. Heldeweg MLA, Liu N, Koh ZX, Fook-Chong S, Lye WK, Harms M, Ong MEH. A novel cardiovascular risk stratification model incorporating ECG and heart rate variability for patients presenting to the emergency department with chest pain. Crit Care. 2016;20(1):179.
5. Six A, Backus B, Kelder J. Chest pain in the emergency room: value of the HEART score. Neth Heart J. 2008;16(6):191–196.
6. Backus B, Six A, Kelder J, Bosschaert MAR, Mast E, Mosterd A, Veldkamp R, Wardeh A, Tio R, Braam R, Monnink SHJ, Van Tooren R, Mast T, Van den Akker F, Cramer MJM, Poldervaart J, Hoes A, Doevendans P. A prospective validation of the HEART score for chest pain patients at the emergency department. Int J Cardiol. 2013;168(3):2153–2158.
7. Six A, Cullen L, Backus B, Greenslade J, Parsonage W, Aldous S, Doevendans PA, Than M. The HEART score for the assessment of patients with chest pain in the emergency department: a multinational validation study. Crit Pathw Cardiol. 2013;12(3):121–126.
8. Antman E, Cohen M, Bernink P, McCabe C, Horacek T, Papuchis G, Mautner B, Corbalan R, Radley D, Braunwald E. The TIMI risk score for unstable angina/non-ST elevation MI - a method for prognostication and therapeutic decision making. JAMA. 2000;284(7):835–842.
9. Granger CB, Goldberg RJ, Dabbous O, Pieper KS, Eagle KA, Cannon CP, Van De Werf F, Avezum A, Goodman SG, Flather MD, Fox KA. Predictors of hospital mortality in the global registry of acute coronary events. Arch Intern Med. 2003;163(19):2345–2353.
10. Liu N, Koh ZX, Chua EC, Tan LM, Lin Z, Mirza B, Ong MEH. Risk scoring for prediction of acute cardiac complications from imbalanced clinical data. IEEE J Biomed Health Inform. 2014;18(6):1894–1902.
11. Liu N, Koh ZX, Goh J, Lin Z, Haaland B, Ting BP, Ong MEH. Prediction of adverse cardiac events in emergency department patients with chest pain using machine learning for variable selection. BMC Med Inform Decis Mak. 2014;14:75.
12. Liu N, Lee MAB, Ho AFW, Haaland B, Fook-Chong S, Koh ZX, Pek PP, Chua EC-P, Ting BP, Lin Z, Ong MEH. Risk stratification for prediction of adverse coronary events in emergency department chest pain patients with a machine learning score compared with the timi score. Int J Cardiol. 2014;177(3):1095–1097.
13. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. J Artif Intell Res. 2002;16(1):321–357.
14. Huang G-B, Wang D, Lan Y. Extreme learning machines: a survey. Int J Mach Learn Cybern. 2011;2:107–122.
15. Rong H-J, Ong YS, Tan AH, Zhu Z. A fast pruned-extreme learning machine for classification problem. Neurocomputing. 2008;72:359–366.
16. Miche Y, Sorjamaa A, Bas P, Simula O, Jutten C, Lendasse A. OP-ELM: optimally pruned extreme learning machine. IEEE Trans Neural Netw. 2010;21:158–162.
17. Liu N, Wang H. Evolutionary extreme learning machine and its application to image analysis. J Signal Process Syst. 2013;73:1–9.
18. Luo J, Vong C-M, Wong P-K. Sparse bayesian extreme learning machine for multi-classification. IEEE Trans Neural Netw Learn Syst. 2014;25(4):836–843.
19. Liu N, Wang H. Ensemble based extreme learning machine. IEEE Signal Process Lett. 2010;17:754–757.
20. Cao J, Chen T, Fan J. Landmark recognition with compact BoW histogram and ensemble ELM. Multimed Tools Appl. 2016;75(5):2839–2857.
21. Wong PK, Gao XH, Wong KI, Vong CM. An analytical study on reasoning of extreme learning machine for classification from its inductive bias. Cogn Comput. 2016;8(4):746–756.
22. Vong C-M, Ip W-F, Chiu C-C, Wong P-K. Imbalanced learning for air pollution by meta-cognitive online sequential extreme learning machine. Cogn Comput. 2015;7(3):381–391.
23. Vong C-M, Ip W-F, Wong P-K, Chiu C-C. Predicting minority class for suspended particulate matters level by extreme learning machine. Neurocomputing. 2014;128:136–144.
24. Rong H-J, bin Huang G, Liang Y-Q. Fuzzy extreme learning machine for a class of fuzzy inference systems. Int J Uncertain Fuzz. 2013;21:51–61.
