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

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.
Fluid intake needs of endurance athletes

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

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. 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. It however emphasises the importance of individualised fluid and electrolyte replacement schedules for athletes. 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. These include mathematical models that were developed to determine the sweat rate of athletes 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. Engineering models 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, it is well accepted that individuals should avoid drinking more fluid than the amount needed to replace their sweat losses, during prolonged exercise with blood osmolality being accepted as the best haematological marker of hydration status.

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, have found widespread use in the fields of medicine and sport, 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.

Introduction

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

Official Journal of FIMS (International Federation of Sports Medicine)
Methods

Data collection

Following approval by the relevant institutional research ethics committee, raw data were obtained from 4 separate field studies 18,32-34 and 3 separate laboratory studies 19,33,34 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 cyclists 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 °C and 54 97%, respectively. The main outcome measures were individual changes in serum osmolality (Sosm), serum sodium (s[Na+]), plasma volume (PV), urine osmolality (Uosm), urine specific gravity (Usg) 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 VO2 in warm, humid (28.2 ± 0.9 °C; relative humidity:72.1 ± 3.3%) and moderately cool (18.3 ± 0.8 °C), windy (4.0 ± 1.0 m/s) conditions 33. Ad libitum fluid intake was recorded. Pre-post trial assessments included BM, Sosm, urine volume and Uosm.

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, Sosm, s[Na+] and Usg.

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

Database 5 (n=32): Changes in BM, TBW, Posm, 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 Posm, Uosm, 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, Illinois) 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 Pposm/Sosm 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:

\[ ESI = 0.63 \times T - 0.03 \times H + 0.002 \times SR + 0.0054 \times T \times H - \frac{0.073}{0.1 + SR} \]  

[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 was used. Based on the guidelines of Walczak & Cerpa, 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.

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).
Fluid intake needs of endurance athletes


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 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 and the output of this network was determined by using the following formula:

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

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:

\[ f(z)_{\text{logsig}} = \frac{1}{1 + \exp(-z)} \tag{3} \]

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

\[ f(z)_{\text{tansig}} = \frac{\exp(z) - \exp(-z)}{\exp(z) + \exp(-z)} \tag{4} \]

The non-linear output \( y_k \) was estimated using the optimisation method of Lavenberg-Marquardt. 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 was used to initialise the weights of the network.

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

[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, \( \| \cdot \|_2 \) denotes the Euclidean norm and \( \Phi \) is the Gaussian activation function, given by:

\[ \Phi(r) = \frac{1}{\sqrt{\pi \sigma^2}} e^{-\frac{r^2}{\sigma^2}} \]  

[6]

With \( \sigma^2 \) 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:

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

[7]

Where \( N \) is the number of data samples used to train the ANN, \( y \) is the true output of the network and \( \hat{y} \) 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 (MSEval).

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 denormalised before comparing them with the actual measured data.

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

[8]

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

[9]

\[ \text{CVRMSE} = \frac{\text{RMSE}}{\text{mean}(\hat{y})} \]  

[10]

Results

The composite data base consisted of 237 individual data sets which were obtained from six smaller databases derived from...
Fluid intake needs of endurance athletes


