Thailand Statistician
January 2017; 15(1): 43-53
http://statassoc.or.th
Contributed paper

# The Classification of Severity Level in Exertional Heatstroke Patients by Applying a Decision Tree Technique

**Uthumporn Domthong*[a], Prem Junsawang [a], Virissorn Wongsrichanalai [b], Prapun Kittiworawittkul [b], and Pornanan Domthong [c]**
[a] Department of Statistics, Faculty of Science, Khon Kaen University, Khon Kaen, 40002 Thailand
[b] Pulmonary and Critical Care Unit, Phramongkutklao Hospital, Bangkok, 10400 Thailand
[c] Pulmonary and Critical Care Unit, Department of Internal Medicine, Khon Kaen Hospital, Khon Kaen, 40000 Thailand
*Corresponding author; e-mail: uthudo@kku.ac.th

Received: 6 September 2015
Accepted: 1 March 2016

**Abstract**

Heatstroke (HS) is a life-threatening illness that can lead to multiorgan dysfunction. It occurs when a body temperature is very high. This disease is rarely found in Thailand. In this paper, we propose a model of exertional heatstroke (EHS) classification using a decision tree for classifying the severity level in this group of patients. The model is formulated based on a data of 42 patients who were considered as EHS patients. The data was obtained from a retrospective single-center study conducted at a medical intensive care unit (ICU), Phramongkutklao hospital, Thailand during 2008-2014. To construct the classification model, a decision tree with 10 relevant variables is used. Those variables are the delayed cooling and temperature management, cardiovascular collapse with vasopressor use, disseminated intravascular coagulation (DIC), acute kidney injury with renal replacement therapy (RRT), therapeutic mild hypothermia, rhabdomyolysis, respiratory failure, concurrent infectious illness, hypoglycemic event, and hepatocellular injury. The data were divided into two datasets. The first one was used to create the model and the other was used to validate the model. The results are acquired with high percentage of classification accuracy (91.0%) and the true positive rate and the true negative rate are 0.97 and 0.73, respectively. The delay cooling and temperature management was ranked as the most relevant variable classifying the severity level in exertional heatstroke appearance. Since early diagnosis and prompt treatment for this group of patients are critical, this classification model using decision tree technique can be useful for the classification of severity level in EHS.

**Keywords**: Classification, decision tree, exertional heatstroke, severity.

## 1. Introduction

Heatstroke (HS) is an uncommon disease in Thailand. However, it carries a high mortality rate, and also HS survivors sometimes have a risk of permanent neurological damage (Fan et al. 2015, and Leon and Bouchama 2015). Heatstroke has been classified as exertional and classic type (Khogali and Weiner 1980). Exertional heatstroke is precipitated by heavy exertion in extremely hot and humid climates and is usually seen in otherwise healthy young people (Bouchama and Knochel 2002). Classic heatstroke results from exposure to high temperatures and humidity (Varghese et al. 2005).

Exertional heatstroke (EHS) typically affects young athletes or military personnel who are pushed to their physical limits, and suffer a clinical and pathological syndrome caused by an inability to dissipate heat produced by strenuous exercise (Brian et al. 2015). This can occur at any time of year. There are usually identifiable risk factors such as dehydration, concurrent illness, lack of sleep, obesity, alcohol ingestion, wearing too much clothing, or poor cardiovascular fitness (Shapiro and Seidman 1990). On the other hand, classical or environmental heat illness occurs in those whose thermoregulatory control mechanisms are inefficient, such as very young or elderly people, and those subjected to extreme temperatures. The main factor in the development of this condition is a high environmental temperature, with clusters of cases occurring after heat waves (700 heat related deaths were reported after the Chicago heat wave of 1995), and the annual pilgrimages in the Middle East (Easterling et al. 2000, Misset et al. 2006, Tibbetts 2007, and Smith 2015).

Varghese et al. (2005) investigated predictors of multiple organ dysfunctions in patients presenting with HS. They concluded that aggressive measures to lower the body temperature with other supportive therapy could substantially reduce the mortality. Sithinamsuwan et al. (2009) reported their 12-years study with Thai EHS patients, and described the incidence, clinical characteristics, laboratory profiles, outcomes of EHS associated with early recognition, and aggressive combined cooling. They proposed that an optimal cooling target of core body temperature of 38.0°C achieved within 3 hours after presentation to the emergency room. Domthong et al. (2015) studied EHS patients in a medical intensive care unit (ICU) of Phramongkutklao Hospital from 2008 to 2014. They emphasize that the early aggressive combined cooling and promptly temperature management are the mainstay treatment of EHS to improve morbidity and mortality outcome.

