ISMJ
International SportMed Journal

Determination of the fluid intake needs of endurance athletes using computational intelligence

[1*]Dr Navin R Singh, PhD, [2]Professor Edith M Peters-Futre, PhD

[1]Department of Electronic Engineering, Faculty of Engineering, Science and the Built Environment, Durban University of Technology, Durban, South Africa
[2]Discipline of Human Physiology, School of Medical and Laboratory Sciences Faculty of Health Sciences, University of KwaZulu-Natal, Westville, Durban, South Africa

*Corresponding author. Address at the end of text.

Abstract

The aim of this study was to assess the efficacy of using artificial neural networks (ANNs) to classify hydration status and predict the fluid requirements of endurance athletes. Hydration classification models were built using a total of 237 data sets obtained from 148 participants (106 males,42 females) in field-and laboratory studies involving running or cycling. 116 data sets obtained from athletes who completed endurance events euhydrated (plasma osmolality: 275-295 mmol.kg^{-1}) following *ad libitum* replenishment of fluid intake was used to design prediction models. A filtering algorithm was used to determine the optimal inputs to the models from a selection of 13 anthropometric, exercise performance, fluid intake and environmental factors. The combination of gender, body mass, exercise intensity and environmental stress index in the prediction model generated a root mean square error of 0.24 L.h^{-1} and a correlation of 0.90 between predicted and actual drinking rates of the euhydrated participants. Additional inclusion of actual fluid intake resulted in the design of a model that was 89% accurate in classifying the post-exercise hydration status of athletes. These findings suggest that the ANN modelling technique has merit in the prediction of fluid requirements and as a supplement to *ad libitum* fluid intake practices. **Keywords:** hydration status, classification and prediction, body mass, gender, exercise intensity, environmental stress index

*Dr Navin R Singh, PhD, MSc, MBA

Navin Singh is a senior lecturer at the Department of Electronic Engineering at the Durban University of Technology, Durban, South Africa. He completed his MSc (Eng) in 1998 and his MBA in 2002 at the University of Natal, Durban, South Africa. He qualified as an exercise specialist/personal trainer with the Health and Fitness Professional Association (HFPA) in 2003. After having successfully completed the pre-requisite coursework in the Masters of Sports Medicine degree, he registered for his PhD in Exercise Physiology at the University of Kwa-Zulu-Natal, Durban, South Africa under the supervision of Prof Peters-Futre in 2010. He was awarded his PhD in Exercise Physiology by the University of Kwa-Zulu Natal in April, 2013. His research interests are fluid and hydration balance in athletes with special focus on the use of engineering computational intelligence techniques in this regard. To date he has authored/co-authored 3 reviewed scientific publications spanning the areas of muscle physiology and fluid replacement/hydration status.

Professor Edith M Peters-Futre, PhD

Edith M Peters-Futre is Professor in the Division of Human Physiology of the School of Laboratory Medicine and Medical Sciences at the University of Kwa-Zulu-Natal, Durban, South Africa. She graduated with an MSc (Med) in Sports Science in the Department of Physiology of the University of Cape Town and completed her PhD in Exercise Immunology

at the University of Pretoria. She has authored/co-authored 46 peer-reviewed scientific publications spanning a range of exercise physiology including exercise immunology, sports nutrition, muscle physiology and fluid replacement/hydration status. These include two ISI original research publications with > 100 citations in Exercise Immunology as well as research papers in the field of fluid and hydration balance in the British Journal of Sports Medicine and Clinical Journal of Sports Medicine. She has successfully supervised >20 post-graduate students in Sports Medicine and obtained NRF rating in 2005.
Email: futree@ukzn.ac.za

Introduction

Maintenance of appropriate hydration status can be crucial during endurance exercise. While excessive dehydration has been associated with an impairment of exercise performance[1,2], exercise-associated overhydration and hyponatraemia can lead to loss of consciousness and be life-threatening [3-5].

Factors which affect the hydration needs of athletes include height, weight, body composition, genetic predisposition and metabolic rate, level of conditioning, exercise intensity and duration, environmental conditions, clothing worn and heat acclimation [6]. During exercise, their combined effect determines an individual's sweat rate and urinary output which are the major contributors to their fluid needs. The most recent position stand of the American College of Sports Medicine (ACSM) suggests *ad libitum* drinking of 0.4 to 0.8 L/h, with the higher rates for faster, heavier individuals competing in warm environments and the lower rates for the slower, lighter persons competing in cooler environments for marathon runners who are euhydrated at the start [3]. It however emphasises the importance of individualised fluid and electrolyte replacement schedules for athletes [3]. This necessitates careful customisation of their requirements which is difficult in view of the numerous above-mentioned confounders.

There is therefore a need for models which are able to make static, pre-event predictions of the hourly fluid requirements of athletes based on a number of physiological and environmental factors [7-9]. These include mathematical models that were developed to determine the sweat rate of athletes [10,11] and have been used widely to predict water needs under the assumption that the fluid intake replaces the expected water lost by sweating, and revisions thereof that factor in exposure time and clothing systems [7,9,12]. Engineering models [13] have also been developed to provide for more accurate sweat predictions over a broader range of conditions and applications. Although the % dehydration associated with optimal performance remains a matter of debate [14-16], it is well accepted that individuals should avoid drinking more fluid than the amount needed to replace their sweat losses, during prolonged exercise[3,8,17] with blood osmolality being accepted as the best haematological marker of hydration status [15,18,19].

Because of the complexity of defining and determining the fluid requirements of athletes, we set out to investigate whether an artificial neural network (ANN) which presents a powerful data modelling tool, can be used to capture and represent the complex relationships between the determinants of fluid requirements and the recommended hourly volume of fluid intake needed to maintain euhydration. In addition to predicting their fluid requirements over a range of exercise intensities and environmental conditions, these biologically inspired computer programs which simulate the way in which the human brain processes information[20,21], have found widespread use in the fields of medicine [22,23] and sport[24-29], and can also offer a simplified method of classifying the hydration status of athletes.

Due to the absence of a previously recorded attempt to design a network which encompasses such a wide range of potential confounders, a null hypothesis was set. It was hypothesised that an ANN will not perform well in classifying or accurately predicting the fluid intake requirements of endurance athletes.

Methods
Data collection
Following approval by the relevant institutional research ethics committee, raw data were obtained from 4 separate field studies [19,30-32] and 3 separate laboratory studies [18,33,34] conducted on cyclists and runners, in which fluid intake was recorded and plasma/serum osmolality measurements were made to determine the hydration status of the participants.

A summary of the databases is provided below.

Database 1(n=63): Twenty-two (7 men, 15 women) amateur runners took part in a three-day trail run in mild environmental conditions with *ad libitum* fluid intake [19]. Distances covered were 29.3 km in Stage 1(S1), 37.9 km in Stage 2 (S2) and 27.8 km in Stage 3 (S3) The range of ambient temperature and relative humidity over the three days was 11.5 - 22.8 ºC and 54 97 %, respectively. The main outcome measures were individual changes in serum osmolality (S_{osm}), serum sodium ($s[Na^+]$), plasma volume (PV), urine osmolality (U_{osm}), urine specific gravity (U_{sg}) and body mass (BM).

Database 2 (n=26): Thirteen well-trained male road cyclists completed two 90-minute trials in our laboratory at 60-65% of peak VO_2 in warm, humid (28.2 ± 0.9 º C; relative humidity:72.1 ± 3.3%) and moderately cool (18.3 ± 0.8 ºC), windy (4.0 ± 1.0 m/s) conditions [33]. *Ad libitum* fluid intake was recorded. Pre-post trial assessments included BM, S_{osm}, urine volume and U_{osm}.