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)**

<table>
<thead>
<tr>
<th></th>
<th>Total data set (n = 237)</th>
<th>Ad libitum subset (n=183)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (±SD)</td>
<td>Min</td>
</tr>
<tr>
<td>Age, y</td>
<td>34 ± 10</td>
<td>18</td>
</tr>
<tr>
<td>Body fat, %</td>
<td>19.6 ± 4.0</td>
<td>8.2</td>
</tr>
<tr>
<td>Mass, Kg</td>
<td>74.9 ± 12.9</td>
<td>49.1</td>
</tr>
<tr>
<td>BMI</td>
<td>24.4 ± 3.2</td>
<td>18.7</td>
</tr>
<tr>
<td>BSA*, m²</td>
<td>1.9 ± 0.2</td>
<td>1.5</td>
</tr>
<tr>
<td>Distance, km</td>
<td>45 ±25</td>
<td>17</td>
</tr>
<tr>
<td>Exercise intensity**</td>
<td>7.9 ± 2.8</td>
<td>4.3</td>
</tr>
<tr>
<td>Duration, h</td>
<td>4.3 ± 2.2</td>
<td>1.5</td>
</tr>
<tr>
<td>Temperature, °C</td>
<td>23.9 ± 13.0</td>
<td>12.3</td>
</tr>
<tr>
<td>Humidity, %</td>
<td>62 ± 25</td>
<td>20</td>
</tr>
<tr>
<td>Solar radiation, W.m⁻²</td>
<td>834 ± 132</td>
<td>0</td>
</tr>
<tr>
<td>Environmental stress index ***</td>
<td>19.4 ± 8.3</td>
<td>9.4</td>
</tr>
<tr>
<td>Drinking rate, L.h⁻¹</td>
<td>0.404 ± 0.403</td>
<td>0.000</td>
</tr>
<tr>
<td>Sweat rate****, L.h⁻¹</td>
<td>0.912 ± 0.433</td>
<td>0.185</td>
</tr>
<tr>
<td>Ratio of Drinking / Sweat rate, %</td>
<td>47 ± 96</td>
<td>0</td>
</tr>
<tr>
<td>Pre-race plasma osmolality, mmol.kg⁻¹</td>
<td>291 ± 6</td>
<td>276</td>
</tr>
<tr>
<td>Post-race plasma osmolality, mmol.kg⁻¹</td>
<td>295 ± 8</td>
<td>273</td>
</tr>
</tbody>
</table>

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 (Pₚₒsm) within the normal reference range for euhydration (275-295 mmol.kg⁻¹), 49% (n=116) completed the events with Pₚₒsm in this reference range, while the remaining 51% (n=121) completed the events dehydrated (Pₚₒsm ≥ 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 (Pₚₒsm < 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 Pₚₒsm in the euhydrated 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.

432

Official Journal of FIMS (International Federation of Sports Medicine)
Fluid intake needs of endurance athletes


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)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Correlation (r)</th>
<th>Statistical significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, y</td>
<td>-0.15</td>
<td>p&lt;0.05</td>
</tr>
<tr>
<td>Gender</td>
<td>0.24</td>
<td>p&lt;0.01</td>
</tr>
<tr>
<td>Body Fat, %</td>
<td>0.30</td>
<td>p&lt;0.01</td>
</tr>
<tr>
<td>Height, m</td>
<td>0.17</td>
<td>p&gt;0.05</td>
</tr>
<tr>
<td>Body mass, Kg</td>
<td>0.38</td>
<td>p&lt;0.01</td>
</tr>
<tr>
<td>BMI</td>
<td>0.40</td>
<td>p&lt;0.01</td>
</tr>
<tr>
<td>BSA, m²</td>
<td>0.34</td>
<td>p&lt;0.01</td>
</tr>
<tr>
<td>Distance, km</td>
<td>-0.15</td>
<td>p&lt;0.05</td>
</tr>
<tr>
<td>Exercise Intensity, km.h⁻¹</td>
<td>-0.47</td>
<td>p&lt;0.01</td>
</tr>
<tr>
<td>Duration, h</td>
<td>0.10</td>
<td>p&gt;0.05</td>
</tr>
<tr>
<td>Environmental stress index</td>
<td>0.40</td>
<td>p&lt;0.01</td>
</tr>
<tr>
<td>Drinking rate, L.h⁻¹</td>
<td>-0.44</td>
<td>p&lt;0.01</td>
</tr>
<tr>
<td>Pre-race hydration status, mmol.kg⁻¹</td>
<td>0.12</td>
<td>p&gt;0.05</td>
</tr>
</tbody>
</table>

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

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)

