American Journal of Industrial Medicine

WILEY

AMERICAN JOURNAL OF INDUSTRIAL MEDICINE

RESEARCH ARTICLE OPEN ACCESS

Assessing the Feasibility of Wearable Devices for Physiological Monitoring and Heat Risk Prediction in Outdoor Agricultural Workers

Sinan Sousan[1,2,3] | Elizabeth Mizelle[4] | Qiang Wu[1] | Rui Wu[5] | Guy Iverson[6] | Ciprian Popoviciu[7] | Dylan Hemedinger[8] | Eliezer Loyola[9] | Takari Hood[10] | Monica Ramirez[11] | Jo Anne Balanay[6]

[1]Department of Public Health, Brody School of Medicine, East Carolina University, Greenville, North Carolina, USA | [2]North Carolina Agromedicine Institute, Greenville, North Carolina, USA | [3]Center for Human Health and the Environment, NC State University, Raleigh, North Carolina, USA | [4]College of Nursing, East Carolina University, Greenville, North Carolina, USA | [5]Department of Information Technology, College of Computing and Software Engineering, Kennesaw State University, Kennesaw, Georgia, USA | [6]Environmental Health Sciences Program, Department of Health Education and Promotion, East Carolina University, Greenville, North Carolina, USA | [7]Department of Technology Systems, East Carolina University, Greenville, North Carolina, USA | [8]Department of Biology, Thomas Harriot College of Arts and Sciences, East Carolina University, Greenville, North Carolina, USA | [9]Department of Foreign Languages & Literatures, College of Arts and Sciences, East Carolina University, Greenville, North Carolina, USA | [10]Department of Public Health Education, College of Health and Sciences, North Carolina Central University, Durham, North Carolina, USA | [11]Department of Psychology, Thomas Harriot College of Arts and Sciences, East Carolina University, Greenville, North Carolina, USA

Correspondence: Sinan Sousan (sousans18@ecu.edu)

Received: 19 December 2025 | **Revised:** 3 April 2026 | **Accepted:** 28 May 2026

Funding: National Institute of Environmental Health Sciences of the National Institutes of Health, Grant/Award Number: P30ES025128; NC GlaxoSmithKline Foundation; North Carolina Agromedicine Institute; National Institute for Occupational Safety and Health (NIOSH), Grant/Award Number: K01OH012697; National Institutes of Health or the Centers for Disease Control and Prevention (CDC)

ABSTRACT

Background: Outdoor agricultural workers experience significant heat exposure, yet few studies have evaluated whether wearable sensors can reliably measure continuous physiological responses in real field conditions. This pilot study examined the feasibility and predictive utility of core temperature, hydration, heart rate, and movement data collected from commercially available wearables.

Methods: Thirty farmworkers in eastern North Carolina wore three devices, the CALERA core temperature sensor, the Nix Hydration Biosensor, and the Garmin Vivoactive smartwatch, during one July work shift. 1 min, physiological and environmental data were aligned to compute the modified Physiological Strain Index, mPSI. Device reliability, physiological responses, and heat exposure were summarized. Multi-linear regression and machine learning, Random Forest (RF) and Gradient Boosting Regression (GBR), models were evaluated to predict mPSI, using raw and smoothed predictors to estimate mPSI.

Results: CALERA retrieved 100 percent of the data, Garmin retrieved 97 percent, and Nix retrieved 80 percent. WBGT remained consistently high, approximately 28 to 29 degrees Celsius, and 12.6 percent of core temperature measurements exceeded 38 degrees Celsius. mPSI ranged from minimal to high heat strain, with 8.3 percent of observations at or above 5. Heart rate showed the strongest association with mPSI. The RF model using 60-min moving averages achieved the highest predictive accuracy with an R^2 of 0.96 and a root mean squared error of 0.25. Wearables captured meaningful physiological patterns that corresponded with environmental heat exposure, and WBGT-based risk levels consistently overestimated heat strain relative to mPSI.

This work was performed at East Carolina University.