Database 3 (n=54): The hydration status of amateur cyclists who drank *ad libitum* during a three-day, 248 km mountain bike race [32] was assessed in 18 amateur male cyclists. Daily stage length varied from 87km (S1) to 90km (S2) and 71km (S3). Temperature ranged from 6.0 - 21.4°C over the 3 race stages with the main outcome measures being stage-induced changes in BM, S_{osm}, $s[Na^+]$ and U_{sg}.

Database 4 (n= 8): The changes in BM, total body water (TBW), plasma osmolality (P_{osm}), plasma sodium ($p[Na^+]$), plasma potassium $[K^+]$, plasma protein concentrations [TP], running performance and *ad libitum* fluid intake in an ultra-marathon mountain race covering 80 km were measured on seven male and one female runner [30].

Database 5 (n=32): Changes in BM, TBW, P_{osm}, $p[Na^+]$), [TP] and *ad libitum* fluid intake were measured in athletes during 21.1km and 56 km foot races [31]. 21 participants (12 women; 9 men) completed the 21.1km event while 12 participants (3 women; 9 men) completed the 56km event.

Database 6 (n=18): The components of biological variation and the accuracy of potential markers in plasma, urine, saliva and BM for static and dynamic dehydration assessment in 18 (13 males, 5 females) healthy participants were evaluated [18]. The exercise comprised 3 to 5 h of work:rest cycles (50 min work:10 min rest) on a treadmill (1.56 m/s; 4–7% grade) or cycle ergometer (85–120 W) inside an environmental chamber set to 40°C and 20% relative humidity with a 1-m/s laminar wind flow and no fluid intake. The main outcome measures were P_{osm}, U_{osm}, saliva osmolality, urine colour and BM.

Database 7 (n=36): 30 males and 6 females performed work in a laboratory according to a similar design as that used in database 6, with only 90 minutes of rest was allowed after 3 hours of intermittent walk/rest in more severe heat (50°C) [34].

Data analyses
All variables were analysed using SPSS version 19 software (SPSS Inc., Chicago, Ilinois) to determine its skewness, kurtosis and normality using the Kolmogorov-Smirnov test. Central tendencies were appropriately presented as mean ± standard deviation (SD) or median (range). Bivariate correlation analyses were used to determine the relationship between the various physiological and environmental factors and post-exercise hydration status, with a P_{osm}/S_{osm} value in the range [275-295] $mmol.kg^{-1}$ being used as indicative of euhydration [35-37], while a value in excess of 295 $mmol.kg^{-1}$ indicated dehydration [18,35,38]. Statistical significance was accepted at the 0.05 level.

Data pre-processing

Gender and post-exercise hydration status were categorised as follows: female=0; male=1; euhydrated=0; dehydrated=1. All the other variables in the data set were normalised to lie in the range [0-1] by making use of the appropriate divisors to avoid rejection of those with smaller magnitudes by the learning algorithm of the ANN.

Furthermore, temperature, humidity and solar radiation were combined into a single environmental stress index (ESI) using the following equation [39]:

$$ESI = 0.63*T - 0.03*H + 0.002*SR + 0.0054*T*H - \frac{0.073}{0.1+SR} \qquad [1]$$

Where T is the temperature (°C), H is the humidity (%) and SR is the solar radiation (W.m^{-2})

In order to make an optimal selection of input variables to the ANN, a filtering method [40] was used. Based on the guidelines of Walczak & Cerpa [41], only statistically significant variables for which the correlation (r) with the post-exercise hydration status exceeded 0.2 were selected as possible inputs. Identification and removal of superfluous variables was undertaken using partial correlation analysis [40].

The composite data set (n=237) consisting of both euhydrated and dehydrated participants, was used to design, train, validate and test various classification models. Only the data from euhydrated participants (n=116) were then retained from the complete data set and used to design, validate and test various prediction models used to estimate the drinking rates of this subset of athletes. The data sets for classification and prediction models were each randomised in turn and subdivided into a training, validation and test subset, respectively. Thirty percent (30%) of the classification (n=71) and prediction (n=36) model data sets were reserved for testing the respective networks, with the remainder used for training and validation (Figures 1a and 1b).
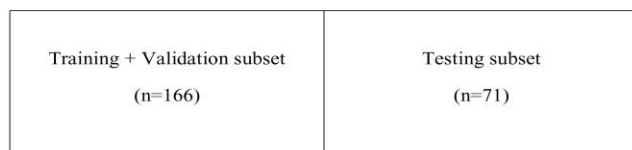
Training + Validation subset (n=166)	Testing subset (n=71)

Figure 1a

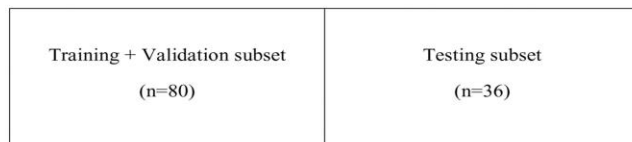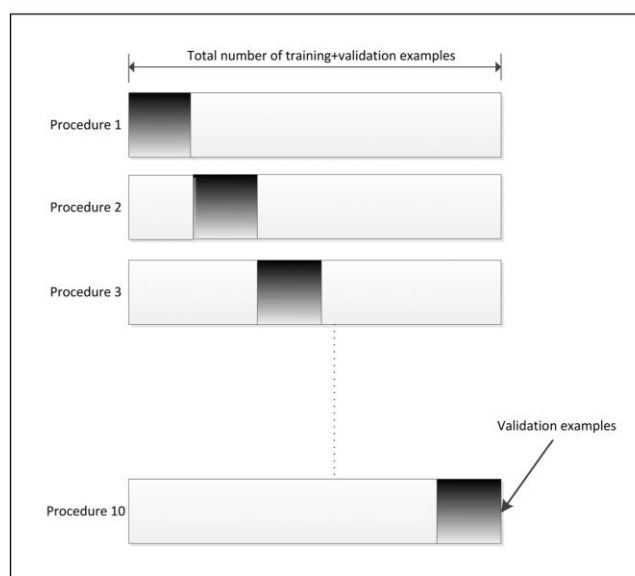| Training + Validation subset (n=80)	Testing subset (n=36)

Figure 1b

Figure 1c

Figure 1: Data split and cross validation methods
(a)   Data split for Artificial Neural Network classification models
(b)   Data split for Artificial Neural Network prediction models
(c)   10-fold cross-validation method

Model design
Randomisation of the classification and prediction model datasets as well as the designing, validation and testing of the ANN models were achieved using MATLAB (R2011b, The Mathworks, Natick, Massachusetts). All possible
.

combinations of the input variables were used to create classification and prediction models using feed forward multilayer perceptron (MLP) and radial basis function (RBF) ANNs with single hidden layers (Figures 2a and 2b)
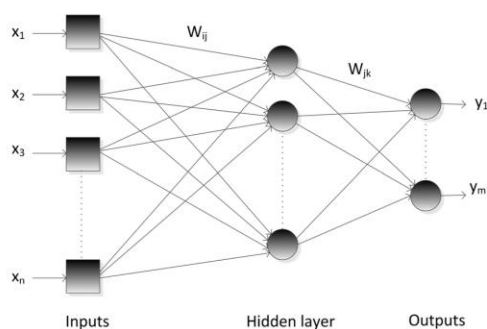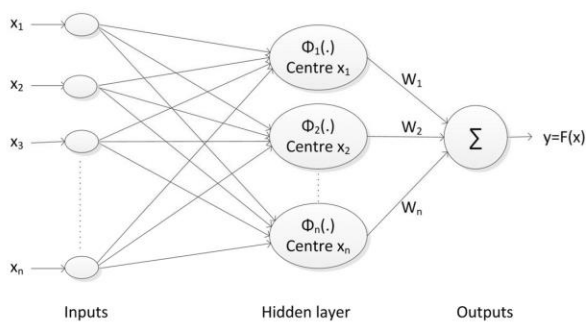
Figure 2a

