IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 8 November 2022, accepted 22 February 2023, date of publication 7 March 2023, date of current version 14 March 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3253384

APPLIED RESEARCH

Sweat Loss Estimation Algorithm for Smartwatches

KONSTANTIN PAVLOV[1], ALEXEY PERCHIK[1], VLADIMIR TSEPULIN[1], GEORGE MEGRE[1], EVGENII NIKOLAEV[1], ELENA VOLKOVA[1], GEORGII NIGMATULIN[1], JAEHYUCK PARK[2], NAMSEOK CHANG[2], WONSEOK LEE[2], AND JUSTIN YOUNGHYUN KIM[2]
[1]Samsung R&D Institute Russia, 127018 Moscow, Russia
[2]Health H/W Research and Development Group, Samsung Electronics, Suwon 16677, South Korea

Corresponding author: Georgii Nigmatulin (g.nigmatulin@partner.samsung.com)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Commission on Biomedical Problems of the Russian Academy of Sciences under Protocol No. 0251, and the Institutional Review Board (IRB) of Kookmin University under IRB No. KMU-202101-HR-253.

ABSTRACT This study presents a newly released algorithm for smartwatches – Sweat loss estimation for running activities. A machine learning model (polynomial Kernel Ridge Regression) is used to estimate the sweat loss in milliliters. A clinical dataset of 748 running tests of 568 people was collected and used for training / validation. The data presents a diversity of factors playing an important role in sweat loss: anthropometric parameters of users, distance, ambient temperature and humidity. The data augmentation technique was implemented. One of the key points of the algorithm is an accelerometer-based model for running distance estimation. The model we developed has a mean absolute percentage error (MAPE) = 7.7% and a coefficient of determination (R2) = 0.95 (at distances in the range of 2–20 km). The performance of the fully automatic sweat loss estimation algorithm provides an average root mean square error (RMSE) = 236 ml; more fundamentally, health-related parameter body weight percentage RMSE (RMSEBWP) = 0.33% and R2 = 0.79. To the best of the authors' knowledge, the algorithm provides the best performance of any existing solution or described in the literature.

INDEX TERMS IMU, PPG, fitness, running, sensors, skin temperature, smartwatch, sweat loss estimation, wearables, wrist-wearable device.

I. INTRODUCTION

Hydration level is an important parameter to maintain the health of your body, as almost every cell in the body needs water to function properly. Knowing the personal physiological need to consume water can be crucial in cases of intense physical exercise (such as long-distance running). Dehydration (lack of drinking) poses a risk of thermoregulation disorders and can lead to heat exhaustion or heat stroke [1]. Overhydration (excessive drinking) in rare cases can lead to hyponatremia [2]. Thirst is not an accurate indicator and incentive for water intake. Most guidelines contain detailed recommendations for best fluid intake practices based on estimates of sweat loss [3].

The associate editor coordinating the review of this manuscript and approving it for publication was Md. Kamrul Hasan.

Therefore, sweat loss estimation appears to be essential for individuals (professional and amateur level) interested in maintaining optimal physical performance and minimizing health risks associated with sports and fitness activities. Knowing how much individuals sweat during exercise allows to plan fluid intake effectively before (prehydration), during, and after (rehydration) exercise.

An amount of sweat loss can be calculated as a difference in individual's body weight (BW) before and after physical activity (nude weight with carefully toweled off sweat) [4]. Figure 1. Factors influencing the amount of a runner's sweat loss grouped by origin. Rectangles mark the factors currently available for analysis with wearable devices (a part of them are used by the reported algorithm).

Although the BW method is commonly used as a reference for research studies, it is impractical for individuals in real life sports and fitness activities.

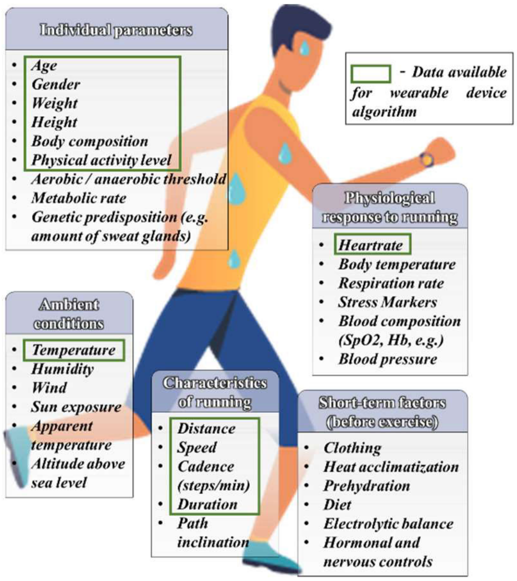


FIGURE 1. Factors influencing the amount of a runner's sweat loss grouped by origin. Rectangles mark the factors currently available for analysis with wearable devices (a part of them are used by the reported algorithm).

A number of studies are devoted to adopting direct sweat loss measurement sensors for wearable devices [5] in the form of absorbent patches [6], filter paper [7], glass capillaries [8] and microfluidic collectors [9]. However, these methods cannot be applied to wearables because of disposable components and need of maintenance by qualified personnel. For the same reason, medical tests such as urine specific gravity are also outside the scope of the study.

The goal of this study was to develop a method for indirect sweat loss estimation during running activities using data from existing wearable sensors and general user information.