Conclusion: Multi-sensor wearable systems are feasible for continuous heat strain monitoring and can support accurate prediction of heat-related risk among agricultural workers. This study is the first to evaluate the CALERA, Nix, and Garmin devices combined for heat strain monitoring in outdoor agricultural workers. By combining all three wearables in a field-based feasibility assessment, it addresses a key gap in validating consumer-grade technology for occupational safety in this high-risk workforce. The findings of the study support the use of wearable devices for real-time heat strain alerts, individualized hydration and work-rest guidance, and early warning systems to help prevent heat-related illness in physically demanding outdoor occupations.

1 | Introduction

The United States Environmental Protection Agency (EPA) [1] and the World Health Organization (WHO) [2] both emphasize that rising global temperatures have intensified heat exposure, creating a growing threat to public health, occupational safety, and productivity. The WHO has recommended programs that integrate engineering and administrative controls, hydration and rest management, medical surveillance, and alignment with global sustainability objectives. Together, these frameworks highlight the pressing need for coordinated, evidence-based strategies that integrate public health and occupational safety approaches to mitigate the increasing human and economic costs of heat in a warming climate [1, 2].

Occupational heat stress refers to situations in which a worker's body accumulates excessive heat due to the combined effects of internal metabolic activity, surrounding environmental conditions, and the insulating properties of their clothing [2, 3]. Physiologically, heat stress alters fluid balance, elevates heart rate [4], increases core body temperature [5], and intensifies thermoregulatory and perceptual strain [6]. Prolonged exposure may lead to a range of heat-related illnesses (HRI), including heat exhaustion, rhabdomyolysis, heat syncope, heat cramps, and heat rash [7]. The most severe outcome, heat stroke, can result in irreversible organ damage or death if not treated immediately. When the body's thermoregulatory mechanisms can no longer maintain thermal equilibrium, the risk of HRI increases substantially [8].

Outdoor workers are particularly susceptible to heat stress due to prolonged physical activity in extreme environmental conditions. High-risk occupational groups include agricultural laborers, forestry and fisheries personnel, construction workers, members of the armed forces, first responders, and athletes [3]. Agricultural work requires especially demanding tasks, such as harvesting, planting, pruning, load carrying, weeding, and digging, which often occur in hot and humid climates, which amplify physiological strain [9, 10]. Preventive measures should intensify as environmental heat levels rise [11]. Heat stress can be assessed using both physiological and environmental metrics. Hydration status, a key determinant of thermoregulation, can be evaluated using urine and blood samples collected before, during, and after heat exposure [9, 12], where several studies have relied on urine analysis to estimate hydration [13–15], and sweat loss is typically calculated from the difference in body weight before and after exposure [16]. Core body temperature is commonly measured using rectal probes [17, 18], tympanic thermometers [12, 13], or oral thermometers [17]. Heart rate can be recorded using electrocardiography or transmissive pulse oximetry by trained professionals [19, 20]. These data are then integrated to compute the PSI, providing an overall indicator of physiological strain [21]. However, these tools are discrete measurements and lack real-time implications that could immediately affect workers.

Technological innovation has significantly enhanced the accuracy and accessibility of heat stress monitoring. Core body temperature can now be measured through ingestible thermometric capsules that wirelessly transmit data from the gastrointestinal tract [22]. Common systems include the CorTemp™ capsule and the BodyCap e-Celsius Performance pill, both validated for occupational field studies [22, 23]. Noninvasive wearable devices, such as CALERA, estimate core body temperature by combining skin temperature and heat-flux measurements [24, 25]. Skin temperature, an additional indicator of thermal strain, can be recorded using Thermochron iButton loggers placed at peripheral and central body sites [26, 27]. The CALERA Research wearable integrates both skin and core temperature data, which can be used in real-time heat strain applications [28]. Parsons et al. [29] used the CALERA (formerly CORE thermal monitors) to monitor 100 Cambodian workers across five sectors, including agriculture, across four climatic zones, for 7 days to record physiological data. Advances in commercial wearables have further expanded the capabilities of heart rate monitoring. Devices such as the Apple Watch, Garmin, Samsung Watch, Fitbit, and Oura Ring utilize electrocardiogram (ECG) and photoplethysmography (PPG) sensors to continuously monitor heart rate [30]. Sweat sensors, such as the Nix Hydration Biosensor, Epicore Connected Hydration patch, and Gatorade Gx Sweat Patch, are advancing noninvasive monitoring of electrolytes and fluid loss by analyzing eccrine sweat, which reflects physiological changes related to hydration and metabolism [31]. Compared with traditional laboratory instruments, these commercially available sensors are affordable, widely accessible, and suitable for large-scale data collection, making them valuable tools for both occupational research and public health applications in heat strain monitoring. To overcome the limitations of studies that use one or two monitors to capture physiological strain, integrating additional sensors is necessary to obtain a comprehensive physiological profile to protect vulnerable worker populations.