The objective of this research is to find a model of exertional heatstroke (EHS) classification using a decision tree to indicate the severity in EHS patients.

## 2. Classification and Regression Tree

Decision trees have been widely accepted as a decision making technique providing high classification accuracy with a simple knowledge representation. During the past 10 years, decision trees have been applied in medical applications. Vili et al. (2002) presented the basic characteristics and applications of decision trees in medicine. Lukas et al. (2008) applied decision tree technique to predict diagnosis and prognosis of dengue disease in the early phase of illness. They concluded that this technique was useful in disease management and surveillance. Additionally, the decision tree techniques could be also used with other techniques to automatic disease diagnosis (Exarchos et al. 2012, Omiotek et al. 2013 and Hong et al. 2016).

Classification and regression tree (CART) is a type of decision tree technique which is constructed to create a predictive model. CART is constructed by recursively partitioning a dataset called learning dataset and fitting the sub-tree model within each partition (Loh 2011). Given a dataset, a group of data is represented by a node called a parent node in a decision tree. The parent node can be partitioned into two groups also represented by nodes called child nodes. The binary partition is performed iteratively (Areerachakul et al. 2010). CART analysis consists of four basic steps (Breiman et al. 1984) as follows:

(1) A tree is built using recursive splitting of nodes. Each resulting node is assigned to a predicted class based on the distribution of classes in the learning dataset which would occur in that node and the assignment of a predicted class to each node occurs whether that node is subsequently split into child nodes. This step is called tree building step.

(2) At this point, a maximal tree has been produced which probably greatly over fits the information contained within the learning data set. This step is called stopping tree building step.

(3) The creation of a sequence of simpler trees through the cutting off increase important nodes. This step is called pruning step.

(4) During the learning process, the optimal tree selection should fit the information in the training dataset, but it should not over-fit the information from the sequence of pruned trees.

## 3. Methodology

### 3.1. Patient selection

In this research, we analyzed a dataset from patients diagnosed with EHS, and admitted to ICU at Phramongkutklao hospital, Bangkok, Thailand from 2008 to 2014. We reviewed medical records from the emergency room (ER), ICU, and medical wards. Our diagnostic criteria for EHS were: (i) the core body temperature of more than 40.0 $^0$c and reliable history of compatible environmental exposure, (ii) evidence of central nervous system dysfunction, seizures or altered of sensorium (disorientation, delirium or coma) that was documented by attending physicians, (iii) history of strenuous exercise. Exclusion criteria were those with proven central nervous system (CNS) infection, systemic sepsis, malaria, malignancy, neuroleptic malignant syndrome, and malignant hyperthermia secondary to anesthetic agents at the time of admission. There were 42 patients who met the criteria of EHS. Thirty-one of them were identified as good outcome status, referred that they survived and recover nearly normal organ function. On the other hand, 11 of them were poor outcome status, referred that they were death, or exhibited organ dysfunction (Table 1).

**Table 1** Treatment outcome

| Profile | Total (n=42) |
|---|---|
| Good treatment outcome (Good) | 31 (73.8%) |
| Poor treatment outcome (Poor) | 11 (26.2%) |
| - Death | 3 (27.3%) |
| - Neurologic sequelae | 8 (72.7%) |

### 3.2. Data preparation

We started from understanding the hospital process of admitting and diagnosing EHS patients who admitted in ICU. Moreover, we collected all clinical and demographic variables that physicians used to diagnose EHS. The data was collected from the medical record division. We selected all relevant variables to the staging area. Then, we validated and transform all variables until we have 10 variables as shown in Table 2.