Figure 2b

Figure 2: Artificial Neural Network feed forward models
(a) Multilayer perceptron (MLP) network
(b) Radial basis function (RBF) network

The MLP network with one hidden layer, incorporating either the logistic sigmoidal (logsig) or hyperbolic tangent sigmoidal (tansig) activation functions, was used as a universal approach element [42] and the output of this network was determined by using the following formula:

$$y_k = \left(\sum_{j=1}^{h} w_{jk} f \left(\sum_{i=1}^{n} w_{ij} x_i - b_j \right) - b_k \right) , \text{k=1,......m} \qquad [2]$$

Where f is the activation function (either logsig or tansig), h is the number of hidden layer neurons (limited to a maximum of 30 during training), w_{jk} and w_{ij} are the weights of the connections between hidden and output layer and between input and hidden layer, respectively, b is the polarisation values (biases) and x is the data vector. For a single output, k is set to 1.

For a logsig function as activation for the neurons in the hidden layer, f was given by:

$$f(z)_{\log sig} = \frac{1}{1 + \exp(-z)} \qquad [3]$$

Similarly, the formula for tansig function as activation for the hidden layer neurons was given by:

$$f(z)_{\tan sig} = \frac{\exp(z) - \exp(-z)}{\exp(z) + \exp(-z)} \qquad [4]$$

The non-linear output y_k was estimated using the optimisation method of Lavenberg-Marquardt [43]. This is a

standard method to minimise the mean square error (MSE), due to its properties of convergence and robustness and the decline method of Nguyen and Widrow[44] was used to initialise the weights of the network.

In order to accommodate the fact that an RBF network also has a feed forward structure consisting of a single hidden layer, with the activation function being mostly built up of Gaussian rather than sigmoid as in MLP networks, the output of the RBF network was given by:

$$y_k(x) = \sum_{j=1}^{M} w_{kj} \Phi_j(x, c_k) = \sum_{j=1}^{M} w_{kj} \Phi_j \left(\left\| x - c_k \right\|_2 \right) \qquad [5]$$

Where x is the input vector, w_{kj} are the weights in the output layer, M is the number of neurons in the hidden layer (limited to a maximum of 100 during training), c_k are the RBF centres in the input vector space, $\|.\|_2$ denotes the Euclidean norm and Φ_j is the Gaussian activation function, given by:

$$\Phi(r) = \frac{1}{\sqrt{r^2 + \sigma^2}} \qquad [6]$$

With σ^2 being the spread parameter (limited to a maximum of 10 during training)

To determine the error calculations used to train an ANN, training of the ANN and performance assessment was done using the following objective function:

$$E = \frac{1}{2N} \sum_{s=1}^{N} \left(\hat{y}(s) - y(s) \right)^2 \qquad [7]$$

Where N is the number of data samples used to train the ANN, y is the true output of the network and ŷ is the estimated output of the network.

In order to train the ANN based on as many examples as possible and obtain the best models, a 10-fold cross validation approach was used to develop the models. The training+validation subset (Figure 1c) was split into ten approximately equal portions, such that each portion was used in turn for validating the classifications/predictions of the ANN models in addition to adjusting the network parameters, while the remainder was used for training. For example, 9/10th of the training + validation subset was used for training and the remaining 1/10th for validation. This procedure was repeated 10 times. The training of the ANN was terminated when a satisfactory compromise was reached between minimisation of the training set error and the quality of the generalisation

of the validation data set. The model selected was the one that had the smallest average mean squared error on the validation data set (MSE_val).

Model performance assessment

The classification and prediction test data sets were used to assess the performance of the classification and prediction ANN models, respectively. Performance of the classification models in classifying the post-race hydration status of athletes was analysed using correlation (r) and sensitivity/specificity analyses, receiver operating characteristics curves (ROC) and the area under the ROC curves (AUC). Performance of the prediction models in being able to correctly predict the drinking rate of the euhydrated athletes were measured using coefficient of determination (R2), root mean squared error (RMSE), mean bias error (MBE) and coefficient of variation of the root mean squared error (CVRMSE). The outputs from the predictive networks were first de-normalised before comparing them with the actual measured data.

$$RMSE = \sqrt{\frac{1}{N} \sum_{s=1}^{N} \left(\hat{y}(s) - y(s) \right)^2} \qquad [8]$$

$$MBE = \frac{1}{N} \sum_{s=1}^{N} \left(\hat{y}(s) - y(s) \right) \qquad [9]$$

$$CV_{RMSE} = \frac{RMSE}{mean(\hat{y})} \qquad [10]$$

Results

The composite data base consisted of 237 individual data sets which were obtained from six smaller databases derived from

148 participants (106 males; 42 females) ranging in age from 18 to 56 years (Table 1). In 34% (n=80) of the data sets, the participants were either amateur or professional cyclists, while the remainder of the data was obtained from amateur runners. There was a wide variation in the anthropometric characteristics of the athletes as well as in the environmental conditions (ESI = 9.4 - 35.6).

Table 1: Environmental factors and physical characteristics of the participants comprising the total data set (n=237) and subset relying on ad libitum fluid replacement (n=183)

	Total data set (n = 237)			*Ad libitum* subset (n=183)		
	Mean (±SD)	Min	Max	Mean (±SD)	Min	Max
Age, y	34 ± 10	18	56	37 ± 8	18	56
Body fat, %	19.6 ± 4.0	8.2	30.6	18.8 ± 4.3	8.2	30.6
Mass, Kg	74.9 ± 12.9	49.1	109.8	72.6 ± 11.9	49.1	103.2
BMI	24.4 ± 3.2	18.7	35.4	23.7 ± 2.5	18.7	30.9
BSA*, m^2	1.9 ± 0.2	1.5	2.4	1.9 ± 0.2	1.5	2.3
Distance, km	45 ±25	17	90	53.2 ± 23.4	21.1	90
Exercise intensity**, $km.h^{-1}$	7.9 ± 2.8	4.3	16.0	8.6 ± 2.8	4.3	16
Duration, h	4.3 ± 2.2	1.5	12.8	4.5 ± 2.4	1.5	12.8
Temperature, °C	23.9 ± 13.0	12.3	50	17.2 ± 3.5	12.3	29.7
Humidity, %	62 ± 25	20	96	74.5 ± 10.8	44.4	95.8
Solar radiation, $W.m^{-2}$	834 ± 132	0	467	108.4 ± 141.4	0.0	467.2
Environmental stress index ***	19.4 ± 8.3	9.4	35.6	15.4 ± 3.6	9.4	27.2
Drinking rate, $L.h^{-1}$	0.404 ± 0.400	0.000	2.000	0.523 ± 0.381	0.874	2.000
Sweat rate****, $L.h^{-1}$	0.912 ± 0.433	0.185	3.067	0.877 ± 0.458	0.185	3.067
Ratio of Drinking / Sweat rate, %	47 ± 36	0	200	62 ± 38	10	327
Pre-race plasma osmolality, $mmol.kg^{-1}$	291 ± 6	276	316	291 ± 6	276	316
Post-race plasma osmolality, $mmol.kg^{-1}$	295 ± 8	273	316	293 ± 7	275	310

Note : Max: maximum, Min: minimum; BMI=body mass index; BSA=body surface area
 *computed using formula of Du Bois & Du Bois [57];
**cycling pace converted to an approximate running pace using a factor of 2.5[58,59];
 ***computed using formula of Moran & Epstein[39];
 ****estimated using the formula [(pre-mass - post mass) + fluids intake - urine voided]/exercise duration; assuming that 1g weight loss is equivalent to 1ml sweat loss

In 77% of the cases (n=183), the athletes were allowed *ad libitum* drinking, with fluid restriction employed in the remaining cases. Of the composite data set, 85% (n=201) of the subjects started the event with a plasma osmolality (P_{osm}) within the normal reference range for euhydration (275- 295 $mmol.kg^{-1}$), 49% (n=116) completed the events with P_{osm} in this reference range, while the remaining 51% (n=121) completed the events dehydrated ($P_{osm} \geq 296$ $mmol.kg^{-1}$). This provided a balanced set of data for training, validating and testing the ANN models. None of the subjects completed their event both overhydrated ($P_{osm} <275$ $mmol.kg^{-1}$) and hyponatraemic (plasma sodium < 134 $mmol.L^{-1}$).