Numerous of factors that influence sweating during running exercise are shown in Figure 1. It is shown that only some of these factors (marked by rectangles), can be used for wearable health services (with no need of manual input of exercise information).

Although environmental conditions can be obtained by wearable devices in the case of outdoor running (for example, from a local online weather forecast), we intentionally excluded most environmental factors in order to describe a single solution for all cases (outdoors and indoors, with or without an Internet connection).

Obviously, an uncertainty about all the ambient conditions, the user's clothes, physical and functional status before and during running leads to the limited performance of the indirect method we describe. The following key components of the study were considered in order to reduce the error of sweat loss estimation:

1) *Collecting a dataset with a variety of external conditions and user parameters*
2) *ML algorithm for accurate running distance estimation*
3) *Data augmentation for dataset expansion*

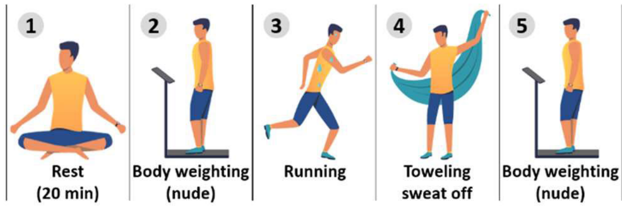


FIGURE 2. Subject's sequence of actions during data collection with running distances <10 km; for longer distances (≥10 km) subject stops running at the middle of distance for extra body weighting, so the sequence is 1-2-3-4-5-3-4-5.

4) *ML algorithm for sweat loss estimation*
5) *Validation of the developed algorithm*

These components are described in separate sections (II-IV) below. The results section (V) confirms the high predictive efficiency of the algorithm, making it suitable for wearable devices providing valuable information to the user. Conclusions and directions for further research are presented in Section VI.

This paper is an expanded version of the publication [9] for the "Biomedical and health informatics (BHI) and the Body Sensor Networks (BSN) conference (IEEE BHI & BSN, 2022). An explanation of the distance estimation algorithm, tests on an independent dataset including samples of different ethnic groups, and a more detailed description of each chapter are the distinctive characteristics of the actual article.

II. COLLECTING A DATASET

A total of 568 human subjects (age 18–53 years) participated in a total of 748 running trials (distances of 2–20 km, both indoor and outdoor). A special test rooms with treadmills and controlled environmental conditions (ambient temperature range of 10–40 °C and relative humidity range of 25–75 %) were prepared for indoor running trials. More information about the subjects' characteristics and ambient conditions can be found in Table 1. All subjects were capable of running distances under the specific environmental conditions (preliminary agreed). Subjects admission to the tests and control of the subject's condition during running were supervised by a medical doctor.

Two remote testing sites were chosen for dataset collection. The majority of the data (549 running trials) was collected with Eastern Asian subjects (South Korea, Kookmin University). The smaller part of the data (199 running trials) was collected with Eastern European subjects (Russian Federation, Institute of Biomedical Problems). The purpose of testing sites diversification was to validate methods with inter-ethnic data collected with different conditions, equipment, and researchers in order to minimize the risk of bias between algorithm performance for training data and any other data (including real-life conditions). The data collection protocol was reviewed and approved by the Commission on Biomedical Problems of the Russian Academy of Sciences (Protocol No 0251 of March 1, 2020) and by the Institutional

TABLE 1. Main values characterizing population sampling and conditions used for algorithm training and validation.

Parameter	Weight, kg			
	42–59	60–72	73–84	85–123
		Men		
N subjects	22	103	61	101
N running trials	29	103	83	160
Age, yr	30±6	28±6	30±7	30±6
Height, cm	167±4	173±4	179±5	182±6
Steps, min^{-1}	172±14	160±12	159±12	146±16
Heartrate, min^{-1}	157±11	153±15	149±14	148±15
Amb. temp., °C	25±3	24±8	25±7	23±8
Amb. r.h., %	46±10	44±17	44±15	44±16
		Women		
N subjects	175	89	16	1
N trails	189	135	38	11
Age, yr	29±8	31±10	32±7	36±0
Height, cm	162±5	169±6	174±4	169±0
Steps, min^{-1}	157±15	144±18	149±16	151±3
Heartrate, min^{-1}	153±16	153±12	157±10	166±4
Amb. temp., °C	24±6	23±7	22±7	21±9
Amb. r.h., %	38±14	40±15	42±6	25±0

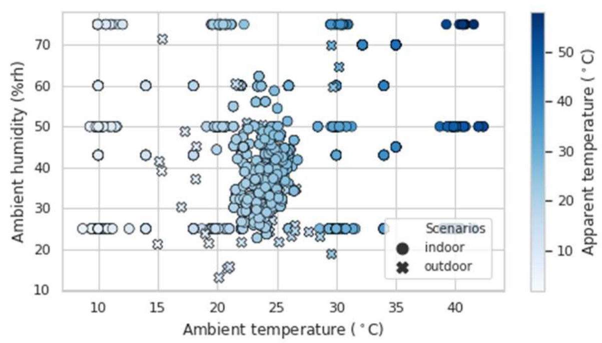


FIGURE 3. Environment conditions used for data collection (air temperature, relative humidity and apparent temperature).

Review Board (IRB) of Kookmin University (IRB registration number KMU-202101-HR-253 of February 15, 2021). Before the participation, all subjects received detailed explanation of the clinical tests and signed the informed consents. The data was kept anonymized and it was used only for the intended research purpose.