The purpose of this pilot study was to evaluate the feasibility of various wearable devices for measuring core body temperature, fluid and electrolyte loss, and heart rate among agricultural workers during the summer season. Assessing the consistency of these devices under real field conditions, characterized by intense physical activity and continuous movement, is critical for determining their suitability for occupational heat monitoring and data collection. Furthermore, predictive models were developed to explore whether these physiological indicators can effectively correlate with and forecast heat-related risk within this worker population.

2 | Methods and Materials

2.1 | Measurement Devices

Three wearable devices were utilized for continuous physiological monitoring during fieldwork.

2.1.1 | CALERA Research

A total of ten noninvasive CALERA Research devices (Greenteg, Rümlang, Switzerland) were used to measure core body temperature. The device has the dimensions of 50 mm × 40 mm × 8.35 mm and weighs 12 grams. Prior to deployment, each CALERA device was fully charged using the USB-A magnetic adapter and connected via Bluetooth to the CALERA Research Tool for firmware updates, factory resets, and synchronization. Each unit was initialized through the Research Tool or the CALERA® smartphone application, where individual device IDs were assigned, and logging sessions were started remotely. During field use, devices recorded core body temperature and skin temperature data every second throughout the workday. At the end of the session, the sensors were retrieved and reconnected to the Research Tool to download the data in. csv format, which included calibrated outputs and time-stamped physiological measurements.

2.1.2 | Nix Hydration Biosensors

A total of ten Nix Hydration Biosensors (NIX, Medford, MA, United States) were used to monitor fluid and electrolyte loss. The device was used in this study to continuously monitor sweat rate, total fluid loss, electrolyte loss rate, total electrolyte loss, sodium loss rate, and total sodium loss every minute throughout the workday. Each Nix™ Pod was paired with its corresponding Nix™ patch and connected to the Nix™ Solo mobile application via Bluetooth on an assigned Android phone. After completion, the session was stopped in the app using the same paired phone, and the data were saved and exported in. xlsx format for analysis.

2.1.3 | Garmin Vívoactive 6 Watch

Physiological Activity: Ten Garmin Vívoactive 6 wristwatches (Olathe, KS, United States) were used to measure various physical activity parameters, including heart rate, cadence, distance, fractional cadence, enhanced speed, enhanced altitude, and GPS data at 1-second intervals during the workday. Before field deployment, each watch was configured using Garmin Express™ to ensure the latest firmware updates were installed and to assign a unique device ID. Custom activity profiles were created on each watch to track relevant parameters throughout the workday. Once initialized, the study team launched an "Activity" session directly from the watch interface, enabling time-stamped physiological data collection before participants began fieldwork. At the end of the workday, data collection was stopped and saved as. fit files on the device's internal storage. These files were then retrieved via USB connection to a laboratory computer, converted to. csv format using the FIT File Viewer online tool for further analysis, and organized by participant ID and collection date.

All devices listed above also featured onboard memory to ensure uninterrupted data collection during field activities.

2.1.4 | Supporting Equipment

Ten Doogee (Note 58, Shenzhen, China) Android smartphones running version 15 were used to initiate and stop the Nix device recordings, synchronize data, and store the collected measurements.

2.2 | Weather Data

High-resolution 1-min frequency WBGT data [32] were retrieved from a locally collected data station 32 km from the farm, downloaded from https://climate.ncsu.edu/. [Correction added on 20 June 2026, after first online publication: A new Reference 32 was added after publication; consequently, the subsequent references have been renumbered from 33–64 (previously 32–63).]