The athletes displayed a wide variability in drinking and sweat rate with mean (±SD) drinking ($L.h^{-1}$) and sweat rates ($L.h^{-1}$) of 0.404 (±0.400) and 0.912 (±0.433), respectively (Table 1). In the group of athletes completing the race with P_{osm} in the euhdrated range (n=116), the mean (±SD) drinking rate ($L.h^{-1}$), sweat rate ($L.h^{-1}$) and drinking/sweat rate ratios (%) were 0.582 (±0.438), 0.944 (±0.518) and 66 (±44), respectively. Of the athletes that were allowed *ad libitum* fluid intake (n=183), 63% (n=116) of them finished the event euhydrated, with the remaining 37% (n=67) falling into the dehydrated category of which 94% (n=63) of them had taken part in either the multiday cycle or trail runs.

Table 2 presents the results of the bivariate correlational analyses conducted on the entire data set. From the entire set of variables listed (n=13), only pre-race

.

hydration status (PH), height (H) and exercise duration (DU) were found to be non-significant ($p > 0.05$) in determining the post-exercise hydration status

Table 2: Results of bivariate correlational analyses with post-exercise hydration status (n=237)

	Correlation (r)	Statistical significance
Age, y	-0.15	$p < 0.05$
Gender	0.24	$p < 0.01$
Body Fat, %	0.30	$p < 0.01$
Height, m	0.17	$p > 0.05$
Body mass, Kg	0.38	$p < 0.01$
BMI	0.40	$p < 0.01$
BSA, m^2	0.34	$p < 0.01$
Distance, km	-0.15	$p < 0.05$
Exercise Intensity, $km.h^{-1}$	-0.47	$p < 0.01$
Duration, h	0.10	$p > 0.05$
Environmental stress index	0.40	$p < 0.01$
Drinking rate, $L.h^{-1}$	-0.44	$p < 0.01$
Pre-race hydration status, $mmol.kg^{-1}$	0.12	$p > 0.05$

Note: BMI=body mass index; BSA=body surface area as computed using formula of Du Bois & Du Bois [57]

The variables exercise intensity (EI), environmental stress index (ESI), body mass (BM), gender (G) and drinking rate (FI) that were identified using the filtering method, allowed for 15 different input combinations to the ANN models (Table 3). As the purpose of the prediction models was to estimate the drinking rate of athletes, FI was removed as an input variable into these models (Table 4).

Table 3: Results for the ANN classification models (n=237)

Model no.	Variables	MSE$_{val}$ MLP	MSE$_{val}$ RBF	(r) MLP	(r) RBF	Sensitivity MLP	Sensitivity RBF	Specificity MLP	Specificity RBF	AUC MLP	AUC RBF	Best model Network	Best model Activation function**	Best model Structure***
C1	FI, ESI	0.11	0.15	0.43	0.49	0.71	0.80	0.71	0.74	0.71	0.74	MLP	tansig	2 : 11 : 1
C2	FI, EI	0.12	0.14	0.49	0.38	0.88	0.62	0.58	0.75	0.73	0.68	MLP	tansig	2 : 8 : 1
C3	FI, EI, ESI	0.11	0.12	0.51	0.52	0.60	0.74	0.89	0.78	0.74	0.76	RBF	Gaussian	3 : 40 : 1
C4	FI, G	0.12	0.13	0.59	0.57	0.91	0.91	0.66	0.63	0.79	0.77	MLP	logsig	2 : 3 : 1
C5	FI, ESI, G	0.12	0.14	0.61	0.60	0.88	0.80	0.72	0.80	0.80	0.80	MLP	logsig	3 : 29 : 1
C6	FI, EI, G	0.11	0.14	0.69	0.52	0.86	0.80	0.83	0.72	0.85	0.76	MLP	tansig	3 : 20 : 1
C7	FI, EI, ESI, G	0.11	0.13	0.74	0.66	0.85	0.85	0.88	0.80	0.87	0.83	MLP	tansig	4 : 30 : 1
C8	FI, BM	0.11	0.14	0.69	0.66	0.97	0.86	0.69	0.81	0.83	0.83	MLP	logsig	2 : 15 : 1
C9	FI, ESI, BM	0.11	0.15	0.64	0.58	0.77	0.83	0.86	0.75	0.82	0.79	MLP	tansig	3 : 18 : 1
C10	FI, EI, BM	0.11	0.14	0.68	0.61	0.94	0.80	0.72	0.81	0.83	0.80	MLP	tansig	3 : 11 : 1
C11	FI, EI, ESI, BM	0.11	0.13	0.75	0.54	0.89	0.89	0.86	0.64	0.87	0.76	MLP	logsig	4 : 16 : 1
C12	FI, G, BM	0.11	0.15	0.55	0.53	0.82	0.85	0.72	0.66	0.77	0.76	MLP	tansig	3 : 12 : 1
C13	FI, ESI, G, BM	0.12	0.16	0.65	0.61	0.91	0.83	0.72	0.77	0.82	0.80	MLP	logsig	4 : 17 : 1
C14	FI, EI, G, BM	0.10	0.13	0.69	0.71	0.83	0.86	0.86	0.86	0.84	0.85	RBF	Gaussian	4 : 30 : 1
C15	FI, ESI, EI, G, BM	0.09	0.14	0.78	0.55	0.83	0.77	0.94	0.78	0.89	0.77	MLP	tansig	5 : 19 : 1

Note: G=gender; BM=body mass; EI=exercise intensity; ESI=environmental stress index; FI=drinking rate; MLP=multi-layer perceptron; RBF=radial basis function; MSEval=average of mean squared error for all models computed on validation dataset; AUC=area under receiver operating curve; r=correlation coefficient.; tansig=hyperbolic tangent sigmoid; logsig=logistic sigmoid; *Best model selection is based on lowest MSEval, highest r and AUC; **This is the activation function for the neurons in the hidden layer; ***First number in the structure is the number of neurons in the input layer, second number is the number of neuron in the hidden layer, whilst the third number is the number of neurons in the output layer.