<table>
<thead>
<tr>
<th>Model no.</th>
<th>Variables</th>
<th>MSEval</th>
<th>(r)</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>AUC</th>
<th>Best model for input combination*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>MLP</td>
<td>RBF</td>
<td>MLP</td>
<td>RBF</td>
<td>MLP</td>
<td>Network</td>
</tr>
<tr>
<td>C1</td>
<td>FI, ESI</td>
<td>0.11</td>
<td>0.15</td>
<td>0.43</td>
<td>0.49</td>
<td>0.71</td>
<td>MLP</td>
</tr>
<tr>
<td>C2</td>
<td>FI, EI</td>
<td>0.12</td>
<td>0.14</td>
<td>0.49</td>
<td>0.38</td>
<td>0.88</td>
<td>MLP</td>
</tr>
<tr>
<td>C3</td>
<td>FI, EI, ESI</td>
<td>0.11</td>
<td>0.12</td>
<td>0.51</td>
<td>0.52</td>
<td>0.60</td>
<td>MLP</td>
</tr>
<tr>
<td>C4</td>
<td>FI, G</td>
<td>0.12</td>
<td>0.13</td>
<td>0.59</td>
<td>0.57</td>
<td>0.91</td>
<td>MLP</td>
</tr>
<tr>
<td>C5</td>
<td>FI, ESI, G</td>
<td>0.12</td>
<td>0.14</td>
<td>0.61</td>
<td>0.60</td>
<td>0.88</td>
<td>MLP</td>
</tr>
<tr>
<td>C6</td>
<td>FI, EI, G</td>
<td>0.11</td>
<td>0.14</td>
<td>0.69</td>
<td>0.52</td>
<td>0.86</td>
<td>MLP</td>
</tr>
<tr>
<td>C7</td>
<td>FI, EI, ESI, G</td>
<td>0.11</td>
<td>0.13</td>
<td>0.74</td>
<td>0.66</td>
<td>0.85</td>
<td>MLP</td>
</tr>
<tr>
<td>C8</td>
<td>FI, BM</td>
<td>0.11</td>
<td>0.14</td>
<td>0.69</td>
<td>0.66</td>
<td>0.97</td>
<td>MLP</td>
</tr>
<tr>
<td>C9</td>
<td>FI, ESI, BM</td>
<td>0.11</td>
<td>0.15</td>
<td>0.64</td>
<td>0.58</td>
<td>0.77</td>
<td>MLP</td>
</tr>
<tr>
<td>C10</td>
<td>FI, EI, BM</td>
<td>0.11</td>
<td>0.14</td>
<td>0.68</td>
<td>0.61</td>
<td>0.94</td>
<td>MLP</td>
</tr>
<tr>
<td>C11</td>
<td>FI, EI, ESI, BM</td>
<td>0.11</td>
<td>0.13</td>
<td>0.75</td>
<td>0.54</td>
<td>0.89</td>
<td>MLP</td>
</tr>
<tr>
<td>C12</td>
<td>FI, G, BM</td>
<td>0.11</td>
<td>0.15</td>
<td>0.55</td>
<td>0.53</td>
<td>0.82</td>
<td>MLP</td>
</tr>
<tr>
<td>C13</td>
<td>FI, ESI, G, BM</td>
<td>0.12</td>
<td>0.16</td>
<td>0.65</td>
<td>0.61</td>
<td>0.91</td>
<td>MLP</td>
</tr>
<tr>
<td>C14</td>
<td>FI, EI, G, BM</td>
<td>0.10</td>
<td>0.13</td>
<td>0.69</td>
<td>0.71</td>
<td>0.83</td>
<td>MLP</td>
</tr>
<tr>
<td>C15</td>
<td>FI, ESI, EI, G, BM</td>
<td>0.09</td>
<td>0.14</td>
<td>0.78</td>
<td>0.55</td>
<td>0.83</td>
<td>MLP</td>
</tr>
</tbody>
</table>

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)