According to the protocol, each subject was examined by the medical staff to measure anthropometric parameters (including height) and to survey for general information (age, gender, medical history, exercise habits, and current medication). After the screening, subjects were asked to put on the smartwatch device on their left wrist and do the following actions (see Figure 2): rest (seating) for 20 minutes in a room with normal temperature (about 23°C); perform the 1st nude body weighing with precise CAS-HB-150 (South Korea) scales; run a distance with predefined conditions.

If the trial distance is shorter than 10 km, then the subject runs the whole distance in one go; carefully towels off sweat; performs the 2nd nude body weighing and completes the test. In this case, sweat loss reference is defined as the

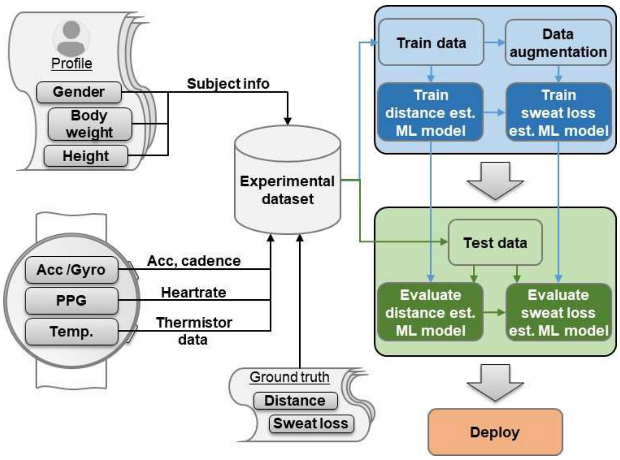


FIGURE 4. Data processing pipeline of sweat loss algorithm training and testing. Sweat loss ML model requires general subject's information, IMU, PPG and temperature data.

difference between the 1st and 2nd body weights, excluding water intake. If the trial distance is equal or longer than 10 km, then the subject stops running in the middle of the distance; performs the 2nd nude body weight; changes into dry clothes; runs the second half of the distance; carefully towels off sweat; performs the 3rd nude body weight; and completes the test. In this case, there are two sweat loss references: for the half distance (difference between the 1st and the 2nd body weights, excluding water intake) and for the whole distance (difference between the 1st and the 3rd body weights, excluding water intake). In the second case, half-distance and whole-distance trials were considered as two samples in the dataset.

Figure 3 illustrates a diversity of environmental conditions at indoor and outdoor running scenarios. Our aim was to cover the vast majority of ambient conditions for real-life running monitoring use cases. Some of the data points structured as a "rectangular grid" correspond to running trials under controlled environmental conditions (treadmill in a special climatic chamber). Vertical points cloud at ambient temperatures of around 23°C corresponds to running trials in regular gym conditions (with treadmill). Data points marked with an 'x' symbol correspond to outdoor running (conditions are defined by local weather). A smartwatch Galaxy Watch Active 2 (Samsung, South Korea) model was used for data collection. This smartwatch includes a set of sensors commonly used in modern wearable devices: photoplethysmography (PPG used for heartrate measurement), an inertial measurement unit (IMU including a three-axis accelerometer and three-axis gyroscope), and internal thermistors (commonly used for monitoring of components temperature). Smartwatch devices were sufficiently tightly fixed on the wrist (not too tight, not too loose) during the running, providing convenient wearing for a subject and correct operation of sensors to obtain a high signal quality.

Besides the raw sensor signals, the dataset contains processed values of heartrate from PPG and running

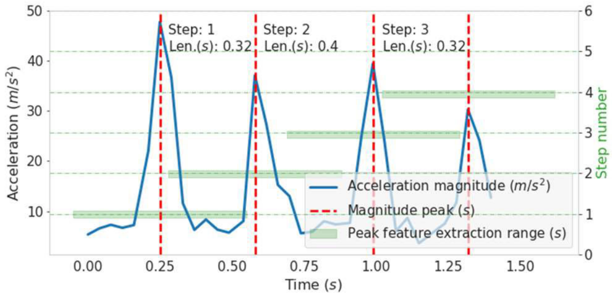


FIGURE 5. Samples of IMU (accelerometer magnitude) data. High peaks correspond to wrist acceleration associated with steps, green rectangles correspond to data windows length used for features extraction.

cadence (steps/min) from IMU. The whole data processing pipeline is presented in Figure 4.

III. ML ALGORITHM FOR RUNNING DISTANCE ESTIMATION

During intense running, human muscles do a lot of work, producing excessive heat. Sweating helps the body to dissipate the heat and maintain the normal temperature (in general, the more work done – the more heat – the more sweat). In physics, work is a product of force and displacement (when force and displacement are co-directional). Obviously, running distance (equal to displacement) is one of the most meaningful features for sweat loss estimation. That's why we place greater focus on the distance estimation ML model in this study.

One way to estimate the runner's distance is GPS. However, this approach has a significant drawback – it's inapplicable when an individual is running on a treadmill. An IMU-based approach is applicable for most running conditions and scenarios.