Table 4: Results for the ANN prediction models (n=116)

Model no.	variables	MSE_{val} $L.h^{-1}$ MLP	RBF	R^2 MLP	RBF	RMSE $L.h^{-1}$ MLP	RBF	MBE $L.h^{-1}$ MLP	RBF	CV_{RMSE} % MLP	RBF	Best model for input combination* Network	Activation function**	Structure***
P1	ESI	0.01	0.01	0.44	0.42	0.35	0.36	-0.08	-0.09	66.25	69.00	MLP	tansig	1 : 27 : 1
P2	EI	0.02	0.03	0.62	0.56	0.30	0.32	-0.07	-0.07	55.42	59.93	MLP	logsig	1 : 11 : 1
P3	EI, ESI	0.01	0.02	0.77	0.74	0.25	0.26	-0.06	-0.08	46.20	50.28	MLP	tansig	2 : 19 : 1
P4	G	0.04	0.04	0.28	0.28	0.42	0.41	-0.12	-0.08	86.53	77.80	RBF	Gaussian	1 : 60 : 1
P5	ESI, G	0.01	0.02	0.53	0.47	0.34	0.35	-0.12	-0.11	68.38	70.66	MLP	logsig	2 : 5 : 1
P6	EI, G	0.02	0.03	0.69	0.11	0.31	0.56	-0.08	0.07	57.31	82.96	MLP	logsig	2 : 5 : 1
P7	EI, ESI, G	0.01	0.01	0.76	0.78	0.25	0.24	-0.11	-0.08	50.41	46.31	RBF	Gaussian	3 : 13 : 1
P8	BM	0.03	0.09	0.03	0.06	0.46	0.45	-0.07	-0.06	84.85	82.11	MLP	tansig	1 : 22 : 1
P9	ESI, BM	0.01	0.04	0.33	0.37	0.39	0.38	-0.11	-0.10	77.75	74.70	MLP	tansig	2 : 29 : 1
P10	EI, BM	0.02	0.03	0.53	0.55	0.33	0.33	-0.06	-0.07	58.96	60.56	MLP	logsig	2 : 17 : 1
P11	EI, ESI, BM	0.01	0.03	0.66	0.69	0.27	0.28	-0.01	-0.08	44.00	52.10	MLP	logsig	3 : 25 : 1
P12	G, BM	0.03	0.04	0.42	0.51	0.38	0.38	-0.05	-0.10	67.36	74.21	MLP	tansig	2 : 5 : 1
P13	ESI, G, BM	0.01	0.02	0.40	0.50	0.37	0.34	-0.10	-0.11	72.86	68.89	RBF	Gaussian	3 : 60 : 1
P14	EI, G, BM	0.02	0.03	0.61	0.59	0.31	0.31	-0.09	-0.07	58.31	58.43	MLP	tansig	3 : 11 : 1
P15	ESI, EI, G, BM	0.01	0.02	0.80	0.71	0.24	0.26	-0.09	-0.06	42.20	47.17	MLP	tansig	4 : 10 : 1

Note: G=gender; BM=body mass; EI=exercise intensity; ESI=environmental stress index; MLP=multi-layer perceptron; RBF=radial basis function; MSE_{val}=average of mean squared error for all models computed on validation dataset; R^2=coefficient of determination; RMSE=root mean squared error; MBE=mean bias error; CV_{RMSE}=coefficient of variation on RMSE; L=litres; h=hours; * Best model selection is based on lowest MSE_{val}, lowest CV_{RMSE} and highest R^2;   ** This is the activation function for the neurons in the hidden layer; *** First number in the structure is the number of neurons in the input layer, the second number is the number of neuron in the hidden layer, whilst the third number is the number of neurons in the output layer.

The results for the ANN classifiers of the post-exercise hydration status of subjects (n=237) as either euhydrated (0) or dehydrated (1), are presented in Table 3. In 87% of the classification models (n=13), MLP networks were superior to the RBF networks, producing lower values of MSE_{val}, and higher correlations (r), sensitivities, specificities and AUC. In these MLP models, use of the tansig activation function for the neurons in the hidden layer gave better performance than the logsig function in 62% of the models (n=8). The best performing model was an MLP network with 5 neurons in the input layer, 19 tansig neurons in the hidden layer and 1 linear neuron in the output layer.   Taking as inputs FI, ESI, EI, G and BM, this model had the lowest MSE_{val} (0.09) and it resulted in the highest AUC (0.89) and correlation (r=0.78) between actual hydration status of the athletes in the test data set and the estimated hydration status generated by this model.

