Machine Learning-based Heat Risk Estimation for Outdoor Sports

Hanna Suzuki

Bedford High School, Bedford, MA 01730, USA
Hanna.S.Suzuki@gmail.com

SUMMARY

This work aids athletes and coaches to be aware of the heat risk of their activity and take precautions early. The proposed method relies on several meteorological parameters that are commonly available for athletes and coaches and performs supervised machine learning to estimate the current heat alert level in compliance to a standard heat safety policy for athletes. Based on over 15,000 weather data samples in the summer of 2024, decision tree (DT) and random forest (RF) models are trained to perform multiclass classification for heat alert levels. After resampling imbalanced data and adjusting hyperparameters, the DT and RF models yield the accuracy of 91% and 99%, respectively.

KEYWORDS

Heat-related illnesses, multiclass data classification, web bulb globe temperature (WBGT)

INTRODUCTION

Heat-related illnesses (HRIs), such as heat cramps, heat exhaustion and heat stroke, are a critical threat for anyone in summer and can cause death if not taken seriously. In the US, at least 2,325 HRI deaths were reported in 2023[5], and over 10,000 annual HRI deaths are estimated from 1997 to 2006[9].

HRIs affect thousands of athletes every year. There have been at least 77 HRI deaths among athletes in the US since 2000, and 65 of them (84%) are teenagers[11]. Younger generations are more vulnerable due to their high metabolism. HRIs are the top cause of preventable deaths for high school athletes[1].

To assess the risk of HRIs, the Wet Bulb Globe Temperature (WBGT) is known to be effective[1]. It quantifies the heat stress on the human body in direct sunlight by measuring solar radiation (sun exposure), air temperature, humidity, and wind speed[10]. WBGT is similar to heat index in that both are apparent ("feels-like") temperatures calculated with air temperature and humidity. However, heat index uses air temperature in the shade. This is not a reasonable assumption for outdoor sports. In contrast, WBGT considers solar radiation to emphasize the impacts of direct sunlight on body temperature.

Due to its effectiveness, professional, national, and state athletic associations have adopted WBGT as a primary means to formulate their heat safety policies. Table 1 shows the most common policy for high school athletes in the US[4]. It is recommended by the National Federation of State High School Associations and its state affiliates. By analyzing regional variations in acclimatization to heat, this policy defines three regions and sets WBGT thresholds for activity modification in each region. Region 1 (northern regions) has lower thresholds because athletes in the North are less acclimatized to heat than in Region 2 (south of Region 1) and Region 3 (southern regions).

This heat safety policy, unfortunately, is not used often in high school sports because WBGT meters are expensive (>$500ea). Thus, several mathematical models have been developed to estimate WBGT using meteorological parameters[6]. However, some of the parameters, such as solar radiation, are not commonly available for non-experts like athletes and coaches.

This work proposes a method to estimate the current heat alert level from a set of common meteorological parameters even for non-experts. It uses supervised machine learning (ML) to classify a set of meteorological parameters to an alert level in compliance to the heat safety policy in Table 1.

Existing work that use ML for heat risk estimation predict WBGT from periodic time series data[3,7], the number of heatstroke cases from meteorological data including solar radiation[8], and the maximum allowable exposure time from personal, real-time physiological data such as heart rate[2]. The proposed method is different in that it does not rely on time series, solar radiation, nor personal physiological data.

Table 1 WBGT-based Heat Safety Policy by Grundstein et al.[4]

WBGT by Region (F)			Activity Guidelines	Alert Level
Region 1	Region 2	Region 3		
<= 76.0	<= 79.7	<= 82.0	Normal activities. At least 3 separate 3-min breaks each hr	0
76.1-81.0	79.8-84.7	82.1-87.0	Use discretion for intense or prolonged exercise. At least 3 separate 4-min breaks each hour.	
81.1-84.0	84.8-87.7	87.1-90.0	Max 2 hrs of practice. 4+ separate 4-min breaks each hr.	1
84.1-86.0	87.8-89.7	90.1-92.0	Max 1 hour of practice. No conditioning activities. 20-min breaks distributed throughout the hour.	2
>= 86.1	>= 89.8	>=92.1	No outdoor workouts. Delay practice/competitions until a cooler WBGT is reached.	3