**Table 2** Variables and Explanation

| Variables | Explanation |
| --- | --- |
| 1.The delayed cooling and temperature management | Fail to achieve core body temperature management (body temperature is less than 38°C longer than 3 hours after onset of symptoms). |
| 2. Cardiovascular collapse with vasopressor use | Hemodynamics instability after adequate volume resuscitation and using vasopressor is needed. |
| 3. DIC (clinical and lab) | Intravascular coagulation with significant clinical bleeding, prolong prothrombin time and thrombocytopenia. |
| 4. Acute kidney injury with RRT | Acute kidney injury patients who need renal replacement therapy (RRT) for the waste product clearance. |
| 5. Therapeutic mild hypothermia | Core body temperature is decreased into 32°C -34°C at least 12-24 hours. |
| 6. Rhabdomyolysis | Serum creatinine kinase linage level is increased and this may lead to myoglobinuria or renal failure. |
| 7. Respiratory failure | The failure of spontaneously ventilated lung and mechanical ventilation support is needed. |
| 8. Concurrent infectious illness | Coexisting of infectious diseases with exertional heatstroke such as bronchitis, pneumonia, urinary tract infection, skin and soft tissue infection. |
| 9. Hypoglycemic event | Blood glucose level below 60 mg/dl. |
| 10. Hepatocellular injury | Elevate the liver enzymes with or without clinical of hepatic encephalopathy. |

### 3.3. Model calibration and validation

In this study, the three-folds cross validation was used to evaluate the created decision tree model by CART technique since there are only eleven cases assigned as poor outcome status. The original sample is randomly divided into three equal parts. Two parts were used to construct a decision tree model, called a training set, and the third part is used to validate the model, called a test set. The training and test selections are performed iteratively four times with distinctive test set.

### 3.4. Performance matrices

In this study, it could be considered as an imbalance classification problem. Most classification performance measure is computed based on a confusion matrix as shown in Table 3.

**Table 3** Confusion matrix

| | | True Class | |
| --- | --- | --- | --- |
| | | Good | Poor |
| Hypothesis Output | Good | TP | FP |
| | Poor | FN | TN |

True Positive (TP) and True Negative (TN) denote the number of positive and negative cases correctly classified. False Positive (FP) and False Negative (FN) denote the number of misclassified positive and negative cases. For model performance measure, classification accuracy, true positive rate and true negative rate are adopted and given by the following formulas:

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \tag{1}$$

$$\text{true positive rate} = \frac{TP}{TP + FN}, \tag{2}$$

and

$$\text{true negative rate} = \frac{TN}{TN + FP}. \tag{3}$$

## 4. Results

From 2008 to 2014, 42 patients met the inclusion criteria of EHS and were recruited to our study (Table 4). All patients were male. Forty-one cases were soldiers (97.6 %) and one case was a college student (2.4%). The median age was 22.2 years (Quartile Deviation (QD) = 2.4), the median body weight was 82.4 kg (QD = 4.5), the median height was 174.8 cm (QD = 3.5), and the median BMI was 26.1 kg/m$^2$ (QD = 1.85).

As mentioned, the three-fold cross validation was used to evaluate the model create by CART technique. Forty-two EHS patients were divided into two groups. The first group was good outcome status (31 cases), and another group was poor outcome status (11 cases). The percentage of good and poor treatment outcome status for 10 clinical profiles was shown in Table 5. Moreover, the two independent sample t-tests along with the Levene's test for equality of variance were used to test the difference of the proportion between these two groups at the 0.05 level of significance as shown in Table 5.

In this study, CART was performed by MATLAB programming. The confusion matrix of three-fold cross validation is given in Table 6.

From Table 6, it is shown that TP and TN values are equal to 30 and 8, respectively. The CART classified correctly in good and poor labels which is equal to 38 cases out of 42 cases. FP and FN values are 3 and 1, respectively. There are only 4 cases out of 42 cases are misclassified. Based on Equations (1)-(3), the classification accuracy, the true positive rate, and the true negative rate are 0.91, 0.97 and 0.73, respectively. From the experiment, three decision tree models are obtained on three-fold cross validation. One of decision tree models is shown in Figure 1. The cycle represents parent node or variable nodes used for classification and rectangle represents an assigned class label.

**Table 4** Demographic Data

| Profile | Number | % |
|---|---|---|
| Sex, male; n (%) | 42 | 100% |
| Occupation; n (%) | | |
| Soldier | 41 | 97.6% |
| Non soldier | 1 | 2.4% |
| | Median | Quartile Deviation(QD) |
| Age (year) | 22.2 | 2.4 |
| Body weight (kg) | 82.4 | 4.5 |
| Height (cm) | 174.8 | 3.5 |
| Body mass index (kg/m$^2$) | 26.1 | 1.85 |