25. Rong H-J, Huang G-B, Sundararajan N, Saratchandran P. Online sequential fuzzy extreme learning machine for function approximation and classification problems. IEEE Trans Syst Man, Cybern B, Cybern. 2009;39(4):1067–1072.
26. Peng X, Lin P, Zhang T, Wang J. Extreme learning machine-based classification of ADHD using brain structural MRI data. PLOS ONE. 2013;8:e79476.
27. Liu N, Cao J, Koh ZX, Pek PP, Ong MEH. Risk stratification with extreme learning machine: A retrospective study on emergency department patients, Mathematical Problems in Engineering, vol. 2014, pp. Article ID 248 938, 6 pages. 2014.
28. Kim J, Shin HS, Shin K, Lee M. Robust algorithm for arrhythmia classification in ECG using extreme learning machine. Biomed Eng Online. 2009;8:31:1–12.
29. Cao J, Zhang K, Luo M, Yin C, Lai X. Extreme learning machine and adaptive sparse representation for image classification. Neural Netw. 2016;81:91–102.
30. Cao J, Hao J, Lai X, Vong C-M, Luo M. Ensemble extreme learning machine and sparse representation classification algorithm. J Franklin Inst. 2016;353:4526–4541.
31. Zhang R, Huang G-B, Sundararajan N, Saratchandran P. Multicategory classification using extreme learning machine for microarray gene expression cancer diagnosis. IEEE/ACM Trans Comput Biol Bioinform. 2007;4:485–495.
32. Lu H-J, An C-L, Zheng E-H, Lu Y. Dissimilarity based ensemble of extreme learning machine for gene expression data classification. Neurocomputing. 2014;128:22–30.
33. Cao J, Wang W, Wang J, Wang R. Excavation equipment recognition based on novel acoustic statistical features. IEEE Trans Cybern. 2016;PP(99):1–13.
34. Wong PK, Zhong J, Yang Z, Vong CM. Sparse bayesian extreme learning committee machine for engine simultaneous fault diagnosis. Neurocomputing. 2016;174:331–343.
35. Niskanen J-P, Tarvainen MP, Ranta-Aho PO, Karjalainen PA. Software for advanced HRV analysis. Comput Methods Programs Biomed. 2004;76(1):73–81.

36. Levick JR. An introduction to cardiovascular physiology: Butterworth-Heinemann; 2013.
37. Task Force of the European Society of Cardiology the North American Society of Pacing Electrophysiology, Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. Circulation. 1996;93:1043–1065.
38. Brennan M, Palaniswami M, Kamen P. Do existing measures of poincare plot geometry reflect nonlinear features of heart rate variability? IEEE Trans Biomed Eng. 2001;48(11):1342–1347.
39. Pincus S. Approximate entropy (ApEn) as a complexity measure. Chaos. 1995;5(1):110–117.
40. Richman JS, Moorman JR. Physiological time-series analysis using approximate entropy and sample entropy. Am J Physiol Heart Circ Physiol. 2000;278(6):H2039–H2049.
41. Hsin H-T, Yang C-Y, Yeih D-F, Shieh J-S, Li, A-H. The detrended fluctuation analysis of acute-phase heart-rate variability in acute coronary syndromes - a pilot study. Int J Cardiol. 2010;140(2):252–255.
42. Tulppo MP, Kiviniemi AM, Hautala AJ, Kallio M, Seppänen T, Mäkikallio TH, Huikuri HV. Physiological background of the loss of fractal heart rate dynamics. Circulation. 2005;112(3):314–319.
43. Tapanainen JM, Thomsen PE, Køber, L, Torp-Pedersen C, Mäkikallio, TH, Still AM, Lindgren KS, Huikuri HV. Fractal analysis of heart rate variability and mortality after an acute myocardial infarction. Am J Cardiol. 2002;90(4):347–352.
44. Penzel T, Kantelhardt JW, Grote L, Peter JH, Bunde, A. Comparison of detrended fluctuation analysis and spectral analysis for heart rate variability in sleep and sleep apnea. IEEE Trans Biomed Eng. 2003;50(10):1143–1151.
45. Huang G-B, Zhou H, Ding X, Zhang R. Extreme learning machine for regression and multiclass classification. IEEE Trans Syst Man Cybern B Cybern. 2012;42:513–529.
46. Huang G-B, Zhu Q-Y, Siew CK. Extreme learning machine: theory and applications. Neurocomputing. 2006;70:489–501.
47. Serre D. Matrices: Theory and applications. New York: Springer; 2002.
48. He H, Garcia EA. Learning from imbalanced data. IEEE Trans Knowl Data Eng. 2009;21:1263–1284.
49. Han J, Kamber M. Data Mining: Concepts and Techniques: Morgan Kaufmann; 2006.
50. Polikar R. Ensemble based systems in decision making. IEEE Circ Syst Mag. 2006;6:21–45.