Table 4 provides the results for the ANN predictors used to estimate the drinking rate of the euhydrated subjects (n=116). The MLP models performed better than the RBF networks in 80% of the cases (n=12), by producing lower MSE_{val}, CV_{RMSE} and larger R^2. All the MLP prediction models underestimated the drinking rates of the athletes as can be seen from the negative values for MBE. Use of the tansig instead of logsig activation function for the hidden layer neurons gave better performance in these MLP models in 58% (n=7) of the cases. The input variables to the best performing model were ESI, EI, G and BM. This was an MLP network with 4 input neurons, 10 tansig neurons in the hidden layer and 1 linear neuron in the output layer. When comparing the fluid estimates generated by this model with the fluid intake of athletes in the test data set, in comparison to the other models, this had the highest R^2 (0.80), lowest RMSE (0.24 L.h^{-1}) and CV_{RMSE} (42.20%). Furthermore, the superior performance of the MLP network in comparison to RBF is evident in the performance graphs (Figure 4).
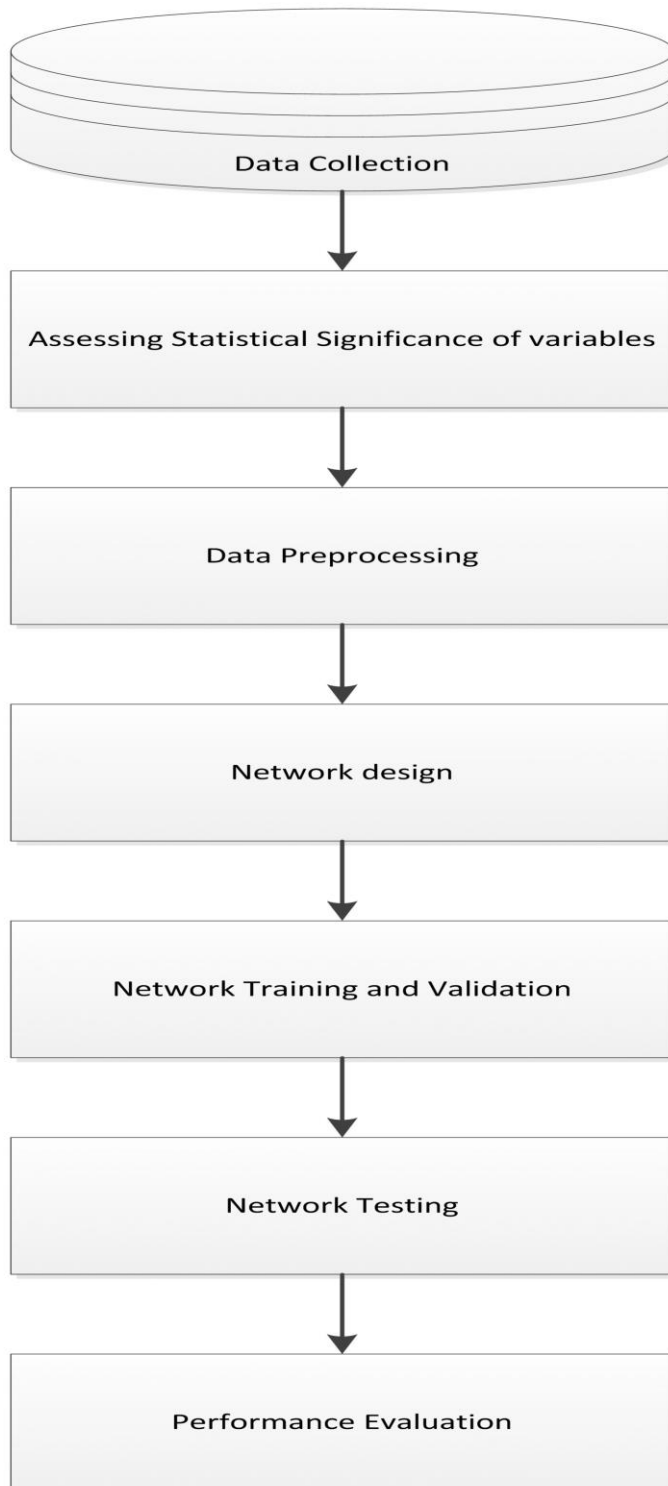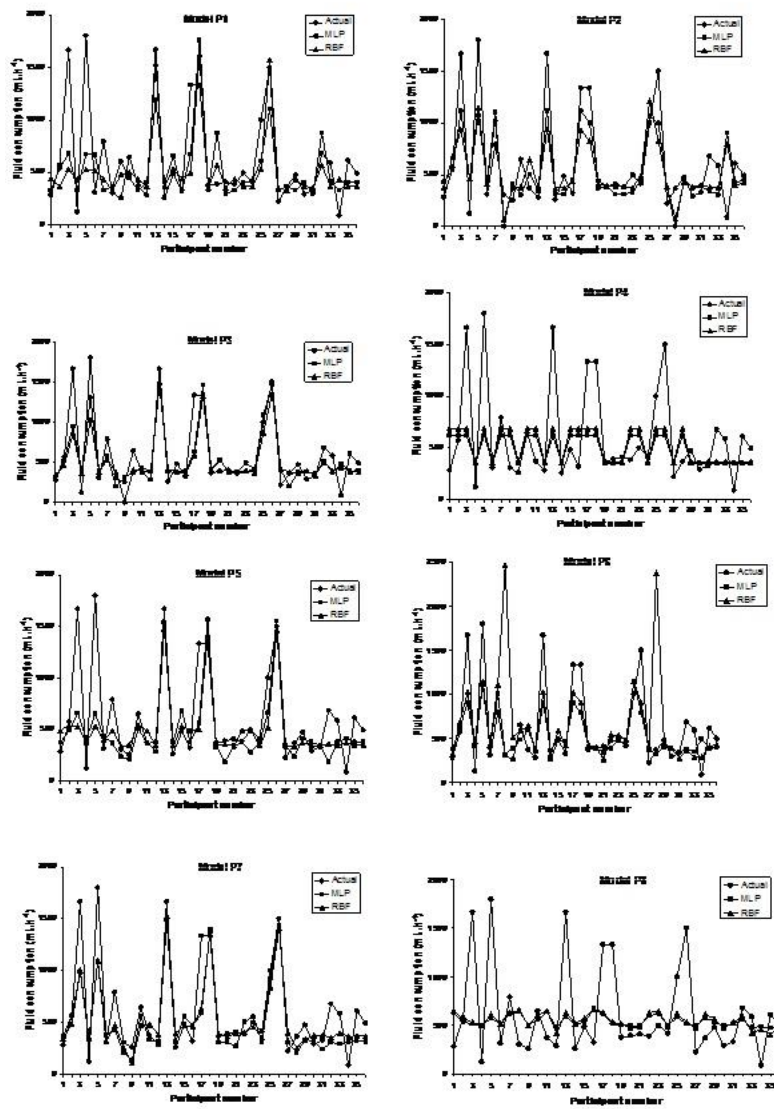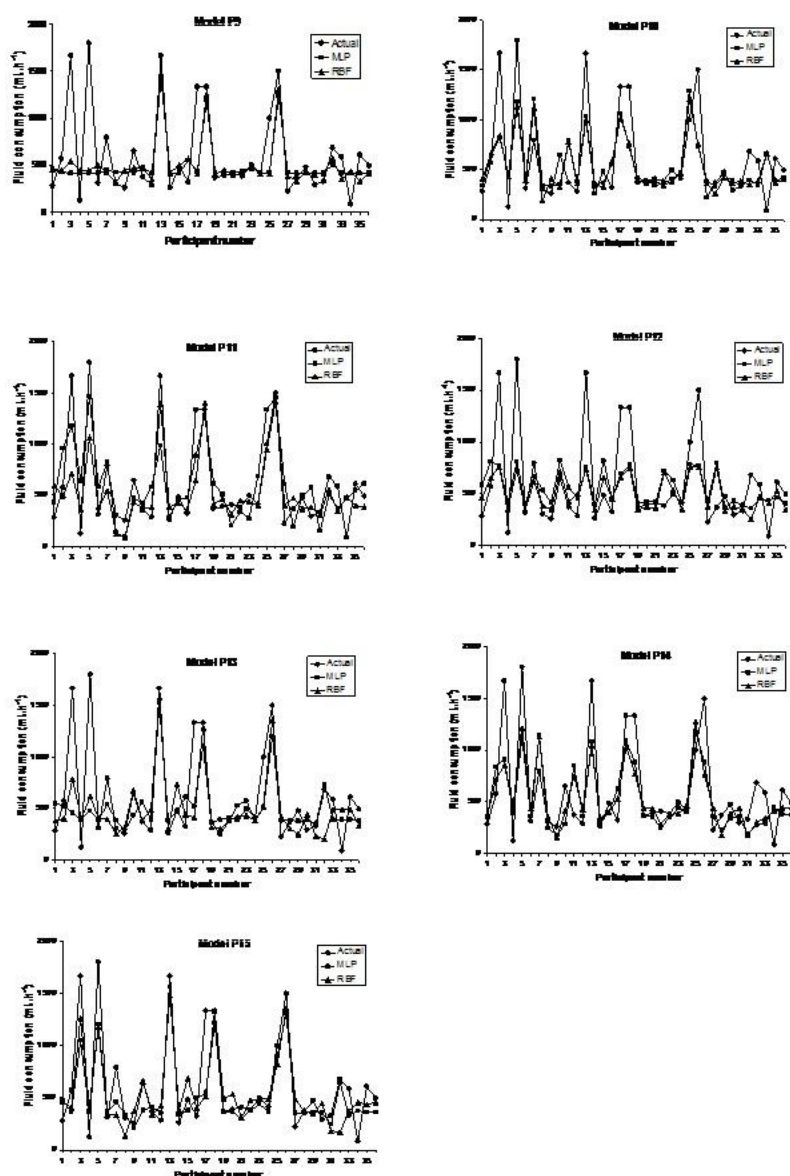
Figure 3: Flowchart of methodology

Figure 4A

Figure 4B
Figure 4A and 4B: Results of the Multilayer Perceptron and Radial Basis Function prediction models

Discussion

When athletes drink *ad libitum*, they have been shown to replace no more than 75% of their total water losses[17,45,46]. As the currently existing hydration models designed for athletes are based on complete replacement of the sweat output and total water losses are primarily made up of sweat when exercising in the heat, these existing models therefore provide an exaggerated estimate of fluid intake of athletes. Instead of estimating sweat rate alone, we used the complete set of physical, performance, training and environmental variables, to both classify the hydration status of athletes and predict their fluid intake using the ANN.

Although it may appear that there are several techniques other than ANNs which could be used in this application, including, but not limited to standard statistics such as regression analyses and expert systems, standard statistics would only have been viable had there been a model that already existed and to which a best fit had to be made. On the other hand, expert systems require the pre-existence of a clear set of criteria for the classification of hydration status and prediction of fluid

439   **Official Journal of FIMS (International Federation of Sports Medicine)**

requirements in athletes, which also do not yet exist. In view of the availability of sufficient training examples and no clearly defined relationship between the input variables and output, the ANN, with its ability to take into account the total interaction between the input variables, was therefore the preferred method for this particular application. As far as the authors are aware, this is the first report of the use of ANN modelling in the classification and prediction of the fluid intake requirements of endurance athletes.