2.3 | Study Location and Participant Selection

Agricultural workers from local farms in eastern North Carolina (NC) were sought for this study. First, farmer permission was sought to recruit farm employees. Eleven farmers were asked, and four provided signed letters of support. Participants from the farm with the largest number of workers were recruited first. Of 47 workers, 34 were interested. The inclusion criteria for participants included adults aged 18–54 years, male or female, who worked outdoors as a farmworker for at least two weeks, and who were English or Spanish speakers. The exclusion criteria were self-reported recent illness, current pregnancy, or history of kidney disease, and current use of diuretics or presence/history of cardiac arrhythmia. Four participants did not meet the inclusion/exclusion criteria.

Health screenings were conducted up to one week before data collection by a registered nurse and a bilingual student researcher. Screening included assessment of heart rate and rhythm, as well as height and weight, to calibrate the sensors. The study team obtained written informed consent, reading the consent form aloud and answered questions. They also collected demographic surveys and explained the purpose of the sensors and data collection procedures. The demographic survey contained 29 questions, as outlined in the Supplemental Material document. The research team first developed the survey in English and back translation was used to develop the version in Spanish. The survey collected information on participants' age, gender, ethnicity, country of origin, education level, housing type, language proficiency, and agricultural work history, including years of experience, weekly work hours, crop type, and payment method. It also gathered details on hydration practices, clothing type, work conditions, and access to sanitation facilities. Given the limited sample size and pilot scope, data analyses were not stratified by age or sex. Incentives in the form of $50 gift cards were provided to the participants upon successful completion of one full workday wearing the sensors.

2.4 | Field Protocol

Field testing took place on one farm in eastern NC from July 15 to 17, 2025, encompassing three days of data collection, targeting one of the hottest months of the year. Data collection began at approximately 5 AM and concluded at approximately 4 PM each day, with varying start and end times for each participant. Each participant wore all three devices (Calera, Nix, and Garmin watch) for one full workday, and data were collected from 10 participants per day across three days. The devices were placed in their recommended positions by the respective manufacturers. The CALERA Research device was affixed to the chest using a company-provided adjustable chest strap, with the back side of the device in direct contact with the skin just below the sternum to ensure accurate temperature readings. The Nix hydration biosensor was placed on the outer bicep using the supplied adhesive. Before application, the participant's upper arm was cleaned with isopropyl alcohol wipes to remove oils and ensure proper adhesion and allowed to dry. The patch was then applied to a hair-free area, and the Nix™ Pod was attached by aligning the LED indicator with the circular mark on the patch until a click was heard, confirming a secure connection. Once attached, the recording session, referred to as a "workout," was initiated through the app by selecting the appropriate activity and location parameters. The Garmin Vívoactive watch was worn on the wrist to capture activity, heart rate, distance, enhanced speed, and enhanced altitude. All devices were sanitized before use and reuse, and new adhesives were applied for each participant.

2.5 | Participant Demographics

Demographics of participants were summarized in mean and standard deviation (SD) or frequency and percentage (%) as appropriate.