**Table 5** The number and the percentage of clinical profiles from good and poor treatment outcome status

| Profile | Good treatment outcome (n= 31) | Poor treatment outcome (n=11) | P-value |
|---|---|---|---|
| 1. Delayed cooling | 1 (3.2%) | 8 (72.7%) | 0.001* |
| 2. CV collapse | 10 (32.3%) | 10 (90.9%) | 0.000* |
| 3. DIC | 5 (16.1%) | 9 (81.8%) | 0.000* |
| 4. AKI with RRT | 2 (6.5%) | 6 (54.5%) | 0.013* |
| 5. Hypothermia | 13 (41.9%) | 3 (27.3%) | 0.402 |
| 6. Rhabdomyolysis | 23 (71.2%) | 11 (100%) | 0.003* |
| 7. Respiratory failure | 25 (80.6%) | 11 (100%) | 0.012* |
| 8. Concurrent infectious illness | 10 (32.3%) | 8 (72.7%) | 0.019* |
| 9. Hypoglycemic event | 2 (6.5%) | 4 (36.4%) | 0.084 |
| 10. Hepatocellular injury | 17 (54.8%) | 10 (90.9%) | 0.009* |

* Significance level at 0.05.

**Table 6** Confusion matrix by CART technique based on three-fold cross validation with three values of model performance measure

| | | True Class | | Classification Accuracy |
|---|---|---|---|---|
| | | Good | Poor | |
| Hypothesis Output | Good | 30 | 3 | 0.91 |
| | Poor | 1 | 8 | |
| True Positive Rate | | 0.97 | | |
| True Negative Rate | | 0.73 | | |

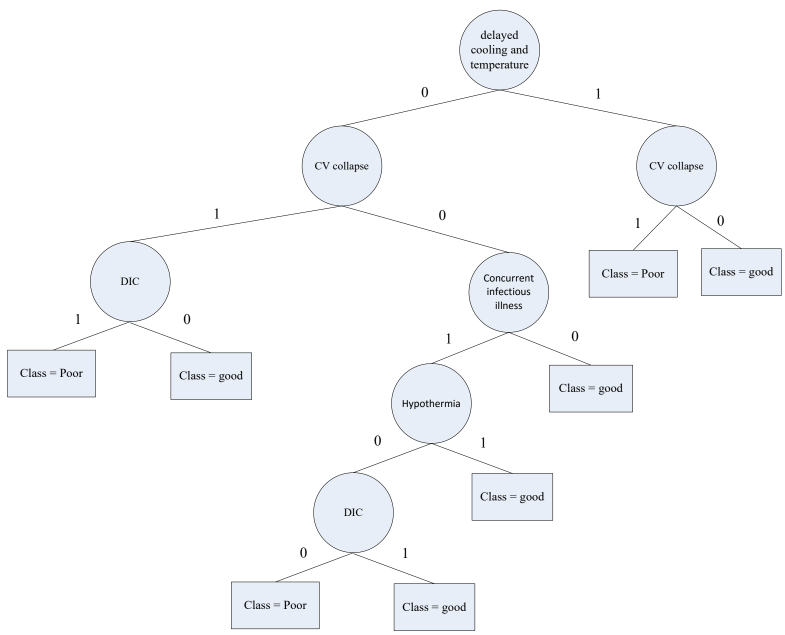


**Figure 1** Example of binary decision tree model from three-fold cross validation (1 = Yes, 0 = No).

Figure 1 illustrates the obtained decision tree model. The delayed cooling and temperature management variable is the most important variable in classification (The first node in the graph) based on Gini Index value. CV collapse with vasopressor use, DIC (clinical and lab), and concurrent infectious illness are also important variables to be used for classification. Moreover, the delayed cooling and temperature management, CV collapse with vasopressor use, DIC (clinical and lab) are the most important variables from three decision tree models.

For deployment, all three decision tree models are used to assign a class by majority vote criterion. A final class label is determined by the majority of answers obtained from three models. If there is equal of answers of good and poor, then the final answer are unclassified. The confusion matrix based on majority vote method is given in Table 7.

**Table 7** The confusion matrix based on majority vote method with three values of model performance measure

| | | True Class | | Classification Accuracy |
|---|---|---|---|---|
| | | Good | Poor | |
| Hypothesis Output | Good | 30 | 0 | 0.98 |
| | Poor | 1 | 11 | |
| True Positive Rate | | 0.97 | | |
| True Negative Rate | | 1.00 | | |

## 5. Discussion

### 5.1. Key findings

According to the decision tree model and ten clinical variables, we found that the most relevant variable that is able to classify the severity level in EHS is the delayed cooling and temperature management. Others predictors were also considered in this model, and we can explain as the following:

(i) Patients who failed to achieve core body temperature and CVS collapse with vasopressor use will be classified as a poor outcome group.