METHODS

This work uses the decision tree (DT) and random forest (RF) algorithms to build supervised classifiers. The proposed classifiers take a set of meteorological parameters as an input: air temperature (F), relative humidity (%), cloud cover (%), precipitation (inch), wind speed (knot; nautical miles per hour), and time of day (24-hour notation). The classification output is a heat alert level shown in Table 1: Level 0, 1, 2, or 3. Level 0 is the safest, and level 3 is the riskiest. The DT classifier is configured with the CART method and Gini impurity index. The RF classifier consists of multiple DTs, has them perform classification individually, and produces the final output through majority voting among those DTs.

The proposed classifiers are trained with 15,273 meteorological data samples in the summer of 2024, which were collected from 290 cities in the US with an online database of the National Oceanic and Atmospheric Administration (NOAA). The raw dataset is imbalanced; the number of samples significantly varies across heat alert levels. 76% of the samples are in Level 0, and only 5% of them are in Level 3. Therefore, the raw dataset is randomly under-sampled so that each alert level has the same number of samples (Figure 1). Then, a subset of the under-sampled dataset is used as the test data (652 samples). The raw dataset is also over-sampled with the SMOTE method to produce the training data (45,777 samples; Figure 1). The training data and test data are combined to generate the cross-validation data, which has 46,429 samples.

This work uses Python and its libraries (e.g., scikit-learn, imbalanced-learn, dtreeviz, yellowbrick, and joblib) to implement, configure and visualize the proposed classifiers.

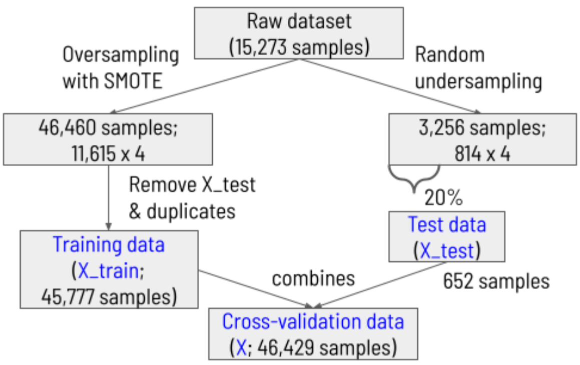

Figure 1 Dataset resampling

	Decision tree		Random forest
	w/o hyperparameter adjustment	w/ hyperparameter adjustment	
Training	100%	91%	100%
Test	90%	84%	95%
Cross validation	95%	91%	99%

Figure 2 Classification accuracy

RESULTS AND DISCUSSION

Figure 2 shows the classification accuracy of the decision tree (DT) and random forest (RF) classifiers that are trained with the training data (`x_train` in Figure 1). Accuracy is measured with the macro F1 score, which considers the precision and recall in each accuracy measurement, because the classifiers perform classification for multiple output classes (i.e., multiple heat alert levels) Training accuracy is the classification accuracy of a trained classifier against the training data. Test accuracy is the accuracy of a trained classifier against the test data (`x_test` in Figure 1). Cross-validation accuracy is the accuracy against the cross-validation data (`x` in Figure 1). Cross-validation performs stratified 5-fold validation, where five folds are made by preserving the percentage of samples for each classification class (i.e., each heat alert level).

With the default hyperparameter settings, the DT classifier's training, test and cross-validation accuracies vary; 100%, 90% and 95%, respectively. (See the "w/o hyperparameter adjustment" column in Figure 2.) This implies that the DT classifier overfits to the training data. Therefore, the following three major hyperparameters are adjusted to avoid this overfitting: `max_depth` (maximum number of layers in a tree), `min_samples_split` (minimum number of samples required to split a tree node into two child nodes), and `min_samples_leaf` (minimum number of samples required to be in a leaf node). Their default values of infinity, 2 and 1 are adjusted to 10, 38 and 18, respectively. This adjustment is performed by observing how cross-validation accuracy changes in comparison to training accuracy as each hyperparameter varies and finding a hyperparameter combination where training and cross-validation accuracies meet. As a result, the training and cross-validation accuracies reach 91%. (See the "/w hyperparameter adjustment" column in Figure 2.)