The number of runners' steps can be easily extracted from the IMU data by calculating the number of peaks corresponding to wrist acceleration and associated with leg acceleration (see Figure 5). Total acceleration (R) was calculated using its projections (a_x, a_y, a_z) measured by 'inertial measurement unit (IMU) sensor:

$$R = \sqrt{a_x^2 + a_y^2 + a_z^2}, \tag{1}$$

However, the number of steps does not indicate the distance run, since the time taken per step may depend on the speed, height and other runner's parameters. The maximum acceleration of hand (R) is a parameter that correlates with the length of each step. So, we can write simple linear equation to approximate the whole track distance (D):

$$D = n \cdot h \cdot (l + c \cdot s(R)), \tag{2}$$

where n – number of steps made by athlete, h – height of the athlete, l – average distance of each step, $s(R)$ – some statistic based on hand acceleration time series. We found out that the third quartile of the maximum acceleration series gives good estimation results. It is simple to calculate and, at the same time, is robust to the presence of pauses in training session. If the user's height is known, we obtain a linear regression model with parameters l, c and $D/(h \cdot n)$ as target value.

The whole dataset was split into train and test sets, keeping an identical distributions of user profiles, features and targets in both splits. Running trials of the same user can present only in one of the train or test sets.

TABLE 2. Performance of distance estimation ML model.

Data splits (number of trials)		Parameters of regression performance			
		MAPE, %	*MAE, km*	*RMSE, km*	R^2
Train	(332)	8.4	0.64	0.88	0.92
Cross-val	(332)	8.8	0.67	0.92	0.91
Test	(416)	7.7	0.61	0.85	0.95

MAPE – mean absolute percentage error
MAE – mean absolute error
RMSE – root mean square error
R^2 – coefficient of determination

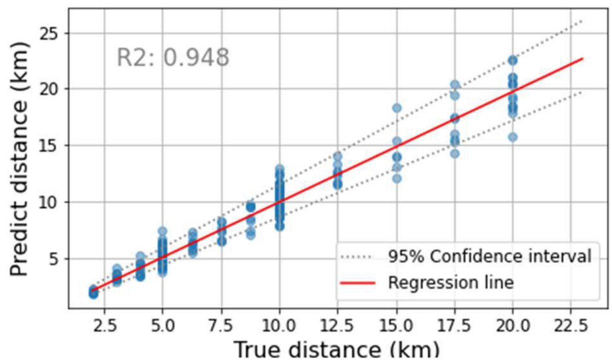


FIGURE 6. Correlation between predicted and true distance. The dotted lines indicate the 95% confidence interval for the regression line.

Experiments with various algorithms showed that better distance estimation doesn't always lead to better estimation of sweat loss. The final model in form (2) was selected to maximize the cross-validation score of sweat loss estimation while taking into account deployment restrictions.

The performance of the distance estimation ML model is presented in Tab. 2 and Figure 6.

Here and elsewhere in this paper a cross-validation procedure is applied in leave one subject out way only for the train set (not for the whole dataset). On each round of cross-validation a distance estimation model is trained and then used to predict distances on for all items in train set. Obtained predictions are used as features for sweat loss estimation model training. After training stage is completed, distance and sweat loss estimation models are consequentially applied to validation set.

The proposed model requires only two parameters to track distance estimation: the number of peaks in acceleration and the value of the third quartile of peak acceleration series. Values of these parameters can be continuously updated during the training session, so storage of all IMU data is not required. Low computational and memory requirements are great advantages of the proposed algorithm in the context of model deployment.

The computation of training session parameters consists of several steps. In the first step, the input signal is downsampled to 12.5 Hz. This operation reduces the computational cost without any decrease in algorithm performance. In the next step, the peak extraction algorithm is applied to find the maximum acceleration and count the number of steps. Since all the sequence of acceleration values required to compute quantiles, we need to add defined acceleration into some storage, for example, an array. We used the t-digest [11] algorithm to obtain an estimation of the third quartile and at the same time meet memory requirements.

Initially, the training set didn't contain whole running trials with short distances <5 km. Short-distance running samples were added to the training set by augmentation.

The accuracy of distance estimation could be increased for outdoor running for which GPS coordinates are available, but this is out of the scope of our study.

IV. DATA AUGMENTATION FOR DATASET EXPANSION

A data augmentation procedure was implemented in order to expand the dataset and make it more diverse in distances. The protocol of clinical tests was limited to obtain only one reference value (targets) of sweat loss at the end (or in the middle) of a running trial. A technique described in this section allows to obtain reference values of sweat loss for any segment of a running trial, thus increasing the number of samples to train the sweat loss estimation ML model.

The augmentation procedure consists of two steps: first – training a model for aggregated output regression; second – augmentation of the training dataset with reference target normalization. Aggregated output regression is a task, where a label is associated with a set of observations in a region. Suppose we have some covariate space X and a response space R. The aggregated output regression model [12] is defined as (3):

$$\int_{X} f(x)\, d\prod\nolimits_{i}(x) + \xi_i,,,\ \xi_i \sim N(0,\sigma^2) \tag{3}$$

where Xi ⊆ X is an observation region with distribution Π_i with Lebesgue density π_i and ξ_i is an independently distributed Gaussian noise with $\sigma > 0$ [13]. A multilayer perceptron (MLP) is used for aggregated output regression (see Figure 7). We use batch normalization right after input. ReLU is used as a dense layer activation function. This model approximates the target conditional distribution $P(\hat{y}|f)$ – where f is our set of features and $\hat{y}$-estimation of target variable. This vector contains the same statistical values as in the sweat loss estimation model, calculated for small segments of a trial. One-minute-long segments were used to train the model. Estimation of sweat loss for each segment is calculated after a forward pass through the neural network. The sum of estimations for all segments of a workout gives the estimation of total sweat loss. Differences between these

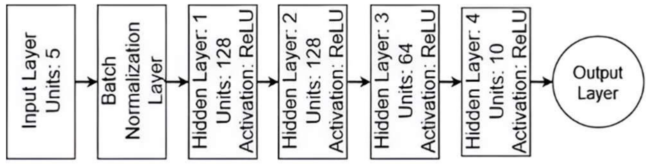


FIGURE 7. Multilayer perceptrons are used for aggregated output regression. It allows estimation of sweat loss targets for short (1 km) running segments for further training of the ML model.

aggregated estimations and ground truth targets for workouts were used to calculate the MSE loss function for error backpropagation.