(ii) Patients who achieve core body temperature, but they also have CVS collapse with vasopressor use, clinical and lab DIC, and concurrent infectious illness, then they will be classified as a poor outcome group as well.

### 5.2. Relationship to previous study

According to Sithinamsuwan et al. (2009), the dataset from 28 EHS patients of Phramongkutklao Hospital indicated that the aggressive combined cooling and maintain hemodynamics were the main factors to decrease morbidity and mortality rate in EHS patients. This conclusion is consistent with our study which used the different statistical techniques and collected the new dataset from different time period. Apparently, the aggressive combined cooling is the mainstay treatment strategy of patients to improve neurologic sequelae and to decrease mortality outcome. Furthermore, to correct volume deficit and maintain hemodynamics, we found that adequate fluid resuscitation at the beginning of the treatment is very important because EHS patients obviously have dehydration, so if they also have the delayed fluid resuscitation, this may lead to CVS collapse. After that they may have vasopressor, and then this will classify them into poor outcome group.

Besides the delayed cooling and temperature management, and CVS collapse which are the related factors to previous study, we also found other clinical factors that can predict poor outcome group. Those is clinical and lab DIC, and concurrent infectious illness. Patients who have DIC may have life-threatening bleeding complication. Thus, this may lead to fatality. Besides EHS itself, the high-grade fever in EHS patients may cause some concurrent infectious illness such as pneumonia, urinary tract infection, skin and soft tissue infection. Therefore, the septic work up and empirical antibiotic treatments have to be administered if concurrent infectious illness cannot be excluded.

### 5.3. Strengths and limitations

This is the study that collected the most number of EHS patients in Thailand. In addition, we use the decision tree model, the different statistical technique from other previous studies. However, even we have collected the dataset for 7 years; we still had the small number of sample size. The reasons are that EHS is found in only some period of a year and only some group of patients such as soldiers in the military units. Moreover, this is a retrospective dataset, so we might have some missing data. Therefore, not only review a medical charts, but we also tried to get more data from computer database, laboratory database, and nurse notes for minimizing the missing data.

### 5.4. Implication of our findings

In this study, we received the clinical factors that can predict the disease severity in EHS, and also we can use this finding to improve and develop the standard of care in ICU for EHS patients as the following:

(i) EHS patients should be treated for the aggressive combined external and internal cooling immediately. This should be a pre-hospital management. Before transferring patients to hospital, we suggest decreasing the body temperature at the accident scene. After patients get to the hospital, we still have to continuously maintain the body temperature to prevent the rebound core body temperature which may lead to brain damage.

(ii) EHS patients should receive the adequate fluid resuscitation to maintain hemodynamics, be controlled a blood pressure, and keep adequate organ perfusion.

(iii) EHS patients with hematologic complication from DIC should receive a special care for concerning of a bleeding complication, and also consider giving the blood component replacement if indicated.

(iv) Fever is the major vital sign that we usually found in EHS patients; however, this may also occur from infection originally. From this study, we found that EHS patients may have concurrent infectious illness such as pneumonia, UTI, skin and soft tissue infection. Therefore, physicians should watchfully checkup patients and also cautiously take laboratory investigations before starting the empirical antibiotic therapy in this group of patients who infectious process may be doubtful.

The result from this study was shown to the Heat Stroke Patient Care Team at Phramongkutklao Army Hospital to develop clinical practice guideline and clinical tracer of EHS patients. This guideline will be distributed to hospitals in Thailand to improve the standard of care and reduce the morbidity and mortality rate.

Since this is the single-center data set with a small number of sample sizes, so if we can collect more data from multi-center, then we may detect more relevant variables that can be useful for updating this model.

## 6. Conclusions

The results are acquired with high percentage of classification accuracy and precision values. The delay cooling and temperature management was ranked as the most relevant variable relating to exertional heatstroke appearance. This predictor is consistent with the previous study. Since early diagnosis and prompt treatment for this group of patient are critical, this classification model using decision tree technique can be useful for the classification of severity level in EHS patients.