<table>
<thead>
<tr>
<th>Model no.</th>
<th>variables</th>
<th>( \text{MSE}_{\text{val}} ) L.h(^{-1} )</th>
<th>( R^2 ) L.h(^{-1} )</th>
<th>( \text{RMSE} ) L.h(^{-1} )</th>
<th>( \text{MBE} ) L.h(^{-1} )</th>
<th>( \text{CV}_{\text{RMSE}} ) %</th>
<th>Best model for input combination*</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>ESI</td>
<td>0.01</td>
<td>0.44</td>
<td>0.35</td>
<td>-0.08</td>
<td>66.25</td>
<td>MLP</td>
</tr>
<tr>
<td>P2</td>
<td>E</td>
<td>0.02</td>
<td>0.62</td>
<td>0.30</td>
<td>-0.07</td>
<td>55.42</td>
<td>MLP</td>
</tr>
<tr>
<td>P3</td>
<td>E, ESI</td>
<td>0.01</td>
<td>0.77</td>
<td>0.25</td>
<td>-0.06</td>
<td>46.20</td>
<td>MLP</td>
</tr>
<tr>
<td>P4</td>
<td>G</td>
<td>0.04</td>
<td>0.28</td>
<td>0.42</td>
<td>-0.12</td>
<td>86.53</td>
<td>RBF</td>
</tr>
<tr>
<td>P5</td>
<td>ESI, G</td>
<td>0.01</td>
<td>0.47</td>
<td>0.34</td>
<td>-0.12</td>
<td>68.38</td>
<td>RBF</td>
</tr>
<tr>
<td>P6</td>
<td>E, G</td>
<td>0.02</td>
<td>0.69</td>
<td>0.31</td>
<td>-0.08</td>
<td>57.31</td>
<td>MLP</td>
</tr>
<tr>
<td>P7</td>
<td>E, ESI, G</td>
<td>0.01</td>
<td>0.76</td>
<td>0.25</td>
<td>-0.11</td>
<td>50.41</td>
<td>RBF</td>
</tr>
<tr>
<td>P8</td>
<td>BM</td>
<td>0.03</td>
<td>0.06</td>
<td>0.46</td>
<td>-0.07</td>
<td>84.85</td>
<td>MLP</td>
</tr>
<tr>
<td>P9</td>
<td>ESI, BM</td>
<td>0.01</td>
<td>0.33</td>
<td>0.39</td>
<td>-0.11</td>
<td>77.75</td>
<td>MLP</td>
</tr>
<tr>
<td>P10</td>
<td>E, BM</td>
<td>0.02</td>
<td>0.53</td>
<td>0.33</td>
<td>-0.06</td>
<td>58.96</td>
<td>MLP</td>
</tr>
<tr>
<td>P11</td>
<td>E, ESI, BM</td>
<td>0.01</td>
<td>0.66</td>
<td>0.27</td>
<td>-0.01</td>
<td>44.00</td>
<td>MLP</td>
</tr>
<tr>
<td>P12</td>
<td>G, BM</td>
<td>0.03</td>
<td>0.51</td>
<td>0.38</td>
<td>-0.05</td>
<td>67.36</td>
<td>MLP</td>
</tr>
<tr>
<td>P13</td>
<td>ESI, G, BM</td>
<td>0.01</td>
<td>0.42</td>
<td>0.37</td>
<td>-0.10</td>
<td>72.86</td>
<td>RBF</td>
</tr>
<tr>
<td>P14</td>
<td>E, G, BM</td>
<td>0.02</td>
<td>0.61</td>
<td>0.31</td>
<td>-0.09</td>
<td>58.31</td>
<td>MLP</td>
</tr>
<tr>
<td>P15</td>
<td>ESI, E, G, BM</td>
<td>0.01</td>
<td>0.80</td>
<td>0.71</td>
<td>-0.09</td>
<td>42.20</td>
<td>MLP</td>
</tr>
</tbody>
</table>

Note: G=gender; BM=body mass; E=exercise intensity; ESI=environmental stress index; MLP=multi-layer perceptron; RBF=radial basis function; \( \text{MSE}_{\text{val}} \)=average of mean squared error for all models computed on validation dataset; \( R^2 \)=coefficient of determination; \( \text{RMSE} \)=root mean squared error; \( \text{MBE} \)=mean bias error; \( \text{CV}_{\text{RMSE}} \)=coefficient of variation on \( \text{RMSE} \); L=litres; h=hours; * Best model selection is based on lowest \( \text{MSE}_{\text{val}} \), lowest \( \text{CV}_{\text{RMSE}} \) 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
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 intake.
Fluid intake needs of endurance athletes

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. 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. 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. Although the sweat rate and fluid loss is related to the metabolic rate, the rate of fluid ingestion is regulated by the osmotically driven thirst centre in the hypothalamus. 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 which may not only be limited to age, pregnancy or presence of diabetes.

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, it does support the hypothesis of Noakes 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.
Fluid intake needs of endurance athletes

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 counteracting 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.

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.

Acknowledgements
Professor Jules-Raymond Tapamo, Professor of Computer Engineering, School of Electrical, Electronic and Computer Engineering, University of KwaZulu-Natal, is thanked for insightful comments pertaining to the artificial neural network modelling technique. Professor Samuel N Cheuvront and Mr Nicholas Tam are thanked for provision of additional original research data sets used in the design and validation of the artificial neural network model.

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

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$^{-1}$.

Reference

2. Cheuvront SN, Montain SJ, Sawka MN. Fluid replacement and