After MLP is trained, the dataset is augmented by splitting the whole running trials into segments of approximately 1 kilometer long. We've also experimented with other lengths of segments (2, 3, 4 km), but we've found that 1 km of augmentation gives the best performance.

MLP estimations are normalized for each running trial as follows:

$$y_{new} = \frac{\hat{y} * y}{\sum_{i=0}^{n} \hat{y}_i}, \tag{4}$$

where $\hat{y}_i$- vector of estimated sweat loss targets for each of 1 km segments, y – our target variable, and y_{new} – normalized target variable, which is used for augmentation.

Data augmentation allowed us to expand the dataset from 748 to 6296 running samples (including whole running trials and their segments). The technique described in this section helped us to improve sweat loss estimation algorithm performance (especially at short running distances <5 km) for both train and test data.

V. ML ALGORITHM FOR SWEAT LOSS ESTIMATION

The amount of data (even with augmentation) has appeared to be insufficient for deep learning architecture implementations. The best estimation performance was achieved with a feature-based kernel ridge regression model with a polynomial kernel:

$$\mathrm{k}(x_i, y_j) = (1 + \sum\nolimits_{k=1}^{d} x_i, ky_j, k)^m, \tag{5}$$

where d is a size of feature vector.

The solution of the Kernel ridge regression model has the following form:

$$\mathrm{f(x)} = \sum\nolimits_{n=1}^{N} \alpha_i k(x, x_i), \tag{6}$$

where N - is a size of training set.

The input feature vector contains a set of features, calculated from each running sample (whole trials and segments):

- maximum heart rate,
- average cadence (steps/min),
- average thermistor temperature,
- user's gender,
- distance run × user's weight × average thermistor temperature.

TABLE 3. Performance of sweat loss estimation ML model.

Data splits (number of trials)		Parameters of regression performance			
		MAE, ml	*RMSE, ml*	*RMSEBWP, %*	*R^2*
Model trained only with whole running trials					
Cross-val	(332)	198	289	0.377	0.744
Test	(416)	156	237	0.330	0.788
Model trained with augmented running samples (trials and segments)					
Cross-val	(332)	193	286	0.374	0.750
Test	(416)	152	236	0.330	0.790

RMSEBWP – body weight percentage RMSE

$RMSEBWP = 100 \cdot \sqrt{\frac{1}{N} \cdot \sum_{i=1}^{N} \left(\frac{y_i - \hat{y}_i}{BW_i}\right)^2}$, were N – number of trials in a data split, y_i – predicted sweat loss (ml), $\hat{y}_i$ – true sweat loss (ml), BW_i – body weight (g).

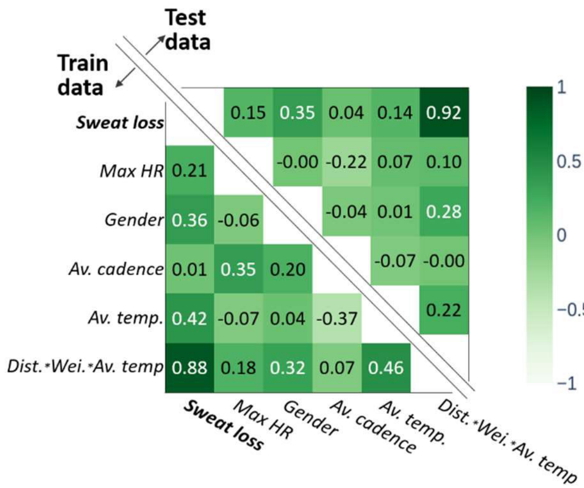


FIGURE 8. Correlation analysis between features and target of the Sweat loss estimation ML model. A high correlation between the target '*Sweat loss*' and the multiplicative feature '*Dist. ∗Wei. ∗Av. temp*' can be observed.