With the default hyperparameter settings, the RF classifier's training, test and cross-validation accuracies are 100%, 95% and 99% (Figure 2). Since these are close enough with other, it is reasonably safe to conclude that no overfitting occurs. Therefore, hyperparameters are not adjusted. The classifier uses an ensemble of 100 DTs.

In order to analyze the contribution of each input parameter to classification performance, permutation importance is measured for individual parameters. The importance of air temperature, humidity, time of day, cloud cover, wind speed and precipitation are 0.67, 0.53, 0.35, 0.11, 0.06 and 0.02 in the DT classifier, respectively. They are 0.63, 0.48, 0.34, 0.08, 0.05 and 0.03 in the RF classifier.

In summary, Figure 2 demonstrates that the proposed DT and RF classifiers are reliable. They are efficient and lightweight as well. Their training time (with 45,777 samples) are 0.29 and 7.46 seconds, respectively. Their classification time (per sample) are <0.01 and 0.02 seconds. Their storage consumption are 60 KB and 8.7 MB.

CONCLUSIONS

This work studies a machine learning approach to estimate the current heat alert level from a set of common meteorological parameters even for non-experts like athletes and coaches. Based on a NOAA weather dataset in the summer of 2024, experimental results demonstrate that the proposed decision tree and random forest classifiers work as reliable early warning indicators for athletes and coaches. They are also efficient and lightweight enough to be deployed in webapps, smartphone apps and Internet-of-Things devices. A prototype webapp runs at `https://wbgt-estimator.streamlit.app/`.

REFERENCES

1. Casa, D.J. and Stearns, R.L. (2017) Preventing sudden death in sport and physical activity. second edition, Jones & Bartlett Learning.
2. Choi, Y., Seo, S., Lee, J. et al. (2024) A machine learning-based forecasting model for personal maximum allowable exposure time under extremely hot environments. *Sustain Cities Soc* **101**:105140.
3. Ding, K., Huang, Y., Tao, M., et al. (2023) WBGT index forecast using time series models in smart cities. In *Proc. of Int'l Conf. of Algorithms and Architectures for Parallel Processing*.
4. Grundstein, A., Williams, C., Phan, M. et al. (2015) Regional heat safety thresholds for athletics in the contiguous United States. *Applied Geography* **56**:55-60.
5. Howard, J.T, Androne, N., Alcover, K.C., et al. (2024) Trends of heat-related deaths in the US, 1999-2023. *JAMA* **332**(14):1203–1204.
6. Liljegren, J.C., Carhart, R.A., Lawday, P. et al. (2008). Modeling the wet bulb globe temperature using standard meteorological measurements. *J Occup Environ Hyg* **5**(10):645–655.
7. Lu, C., Yun, Y. and Yoon, M. (2021) Application of machine learning to the prediction of WBGT, In *Proc. of Annual Conference of the Society of Instrument and Control Engineers of Japan*.
8. Ogata, S., Takegami, M., Ozaki, T. et al. (2021) Heatstroke predictions by machine learning, weather information, and an all-population registry for 12-hour heatstroke alerts. *Nat Commun* **12**:4575.
9. Weinberger, K.R., Harris, D., Spangler, K.R. et al. (2020) Estimating the number of excess deaths attributable to heat in 297 United States counties. *Environ. Epidemiol.* **4**(3):e096.
10. Yaglou, C.P. and Minard, D. (1957) Control of heat casualties at military training centers. *AMA Arch Ind Health* **16**(4):302–316.
11. Yancey-Bragg, N. and Pulver D.V. (2024) Heat is killing student athletes far too often. Experts say we can reverse the trend. USA Today, September 26.