2.6 | Data Cleaning

Due to variations in the times when the devices were placed on and removed from the farmworkers, a fixed time window from 6:30 AM to 3:30 PM was selected for each day. This resulted in a total of 540 measurements per participant across all variables, except for the hydration-loss variables, as the Nix sensors require an accumulation of sweat before recording initiates. The mean delay before the Nix sensor generated usable data was 71.6 min (SD = 30.3 min, Median = 71.0 min), with a minimum observed delay of 20 min and a maximum of 155 min. Consequently, this yielded between 385 and 520 usable data points per participant each day, with an average of 468 min of recorded hydration data. Additionally, further data cleaning was necessary based on visual inspection to identify steady-state values per participant. The total electrolyte loss, electrolyte loss rate, and total sodium loss showed a very strong correlation ($r = 0.99$) with other hydration loss variables and were therefore identified as redundant and excluded from further analysis. For the Garmin data, and due to substantial missing data stemming from sensor thresholding and algorithmic limitations inherent to accelerometry [33, 34], the cadence and fractional cadence variables were excluded from the final dataset. Finally, to ensure data quality, outlier and invalid measurements were removed according to predefined criteria. For core body temperature, all zero values were replaced with 'Not a Number' (NaN), and readings below 34°C or above 42°C were considered physiologically implausible and excluded. The implausible or missing values accounted for only 0.82% of the total dataset. Analysis of these removed values revealed a non-random pattern, where 100% of the missing data occurred consecutively at the immediate onset of the recording period, indicating an initial sensor stabilization phase rather than random signal loss. Heart rate values below 40 beats per minute (bpm) or above 200 bpm were also deemed invalid and replaced with NaN. Similarly, sweat rate values exceeding 5000 mL/h were identified as unrealistic and excluded from further analysis. GPS altitude data showed clear errors, with unrealistic values falling far outside the farm's true elevation range (15–27 m). Therefore, any altitude measurements below 15 m or above 30 m were removed or replaced with NaN to exclude erroneous GPS readings. Speed values above normal walking pace ($\approx$1.5–1.7 m/s) are likely errors or moments when participants were in a vehicle. Therefore, any speed measurement greater than 2 m/s was replaced with NaN to exclude non-walking, non-work-related activity from the analysis.

2.7 | Data Reliability

The current pilot study did not assess data reliability due to funding limitations, so no reference monitors or methods were available to assess core body temperature, hydration loss, and heart rate. However, these devices have been evaluated with reference instruments and methods in previous studies. In both real-world and controlled settings, the CALERA core temperature sensors demonstrated high agreement with reference methods. In a field study involving 14 participants monitored for 3–5 days [25], the CALERA Research sensor overestimated mean core body temperature by approximately 0.10°C compared with a wired rectal probe during nocturnal sleep. In a specialized sleep laboratory, Xu et al. [35] found that the CALERA showed a mean bias of −0.06°C compared to an ingestible gastrointestinal capsule, with 95% limits of agreement of ±0.51°C. Similarly, in a climate-controlled environment (32°C) [24], the CALERA Research Sensor, tested on 15 cyclists over a 95-min rest–exercise–recovery protocol, exhibited a negligible mean bias of 0.01°C and 95% confidence limits of ± 0.36°C relative to gastrointestinal temperature measurements. Verdel et al. [36] reported that the sensor achieved high reliability, with a mean bias of only 0.02°C between trials, though it showed greater variability when compared strictly with rectal thermometers during intense cycling. Finally, Jolicoeur Desroches et al. [37] noted that the sensor may underestimate core temperature during rapid internal cooling, such as cold water ingestion, or high-intensity athletic exertion. However, for occupational monitoring where workload is more continuous and less explosive than competitive sports, these deviations remain within an acceptable range for feasibility assessment.

The Garmin watch has also been used and evaluated in various settings [30]. In clinical and laboratory evaluations, Garmin heart rate monitors demonstrated varying accuracy across settings and activities when compared with electrocardiography. In a hospital post-anesthesia care unit, Helmer et al. [19] found that the Garmin underestimated heart rate by a mean bias of −1.21 beats per minute (bpm) relative to ECG. In a cardiac rehabilitation facility, the Garmin overestimated heart rate with a mean bias of +3.8 bpm overall, increasing from +3.6 bpm on the treadmill to +5.9 bpm on the stationary cycle [38]. Laboratory testing of the Garmin watch on healthy volunteers revealed activity-dependent variation [20], showing minimal bias on the treadmill (−0.3 bpm), modest overestimation on the stationary bicycle (+ 3.4 bpm), and a pronounced overestimation on the elliptical trainer without arm use (+ 7.7 bpm). For the Nix, no current studies have used or evaluated the sensor for primary data collection.