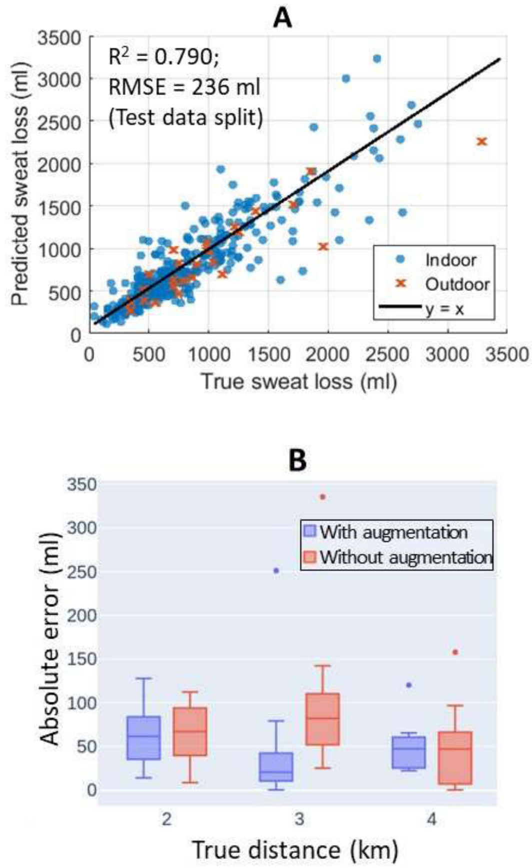


FIGURE 9. (A) – Scatter plot of predicted versus true sweat loss for the model trained with augmentation. (B) – Comparative boxplot of absolute sweat loss error for models trained with and without augmentation.

The last multiplicative feature was found through a deep data analysis (it can be associated with work done by a runner under certain temperature conditions). A correlation plot between those features and the target is shown in Figure 8. We have found that the multiplicative feature has the best correlation with the target (a very close equation for sweat loss estimation was described in [14]). Each feature is linearly transformed to range from 0 to 1.

Regularization parameter and parameters of kernel (5) were selected by optimization of RMSE error calculated using cross-validation with the training part of dataset. Implementation, provided by Optuna framework [15], was used to perform hyper parameters tuning. A Tree parzen estimator model was selected as the surrogate model.

The performance of the ML model trained with and without the data augmentation technique is presented in Table 3. The Ridge regression model with a polynomial kernel was trained on a training data split and evaluated on both training (cross-validation) and test data splits. A model that was trained using an augmentation technique outperforms a model with no augmentation at all performance metrics.

An important health-related parameter is the body weight percentage of sweat loss. A number of sports medicine publications [16] state that 2% of body mass water loss with sweat can be harmful for a human with some changes in mental performance and endurance. Our ML model provides low RMSEBWP error (<0.4%), thus it can reliably inform user about upcoming dehydration threat.

Figure 9A shows concordance between sweat loss predicted by the ML model and ground truth values (from changes in body weight). Only a few data samples have high sweat loss error (up to 1000 ml), for the most of cases predictions are well correlated with true sweat loss. Figure 9B illustrates that augmentation is especially beneficial at short running distances, e.g., at 3 and 4 km.

VI. INDEPENDENT EXTERNAL EVALUATION OF ALGORITHM

The sweat loss algorithm was trained with Eastern Asian and Eastern European subjects (with data collection at two geographical regions).

TABLE 4. Performance of sweat loss estimation on the independent dataset.

Data hand-splits (number of trials)		Parameters of regression performance			
		MAE, ml	*RMSE, ml*	*RMSEBWP, %*	R^2
Left	(320)	170	234	0.284	0.760
Right	(320)	156	220	0.273	0.750

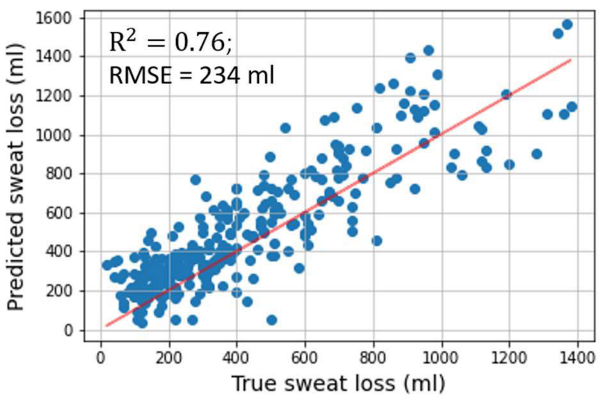


FIGURE 10. Scatter plot of predicted versus true sweat loss for the independent dataset (Left hand).

An independent external evaluation of the algorithm was performed with data collection at the third geographical region in South America with different ethnical groups of subjects (including Latino and Black). The dataset represents running trials of people from a wide range of anthropometric characteristics. A total of 320 experiments were obtained in a similar manner as described in Section II. In addition, to test the versatility of the algorithm, metrics were calculated based on data that was collected from both the left and right hands. The distances of the running trials were 2.5, 5, and 7.5 km. It is important to note that when calculating metrics on the independent dataset, no training of a new model was carried out — only the model originally obtained by the R&D process was employed.

The performance of the ML model for sweat loss estimation on the independent dataset is presented in Table 4.

The algorithm demonstrates good predictive ability. The RMSEBWP error is 0.284% and 0.273% for the right and left arms, respectively, which indicates that the system is able to provide reliable information about the upcoming threat of dehydration. The error in RMSEBWP of less than 0.1% was predicted for 96 (124), less than 0.2% for 194 (214) and less than 0.5% for 289 (295) people on the left (right) hand, respectively.

As compared to the metrics in Table 3 for the initial dataset, the MAE and RMSE have decreased. This happened as a result of the shorter running trials in the independent dataset (maximum 7.5 km, instead of 20 km distance). Sweat losses were reduced as a result, and measurement errors were reduced. However, the R2 metric has remained approximately the same as the performance score. True sweat loss and predictions are often well correlated; for both arms, the R2 is greater or equal to 0.75. The algorithm performs well for both the left and right arms simultaneously.