The most important finding of this series of classification ANN models was that the optimal set of input variables which display high accuracy, include BM, EI, ESI, G and FI, while the optimal set of inputs variables with high predictive precision of FI are BM, EI, ESI and G. This was confirmed in the classification model, C15, which displayed an accuracy of 89% in being able to correctly identify the post exercise hydration status of the athletes that consumed fluids *ad libitum*, and the prediction model, P15, which produced a 90% correlation between the actual and predicted drinking rates of the athletes.

This first extensive comparative analysis of the 13 established variables that are known to affect fluid replacement needs supports 3 of the factors regarded as primary factors governing fluid loss during exercise identified by previous studies viz. body mass, exercise intensity and ambient temperature, [47,48]. However, the filtering algorithm applied to the input data set as well as the results of the ANN modelling technique identified gender as a fourth primary determinant of fluid intake needs in endurance athletes. Physiologically this could be attributed to the fact that women typically have lower sweating rates and electrolyte losses than men due to their smaller stature and lower metabolic rates when performing the same task as men[49,50]. As these findings were restricted to the number of data obtained from females, further work is required to verify these data.

The importance of gender as a variable is verified in Figure 4. For example, subject 15 in model P15, a male athlete of mass 67.2kg, running at 8.2 min.km^{-1} in environmental conditions resulting in an ESI of 13.44 (average ambient temperature 15.0 °C, average relative humidity 64.1%, average solar radiation 360.9 W.m^{-2}) had an average drinking rate of 0.48 L.h^{-1} and completed the race with a P_{osm} of 293 mmol.kg^{-1}. Model P15 predicted that his required fluid consumption to maintain euhydration was, however, only 0.38 L.h^{-1}. On the other hand, subject 31 in model P15, a female of mass 63.0kg, running at 9.9 min.km^{-1} in environmental conditions resulting in an ESI of 16.89 (average ambient temperature 17,9 °C, average relative humidity 76.7%, average solar radiation 233.5 W.m^{-2}) had an average drinking rate of 0.33 L.h^{-1} and completed the race with a P_{osm} of 294 mmol.kg^{-1}. According to model P15, although running in a higher ESI, her necessary fluid consumption to avoid a state of clinical dehydration, was only 0.25 L.h^{-1}.

The data sets also confirm that the differences in weather conditions, shape, size and performance of these athletes, result in a wide variability in their sweat rates and fluid intake. The clinically euhydrated subset of participants replenished on average 66 (±44)% of their fluid lost to sweating, confirming previous findings on *ad libitum* drinkers [17,45,46]. Although the sweat rate and fluid loss is related to the metabolic rate [51], the rate of fluid ingestion is regulated by the osmotically driven thirst centre in the hypothalamus [52]. The large variability in the degree to which participants allowed *ad libitum* fluid intake replaced their sweat losses during exercise, however points to marked differences in physiologic response to changes in S_{osm} between individuals [15,53] which may not only be limited to age [54], pregnancy [55] or presence of diabetes [56].

Of additional interest was the non-significant effect of pre-hydration status and exercise duration in determining post-exercise hydration status. While this may conflict with former conventional theories [1,6,10], it does support the hypothesis of Noakes [15] in which he predicts that *ad libitum* ingestion of fluid will compensate for low pre-exercise fluid status and be adjusted according to the duration of exercise.

Conclusion

Although sample size could be regarded as a possible limitation of this study, the generally accepted criteria regarding input data for each input variable to the network, is exceeded within each data set 21. The limited range of the input variables which did not include extreme environmental conditions or wide range of clothing ensembles, must however be acknowledged as a limitation of this first work exploring the uses of ANNs in the determination of the fluid intake needs of endurance athletes.

As the possibility always exists that ad libitum fluid replacement can be biased according to prior beliefs and misconceptions which athletes may have obtained, the findings of this initial study indicate that the static artificial neural network modelling technique may be valuable in providing accurate estimates of fluid intake which will maintain plasma osmolality within the 275-295 mmol.kg-1 range. These may serve as a pre-event guideline to athletes not wanting to rely solely on their dynamic thirst–induced biological neural network and can play an important role in countering the possibility of overhydration during endurance events.. It can therefore be concluded that artificial neural network modelling which can be used in conjunction with ad libitum fluid replacement has merit and can be refined further using different model architectures as well as data sets in which the input variables span a wider range.

Application

Using MATLAB (R2011b, The Mathworks, Natick, Massachusetts), model P15 was applied to 2 hypothetical cases not part of the test data set.

A 50kg female running at a speed of 8 km.h-1 in temperate environmental conditions (an ambient temperature of 25°C, relative humidity of 70% and zero solar radiation), the model predicted that the fluid intake to maintain euhydration was 0.45 L.h-1.

A 65kg male running at 12 km.h-1 in the same environmental conditions, the model predicted that the fluid intake to maintain euhydration was 0.76 L.h-1.

Key points

The advantages of the artificial neural network modelling technique over standard statistics and expert systems lies in its ability to formulate a model that takes into account the total interaction between the most important input variables.

Body mass, gender, exercise intensity and environmental stress index were identified as the primary variables for the prediction of fluid intake in endurance athletes.

Additional inclusion of fluid intake allowed for the accurate classification of post-exercise hydration status of endurance athletes.

The artificial neural network modelling technique provides a more accurate method of predicting the fluid intakes of endurance athletes as well as classifying their hydration status, than existing models. It has merit in this field and warrants further investigation.

Address for correspondence:

Dr Navin R Singh, Department of Electronic Engineering, Faculty of Engineering, Science and the Built Environment, Durban University of Technology, Durban, South Africa
Tel: +27313732118; Fax: +27313732744
Email: navins@dut.ac.za

References

1. Gonzalez-Alonso J, Mora-Rodriguez R, et al. Dehydration markedly impairs cardiovascular function in hyperthermic endurance athletes during exercise. J Appl Physiol 1997;82(4):1229-1236.
2. Cheuvront SN, Montain SJ, Sawka MN. Fluid replacement and

performance during the marathon. Sports Med 2007;37(4-5):353-357.

3. Sawka MN, Burke LM, Eichner ER, et al. American College of Sports Medicine position stand. Exercise and fluid replacement. Med Sci Sports Exerc 2007;39(2):377-390.

4. Noakes TD. Overconsumption of fluids by athletes. BMJ 2003 2003;327:113-114.

5. Noakes TD, Sharwood K, Speedy DB, et al. Three independent biological mechanisms cause exercise-associated hyponatremia: evidence from 2,135 weighted competitive athletic performances. Proc Natl Acad Sci USA 2005;102:18550-18555.

6. Sawka MN, Wenger CB, Pandolf KB. Thermoregulatory responses to acute exercise- heat stress and heat acclimation. In: Society OUPftAP, ed. Environmental Physiology. New York, 1996;157-186.

7. Gonzalez RR, Cheuvront SN, Montain SJ, et al. Expanded prediction equations of human sweat loss and water needs. J Appl Physiol 2009;107(2):379-88.

8. Hew-Butler T, Verbalis JG, Noakes TD. Updated fluid recommendation: Position statement from the international marathon medical directors association (IMMDA). Clin J Sport Med 2006;16(4):283-291.

9. Institute of Medicine. Water. In: Dietary references intakes for water, sodium, chloride, potassium and sulfate. Washington, D.C: National Academies Press, 2005;73-185.

10. Shapiro Y, Pandolf K, Goldman R. Predicting sweat loss response to exercise, environment and clothing. European Journal of Applied Physiology and Occupational Physiology 1982;48(1):83-96.

11. Shapiro Y, Moran D, Epstein Y, et al. Validation and adjustment of the mathematical prediction model for human sweat rate responses to outdoor environmental conditions. Ergonomics 1995;38(5):981-986.