2.8 | Modified Physiological Strain Index (mPSI)

Using the physiological data from the wearable sensors, the mPSI was calculated using the following equation, which was adopted from Mac et al. [39] (Equation 1):

$$mPSI_i = 5\frac{(temp_i - temp_{\min})}{(39.5 - temp_{\min})} + 5\frac{(heart\ rate_i - heart\ rate_{\min})}{(180 - heart\ rate_{\min})}$$

where

$$temp_i = is\ the\ temperature\ in\quad °C$$

$$\mathrm{temp}_{\min} = \text{is the participant-specific minimum temperature in °C calculated from one-minute averaged data.}$$

[Correction added on 20 June 2026, after first online publication: The physiological data for "$\mathrm{temp}_{\min}$" was revised for clarity.]

$$\mathrm{heart\ rate}_i = \text{is the heart rate in beats per minute}\quad \text{(bpm)}$$

$$\mathrm{heart\ rate}_{\min} = \text{is the minimum heart rate in beats per minute (bpm) collected for the participant at an averaging time of one minute}$$

Because mPSI incorporates both core temperature and heart rate, two of the most sensitive indicators of heat stress, it provides a reliable measure of the combined thermal and cardiovascular burden. The mPSIs were translated into a corresponding heat strain risk level as follows: no/little heat strain (0–2), low heat strain (3–4), moderate (5–6), high (7–8), and very high (9–10) [16].

2.9 | Data Analysis

The data were time-paired with a mean interval of 1 min using MATLAB version 24.2.0.2863752 (R2024b), Update 5. Additionally, mean values for 15 min and 1 h were computed to better understand the results. Box-and-Whisker plots were created for WBGT, physiological parameters, and mPSI, showing variation across days and participants. The physiological parameters used were core body temperature, skin temperature, sweat rate, total fluid loss, sodium loss rate, heart rate, distance, enhanced speed, and enhanced altitude.

2.9.1 | AIHA App Risk Analysis Comparison With mPSI

Environmental heat metrics are essential for contextualizing physiological responses to heat exposure. The heat index integrates air temperature and relative humidity to estimate perceived ambient heat [40], while the Wet Bulb Globe Temperature (WBGT) incorporates dry-bulb, wet-bulb, and globe temperatures to account for ambient conditions, humidity, radiant heat, and air movement [41]. Although WBGT is traditionally measured in occupational settings using dedicated heat stress monitors, several mobile applications estimate WBGT using meteorological data. One such tool is the AIHA Heat Stress App, developed by the American Industrial Hygiene Association, which calculates current and forecasted WBGT values using location-specific weather inputs and converts these values into categorical heat stress risk levels based on workload assumptions [42, 43]. While these applications are useful for daily heat exposure planning among outdoor workers, they generalize environmental conditions at the area level and do not capture inter-individual variability in physiological heat strain.

The AIHA Heat Stress App (version 2.1.0) was used to collect data on current hourly ambient temperature (°C), relative humidity (%), heat index (°C), wet bulb globe temperature (WBGT, °C), and WBGT-derived risk levels each monitoring day based on both moderate and heavy workload assumptions. App data was recorded at the start of every hour from 7:00 AM to 3:00 PM. The zip code corresponding to the study site was used in the app to obtain weather data that was specific to the study site location. The WBGT values calculated by the app were based on the following constant settings: (1) 'outdoors' for the work site; (2) 'work clothes' for clothing worn; and (3) 'none' for non-solar heat irradiance. The app settings used for cloud coverage were based on the hourly cloud forecast, which was either "mostly cloudy" or "cloudy". Cross-tabulation was used to analyze the relationship between risk level categories based on the app WBGT and mPSI. For this comparison, two mPSI data sets were calculated: one used the hourly mean CBT, and the other used the hourly maximum CBT as CBTi in equation 1.

2.9.2 | Geospatial Analysis

For each participant, activity pathways were converted from FIT to GPX format. GPX files were integrated into ArcGIS Pro 3.5.3 (Esri, Inc.) to create point feature classes. Physiological data were integrated into the geodatabase by joining the minutely data to the point features. Site-specific polygons were created using a basemap to visualize fields and buildings that participants worked in, the entrance to the farm (i.e., entry), and the roads that participants traversed. The entrance/entry to the farm included the parking area and footpaths surrounding the farm office, sheds, and other equipment barns. All other locations where participants were observed were classified as "Other", which occurred when participants were standing adjacent to one of the aforementioned locations. These polygons were used to create a database of physiological data based on location. A boxplot and time series of mPSI data for pooled participants were created using the R Statistical Software and RStudio in conjunction with the 'ggplot2', 'dplyr', 'cowplot', and 'readxl' packages.