The correspondence between the ground truth values and the predicted sweat loss by the ML model is illustrated in Figure 10. This demonstrates that the algorithm is tolerant of the wide range of anthropometric and ethnic traits of the participants in the independent dataset because with no bias in predictions (no regularities of underestimates or overestimates at independent dataset). All the prediction errors within the independent dataset are below 500 ml.

Each of these results makes it possible to state the system's reliability when using various types of data and the high accuracy of the predictions, which allows to provide valuable advice on user's hydration.

VII. DISCUSSIONS AND CONCLUSION

The described method of sweat loss estimation is based on sensors that are currently available in most smartwatches and fitness trackers. Although we used indirect estimation (no direct measurements of sweat amount), the approach showed high performance.

It was shown that multiple factors influence sweating during the running exercise and only a part of those factors can be used for the smartwatch algorithm. A set of measures to overcome the uncertainty of unknown factors were implemented: dataset with a variety of external conditions and user parameters; improved running distance estimation; data augmentation technique; and sweat loss ML model optimization.

The essence of the proposed algorithm is the top performance among existing solutions or ever described in literature in the area of smartwatch-based fully automatic sweat loss estimation (to the authors' knowledge).

There is still a field for further research. Probably the algorithm performance can be improved with access to additional data, e.g. geolocation, altitude above sea level, weather info (from on-line services), and also the user's sports habits and some individual information (directly from the user). Recent appearance of BIA sensor [17] in smartwatch could be beneficial in sweat loss and body hydration monitoring (considering body composition changes).

The sweat loss estimation research should be directed to other activities and user scenarios (expand it from running to any daily life and sports activities).

Monitoring the body's hydration and reminders of timely consumption of water will motivate users to achieve and maintain a healthy lifestyle and high results in physical exercises.

REFERENCES

[1] V. A. Convertino, L. E. Armstrong, E. F. Coyle, G. W. Mack, M. N. Sawka, L. C. Senay, and W. M. Sherman, "ACSM position stand: Exercise and fluid replacement," *Med. Sci. Sports Exerc.*, vol. 28, no. 10, pp. 1–9, Oct. 1996.

[2] T. M. Mclellan, S. S. Cheung, W. A. Latzka, M. N. Sawka, K. B. Pandolf, C. E. Millard, and W. R. Withey, "Effects of dehydration, hypohydration, and hyperhydration on tolerance during uncompensable heat stress," *Can. J. Appl. Physiol.*, vol. 24, no. 4, pp. 349–361, Aug. 1999.

[3] R. J. Shephard, "American college of sports medicine position stand: Exercise and fluid replacement," *Yearbook Sports Med.*, vol. 2007, pp. 254–255, Jan. 2007.

[4] E. O'Neal, T. Boy, B. Davis, K. Pritchett, R. Pritchett, S. Nepocatych, and K. Black, "Post-exercise sweat loss estimation accuracy of athletes and physically active adults: A review," *Sports*, vol. 8, no. 8, p. 113, Aug. 2020.

[5] B. Zhong, K. Jiang, L. Wang, and G. Shen, "Wearable sweat loss measuring devices: From the role of sweat loss to advanced mechanisms and designs," *Adv. Sci.*, vol. 9, no. 1, Jan. 2022, Art. no. 2103257.

[6] L. Ortega, A. Llorella, J. P. Esquivel, and N. Sabaté, "Self-powered smart patch for sweat conductivity monitoring," *Microsyst. Nanoeng.*, vol. 5, no. 1, Jan. 2019.

[7] K. Van Hoovels, X. Xuan, M. Cuartero, M. Gijssel, M. Swarén, and G. A. Crespo, "Can wearable sweat lactate sensors contribute to sports physiology?" *ACS Sensors*, vol. 6, no. 10, pp. 3496–3508, Oct. 2021.

[8] P. W. R. Lemon, "A simple and inexpensive method for making sweat collection capsules," *Res. Quart. Exerc. Sport*, vol. 54, no. 3, pp. 299–301, Sep. 1983.

[9] L. B. Baker, J. B. Model, K. A. Barnes, M. L. Anderson, S. P. Lee, K. A. Lee, S. D. Brown, A. J. Reimel, T. J. Roberts, R. P. Nuccio, J. L. Bonsignore, C. T. Ungaro, J. M. Carter, W. Li, M. S. Seib, J. T. Reeder, A. J. Aranyosi, J. A. Rogers, and R. Ghaffari, "Skin-interfaced microfluidic system with personalized sweating rate and sweat chloride analytics for sports science applications," *Sci. Adv.*, vol. 6, no. 50, Dec. 2020, Art. no. eabe3929.

[10] K. Pavlov, A. Perchik, V. Tsepulin, G. Megre, E. Nikolaev, E. Volkova, J. Park, N. Chang, W. Lee, and J. Y. Kim, "Sweat loss estimation solution for smartwatch," presented at the IEEE-EMBS Int. Conf. Biomed. Health Inform. (BHI) Jointly Organised 17th IEEE-EMBS Int. Conf. Wearable Implantable Body Sensor Netw. (BSN), Ionnina, Greece, Sep. 2022.

[11] T. Dunning, "The t-digest: Efficient estimates of distributions," *Softw. Impacts*, vol. 7, Feb. 2021, Art. no. 100049.