12. Cheuvront SN, Montain SJ, Goodman DA, et al. Evaluation of the limits to accurate sweat loss prediction during prolonged exercise. Eur J Appl Physiol 2007;101(2):215-24.

13. Rollins D, Bhandar N, Hulting S. System identification of the human thermoregulatory system using continuous-time block-oriented predictive modeling. Chemical Engineering Science 2006;61:1516-1527.

14. Nolte HW, Noakes TD, Van Vuuren B. Trained humans can exercise safely in extreme dry heat when drinking water ad libitum. J Sports Sci 2011b;29(12):1233-1241.

15. Noakes TD. Waterlogged - The serious problem of overhydration in endurance sports. Human Kinetics, Champaign, 2012.

16. Sawka MN, Noakes TD. Does dehydration impair exercise performance? Med Sci Sports Exerc 2007;39(8):1209-17.

17. Cheuvront SN, Haymes EM. Ad libitum fluid intakes and thermoregulatory responses of female runners in three environments. J Sports Sci 2001;19:845-854.

18. Cheuvront SN, Ely BR, Kenefick RW, et al. Biological variation and diagnostic accuracy of dehydration assessment markers. Am J Clin Nutr 2010;92(3):565-73.

19. Singh N, Peters E. Markers of hydration status in a three-day trail running event. Clin J Sport Med (in print) 2013.

20. McCulloch W, Pitts W. A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biophysics 1943;7:115-133.

21. Rumelhart D, McClelland J. Parallel distributed processing. Vol. 2. Cambridge, MA.: MIT Press, 1986.

22. Agatonvic-Kustrin S, Beresford R. Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. J Pharm Biomed Anal 2000;22(5):717-727.

23. Jayalakshmi T, Santhakumaran A. Statistical normalization and back propagation for classification. International Journal of Computer Theory and Engineering 2011;3(1):89-93.

24. Hoseini S, Soltani A, Pourahmadi-Nakhli M. Application of artificial neural network (ANN) in estimation of body mass index (BMI) based on the connection between environmental factors and physical activity. IJAIA 2012;3(4):107-115.

25. McCullagh J. Data mining in sport: a neural network approach. International Journal of Sports Science and Engineering 2010;4:131-138.

26. Schollhorn W. Applications of artificial neural nets in clinical biomechanics. Clinical Biomechanics 2004;19(9):876-898.

27. Bartlett R. Artificial intelligence in sports biomechanics: new dawn or false hope? Journal of Sports Science and Medicine 2006;5:474-479.

28. Silva A, Costa A, Oliveira P, et al. The use of neural network technology to model swimming performance. Journal of Sports Science and Medicine 2007;6:117-125.

29. Memmert D, Perl J. Game creativity analysis using neural networks. Journal of Sports Sciences 2009;27(2):139-149.

30. Tam N, Hew-Butler T, Papadopoulou E, et al. Fluid intake and changes in blood chemistry, running speed and body mass during an 80km mountain trail race. Medicina Sportiva 2009;13:108-115.

31. Tam N, Nolte H, Noakes TD. Changes in Total Body Water Content During Running Races of 21.1 km and 56 km in Athletes Drinking Ad libitum. Clin J Sport Med 2011;21(3):218-225.

32. Rose S, Peters E. Ad libitum adjustments to fluid intake in cool environmental conditions maintain hydration status in a three-day mountain bike race. Br J Sports Med 2010;44(6):430-436.

33. Pillai P. Non-osmotic regulation of hydration balance during endurance cycling in cool environmental conditions. University of KwaZulu-Natal, 2009.

34. Cheuvront SN, Kenefick RW, Sollanek K, et al. Water deficit equation: systematic analysis and improvement. Am J Clin Nutr 2013; 97:79-85.

35. Fischbach F. A manual of laboratory and diagnostic tests. 4th ed. Philadelphia, PA: Lippincott Company, 1992.

36. Miller R, Eriksson L, Fleisher L, et al. Miller's anesthesia. 7th ed. Philadelphia: Churchill Livingstone, 2010.

37. Boyd A, Modi S, Howard S, et al. Adverse reactions to voriconazole. Clinical Infectious Diseases 2004;39:1241-1244.

38. Senay LC. Effects of exercise in the heat on body fluid distribution Med Sci Sports 1979a;11:42-48.

39. Moran D, Epstein Y. Evaluation of the environmental stress index (ESI) for hot/dry and hot/wet climates. Industrial Health 2006;44:399-403.

40. Guyton I, Elisseeff A. An introduction to variable and feature selection. The Journal of Machine Learning Research 2003;3:1157-1182.

41. Walczak S, Cerpa N. Heuristic principles for the design of artificial neural networks. Information and Software Technology 1999;41(2):109-119.

42. Hornik K, Stinchcombe M, White H. Multilayer feedforward networks are universal approximators. Neural Networks 1989;2(5):359-366.

43. Marquardt D. An algorithm for least-squared estimation of nonlinear parameters. SIAM Journal on Applied Mathematics 1963;11:164-168.

44. Nguyen D, Widrow B. Improving the learning speed of a 2-layer neural network by choosing initial values of the adaptive weights. International Joint Conference on Neural Networks. San Diego, USA, 1990;2063-2068.

45. Greenleaf JE. Problem: thirst, drinking behaviour, and involuntary dehydration. Med Sci Sports Exerc 1992;24:645-656.

46. Maresh CM, Gabaree-Boulant CL, Armstrong LE, et al. Effect of hydration status on thirst, drinking, and related hormonal responses during low-intensity exercise in the heat. J Appl Physiol 2004;97(1):39-44.

47. Cheuvront SN, Haymes EM, Sawka MN. Comparison of sweat loss estimates for women during prolonged high-intensity running. Med Sci Sports Exerc 2002;34(8):1344-50.

48. Barr SI, Costill DL. Water: can the endurance athlete get too much of a good thing? J Am Diet Assoc 1989;89:1629-1632.

49. Shapiro Y, Pandolf KB, Avellini B, et al. Physiological responses of men and women to humid and dry heat. J Appl Physiol 1980;49:1-8.

50. Sawka MN, Toner MM, Francesconi RP, et al. Hypohydration and exercise: effects of heat acclimation, gender, and environment. J Appl Physiol 1983b;55(4):1147-53.
51. Rehrer N. Fluid and electrolyte balance in ultra-endurance sport. Sports Med 2001;31:701-715.
52. Verbalis JG. How does the brain sense osmolality? J Am Soc Nephrol 2007;18:3056-3059.
53. Thompson C, Edwards C, Baylis P. Osmotic and non-osmotic regulation of thirst and vasopressin secretion in patients with compulsive water drinking. Clinical Endocrinology 1991;35(3):221-228.
54. Phillips P, Bretterton M, Johnston C, et al. Reduced osmotic thirst in healthy elderly men. Am J Physiol 1991;261:R166-171.
55. Lindheimer M, Davison J. Osmoregulation, the secretion of arginine vasopressin and its metabolism during pregnancy. Eur J Endocrinol 1995;132:133.
56. Thompson C, Davis S, Baylis P. Effect of blood glucose concentration on osmoregulation in diabetes mellitus. AJP - Regu Physiol 1989;256(3):R597-R604.
57. Du Bois D, Du Bois E. The measurement of the surface area of man. Arch Intern Med 1916;17:863.
58. Hall C, Figueroa A, Fernhall B, et al. Energy expenditure of walking and running: comparison with prediction equations. Med Sci Sports Exerc 2004;36(12):2128-2134.
59. Wilson DG, Papadopoulos J. Bicycling Science 3e Mit Press, 2004.