2.9.3 | Regression and Prediction Models

Similar to correlation analysis, prediction models were used to analyze mPSI in relation to WBGT and physiological parameters.

2.9.3.1 | Moving Averages

Machine learning models require a set of input features to predict the target variable, mPSI. In this study, the input features include sensor-based measurements discussed in other sections (e.g., WBGT and heart rate). In addition to the raw sensor readings, we incorporated moving-average features as supplementary model inputs.

The inclusion of moving averages is motivated by the potential lagged effects of certain physiological or environmental variables on mPSI. For example, an increase in heart rate may not immediately correspond to a change in mPSI, and the effects may appear after a delay (e.g., within 15 min). To allow the model to capture these temporal dependencies, we computed moving averages of the input features across multiple time windows. Specifically, moving averages were calculated using 15-min, 30-min, 60-min, and 90-min windows for all features. These additional temporal features provide the model with smoothed trends and short-term historical context, potentially improving its ability to learn time-dependent relationships relevant to mPSI dynamics.

2.9.3.2 | Correlation Between Study Variables

The R^2 was calculated with respect to mPSI and included skin temperature, sweat rate, total fluid loss, sodium loss rate, heart rate, distance, enhanced speed, and enhanced altitude. Heart rate correlations were also calculated because both the correlation and prediction models will use heart rate as an input, and this variable is reliably available from modern smartwatches. Devices such as the CALERA Research and Nix may not be widely accessible. Discrete skin temperature data are already available on the Garmin watch when the device is worn continuously overnight during sleep, and other devices that are not readily available on the market provide continuous data through wrist-worn devices [44, 45]. Box-and-whisker plots of the Pearson correlation coefficients (r) with mPSI were developed for each physiological parameter and each moving average to illustrate the range and distribution of r values across all participants.

2.9.3.3 | Model Evaluation via Holdout Testing

The performance of all machine learning models was evaluated using a holdout testing approach. In this method, the dataset was randomly divided into training and testing subsets, consisting of 13,992 and 3240 records, respectively. The training subset was used for model fitting, while the testing subset was used for performance evaluation, ensuring that the reported results reflect the model's ability to generalize new data.

2.9.3.4 | Multi-Linear Regression

For the purpose of predicting mPSI, participant-level Pearson correlation coefficients between mPSI and the predictors of skin temperature, hydration, heart rate, and movement, as well as their 15-min, 30-min, 60-min, and 90-min moving averages, were calculated separately and summarized as means and 90% confidence intervals.

Next, participant-level multi-linear regression models were run separately to predict mPSI using skin temperature, hydration, heart rate, and movement, as well as their moving averages. For each participant, five multi-linear regression models were run, one using all original predictors and four using all moving average predictors for 15-min, 30-min, 60-min, and 90-min. The t-statistics of the predictors were summarized in means and 90% confidence intervals.

Lastly, the prediction accuracy of the participant-level multi-linear regression models was summarized using model root mean squared error (RMSE) and model R^2. The model RMSEs were summarized as means and 90% confidence intervals, and the model R^2 values were summarized using box-and-whisker plots. [Correction added on 20 June 2026, after first online publication: In the preceding sentence, "R^2's" has been corrected to "R^2 values" in this version.] We considered both models with and without the hydration variables from the NIX sensors, as they may not be readily available in the future. The training sets were used to fit the models, while the holdout testing sets were used to evaluate model prediction accuracy. For each participant, five multi-linear regression models, as described above, were run with the NIX variable and another five were run without the NIX variable.

2.9.3.5 | Building Machine Learning Models

Two machine learning models, Random Forest (RF) Regressor and Gradient Boosted Regressor (GBR), were implemented in this study to predict mPSI [46, 47]. These models were selected because ensemble tree-based methods have demonstrated strong predictive performance and robustness, particularly when applied to small and medium-sized datasets. In contrast, more complex architectures, such as transformer-based models, typically require substantially larger training datasets to achieve comparable performance and to avoid overfitting.