[12] Y. Tanaka, T. Tanaka, T. Iwata, T. Kurashima, M. Okawa, Y. Akagi, and H. Toda, "Spatially aggregated Gaussian processes with multivariate areal outputs," 2019, *arXiv:1907.08350*.

[13] H. Zhu, A. Howes, O. van Eer, M. Rischard, Y. Li, D. Sejdinovic, and S. Flaxman, "Aggregated Gaussian processes with multiresolution Earth observation covariates," 2021, *arXiv:2105.01460*.

[14] K. J. Sollanek, M. Liu, A. Carballo, A. R. Caldwell, and S. N. Cheuvront, "The accurate prediction of sweat rate from energy expenditure and air temperature: A proof-of-concept study," *Appl. Physiol., Nutrition, Metabolism*, vol. 45, no. 11, pp. 1299–1305, Nov. 2020.

[15] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Jul. 2019, pp. 2623–2631.

[16] S. Sagawa and M. K. Yousef, "Physiology of man in deserts: A review on hypohydration and rehydration," *J. UOEH*, vol. 8, no. 1, pp. 85–92, 1986.

[17] (Aug. 16, 2021). *Measure Your Body Composition With the Galaxy Watch4 Series*. Samsung Levant. Accessed: Sep. 20, 2022. [Online]. Available: https://www.samsung.com/levant/support/mobile-devices/measure-your-body-composition-with-the-galaxy-watch4-series/

KONSTANTIN PAVLOV received the Ph.D. degree in electronics and microelectronics from the Department of Biomedical Systems, National Research University of Electronic Technology, Moscow, Russia, in 2011. Since 2012, he has been with the Samsung R&D Institute Russia. He is an author and a coauthor of 15 articles and 12 patents. His main research interests include wearable sensors, mobile health, circuit engineering, signal processing, and ML algorithms.

ALEXEY PERCHIK received the specialist degree in optical engineering and the B.Sc. degree in computer science from Bauman Moscow State Technical University and the Ph.D. degree in scientific instruments for experimental physics from the Russian Academy of Science. During his Ph.D., he worked in the field of optical spectroscopy and polarimetry. Since 2016, he has been an Engineer with the Samsung R&D Institute Russia. His main research interests include biophotonics and biosignal processing applications for wearables.

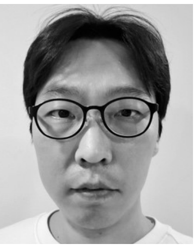
VLADIMIR TSEPULIN received the M.Sc. and Ph.D. degrees in optical and electronic devices and systems from Bauman Moscow State Technical University, in 2014 and 2019, respectively. His main research interests include optical sensors, machine learning methods for NLP, time series, and sensors signal processing.

GEORGE MEGRE received the M.Sc. degree in intelligence systems from the Moscow Aviation Institute, in 2020. From 2020 to 2022, he was a Research Engineer with the Lifestyle Analysis Laboratory, Samsung R&D Institute Russia. His research interests include machine learning and deep learning for signal, image, speech processing, and computer vision.

EVGENII NIKOLAEV received the degree in mathematics and computer science from the Chuvash Pedagogical Institute. Since 2021, he has been a Machine Learning Engineer with the Samsung R&D Institute Russia. His main research interests include machine learning and deep learning for biomarker, NLP, and sound processing.

ELENA VOLKOVA received the master's degree in optics and the Ph.D. degree in biophysics from Saratov State University (SSU). She was a Research Fellow with the Optoelectronics and Measurement Techniques Laboratory, University of Oulu, Finland; an Engineer with the Research and Educational Institute of Nanostructures and Biosystems, SSU; and a Research Associate with the Laboratory of Biomedical Optics of Research Educational Institute Optics and Biophotonics, SSU. Since 2018, she has been with the Samsung R&D Institute Russia. She is an author and a coauthor of 26 articles and three patents. Her main research interests include wearable sensors, mobile health, optics and biophotonics, absorption, and luminescence spectroscopy of nanoparticles and crystals.

GEORGII NIGMATULIN received the M.Sc. degree in applied mathematics and physics from the Institute for Information Transmission Problems, Moscow Institute of Physics and Technology, in 2022. Since 2021, he has been an Machine Learning Engineer with the Samsung R&D Institute Russia. His research interests include machine learning, digital signal processing, causal and Bayesian methods, and optimization.



JAEHYUCK PARK received the B.Sc. degree. He has been a BIA-Sensing Hardware Engineer with Samsung Electronics, since 2011. His main research interests include bio-electrical impedance and the body hydration technologies.

NAMSEOK CHANG received the Ph.D. degree in information and communications engineering from the Korea Advanced Institute of Science and Technology, in 2012. He has been a Bio-Sensing Hardware Engineer with Samsung Electronics, since 2012. His main research interest includes the bio-electrical impedance technology, which measures the body composition.

WONSEOK LEE received the M.Sc. degree in electrical and computer engineering from Hanyang University, South Korea, in 2016. Since 2004, he has been with the Mobile Business Division, Samsung Electronics. His main research interests include biomedical sensors, mobile health, and circuit engineering.

JUSTIN YOUNGHYUN KIM received the Ph.D. degree in electrical engineering from the Department of Electrical Engineering, California Institute of Technology, Pasadena, USA, in 2012. Since 2012, he has been with the Samsung Mobile eXperience Business Division. He is an author and a coauthor of ten articles and 69 patents. His main research interests include wearable health, mobile health, and accessory health sensor.