**Acknowledgements**

We would like to thank all staffs in Pulmonary and Critical Care Division, Department of Medicine, Phramongkutklao hospital, Bangkok, Thailand for helping and finding the dataset, and also deeply thank to all patients in this study. Also, thanks to Department of Statistics, Faculty of Science, Khon Kaen University for the supporting place and statistics programs to authors to working in this research. Finally, we would like to thanks to Faculty of Science, Khon Kaen University for the financial support.

**References**

Areerachakul S, Sanguansintukul S. Classification and regression trees and MLP neural network to classify water quality of canals in Bangkok, Thailand, Int. J. Intell. Comp. Res. 2010; 1(1/2): 43-50.

Bouchama A, Knochel JP. Heat stroke, N. Engl. J. Med. 2002; 346: 1978-88.

Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and Regression Tree, Champman &Hall, 1984; New York.

Brian KS, Emily MK, Dave C, Scott WS, Brian CB, Daniel ER. On-Site treatment of exertional heat stroke, Am. J. Sports Med. 2015; 43: 823.

Domthong P, Wongsrichanalai V, Kittiworawittkul P. Factors associated with neurologic outcome and mortality in exertional heatstroke treated ICU patients during 2008-2014 in Phramongkutklao hospital, Proceeding of the 31st Annual meeting of the Royal College of Physicians of Thailand, 2015 March 26-29; Bangkok, Thailand. 2015.

Easterling DR, Meehl GA, Parmesan C, Changnon SA, Karl TR, Mearns LO. Climate extremes: observations, modelling, and impacts. Science. 2000; 298: 2068-2074.

Exarchos TP, Tzallas AT, Baga D, Chaloglou D, Fotiadis D, Tsouli S, Diakou M, Konitsiotis S. Using partial decision trees to predict Parkinson's symptoms :A new approach for diagnosis and therapy in patients suffering from Parkinson's disease, Computers in Biology and Medicine. 2012; 42: 195-204.

Fan H, Zhu JH, Ye JH. Cold Hemodiafiltration for exertional heat stroke resulting in multiple organ failure, Ther Apher Dial. 2015; 19: 522.

Hong KS, Melanie PO, Ye CK, Serge D. Multivariate alternating decision trees, Pattern Recog. 2016; 50: 195-205.

Khogali M, Weiner JS. Heat stroke: report on 18 cases, Lancet.1980; 2: 276-278.

Leon LR, Bouchama A. Heat stroke, Comprehensive Physiology. 2015; 5: 611-647.

Loh WY. Classification and regression trees, WIREs Data Mining and Knowledge Discovery. 2011; 1: 14-23.

Misset B, De Jonghe B , Bastuji-Garin S. Mortality of patients with heatstroke admitted to intensive care units during the 2003 heat wave in France: a national multiple-center risk factor study, Crit. Care. Med. 2006; 34: 1087-1092.

Omiotek Z, Burda A, Wójcik W, The use of decision tree induction and artificial neural networks for automatic diagnosis of Hashimoto's disease, Expert Sys. Appl. 2013; 40: 6684-6689.

Shapiro Y, Seidman DS. Field and clinical observations of exertional heat stroke patients, Med. Sci. Sports Exerc. 1990; 22: 6-14.

Sithinamsuwan P, Piyavechviratana K, Kitthaweesin T, Chusri W, Orrawanhanothai P, Wongsa A, Wattanathum A, Chinvarun Y, Nidhinandana S, Satirapoj B, Supasyndh O, Sriswasdi C, Prayoonwiwat W. Exertional Heatstroke: Early recognition and outcome with aggressive combined cooling- a 12-year experience, Mili. Med. 2009; 174(5): 496-502.

Smith JE. Cooling methods used in the treatment of exertional heat illness, Br. J. Sports Med. 2005; 39: 503-507.

Tanner L, Schreiber M, Low JGH, Ong A, Tolfvenstam T, et al. Decision tree algorithms predict the diagnosis and outcome of dengue fever in the early phase of illness, PLoS Neglected Tropical Disease. 2008; 2(3), doi: 10.1371/journal.pntd.0000196.

Tibbetts J. Driven to extremes health effects of climate change, Environ. Health Perspect. 2007; 115: A196-203.

Varghese GM, John G, Thomas K, Abraham OC, Mathai D. Predictors of multi-organ dysfunction in heatstroke, Emerg. Med. J. 2005; 22: 185-187.

Vili P., Peter K., Bruno S., Ivan R., Decision trees: an overview and their use in medicine, J. Med. Sys. 2002; 26(5): 445-463.