Both RF and GBR are ensemble learning algorithms that combine the outputs of multiple decision trees to improve generalization. The fundamental difference between the two lies in their ensemble strategies: the RF algorithm uses bagging (bootstrap aggregating), where individual trees are trained independently on random subsets of the data, and predictions are averaged to reduce variance. In contrast, GBR uses a boosting strategy, where trees are built sequentially, with each new tree attempting to correct the residual errors of the ensemble's previous iterations. In general, RF models tend to be more resistant to overfitting and provide stable results across different datasets. GBR, while more sensitive to overfitting, can often achieve higher predictive accuracy on smaller or more complex datasets when appropriately regularized and tuned.

3 | Results

3.1 | Device Reliability

A total of 30 participants contributed 1-min data on physiological and movement metrics on July 15–17, 2025, collected between 6:30 AM and 3:30 PM, are summarized in Supporting Table S1. The CALERA devices achieved 100% data retrieval across all participants throughout the study. The Nix sensors experienced issues retrieving data from the mobile application on Android devices; as a result, data were successfully collected from 24 participants, corresponding to an overall retrieval rate of 80%. Additionally, one participant only had sweat rate data available. For the Garmin watches, one participant accidentally turned off data collection, resulting in a 97% retrieval rate during the study. Additionally, four participants disabled the GPS function on the Garmin watch, which affected speed and altitude. Additionally, another participant lacked altitude data, resulting in 26 participants with speed data and 25 with altitude data.

The number of 1-min data points for each metric and participant is shown in Supporting Table S2. Core body temperature, skin temperature, and heart rate showed near-complete data capture, with most participants contributing 537 to 540 min of data for each metric. Hydration-related metrics demonstrated more variability across individuals. Sweat rate, total fluid loss, and sodium loss rate each showed reduced data availability for several participants. Seventeen participants contributed full or near full hydration datasets (ranging from 346 to 487 min, depending on individual runtime), while the remaining participants had no hydration data recorded for these metrics. The absence of hydration data for specific participants reflected known device connectivity issues that occurred during field deployment rather than participant noncompliance.

Movement-related data were consistently available across most individuals. Distance, enhanced speed, and enhanced altitude each had between 401 and 540 data points per participant. Enhanced altitude and enhanced speed showed slightly greater variability, particularly for Participants 13, 15, 17, 20, 22, and 24, who had reduced data due to intermittent GPS signal loss. Despite these instances, the movement sensors generally provided stable data streams across the study period. Overall, most participants contributed complete physiological and movement datasets, whereas hydration metrics exhibited the greatest variability in completeness. These patterns reflect differences in device reliability among sensor types during field deployment.

TABLE 1 | Participant demographics ($n = 30$).

Variable	Level	*n* (%)
Age (years)	Mean (SD)	32.5 (6.0)
Sex	Female	3 (10.0)
	Male	27 (90.0)
Marital status	Married	17 (56.7)
	Single/Divorced	13 (43.3)
Education level	Some college or more	6 (20.0)
	High school graduates	16 (53.3)
	Some high school or less	8 (26.7)
English level	Intermediate	6 (20.0)
	None/low	24 (80.0)
Years worked in agriculture	< 1	3 (10.0)
	1–5	21 (70.0)
	> 5	6 (20.0)
Weeks worked this season	≤ 20	22 (73.3)
	> 20	8 (26.7)
Days worked this week	5	17 (56.7)
	6	11 (36.7)
	7	2 (6.7)
Work type	Irrigation	2 (6.7)
	Loading	2 (6.7)
	Pruning	9 (30.0)
	Spraying	6 (20.0)
	Tobacco	11 (36.7)
Use double-layered knit clothes	No	21 (70.0)
	Yes	9 (30.0)
Use overalls	No	26 (86.7)
	Yes	4 (13.3)
Distance to nearest bathroom	3-min or less walk	18 (60.0)
	> 3-min walk	12 (40.0)
Baseline heart rhythm (BPM)	Mean (SD)	74.5 (12.7)

3.2 | Participant Demographics

The demographic results from the survey data collected are shown in Table 1. A total of 30 agricultural workers participated in the study across three consecutive days of data collection, with an equal distribution of participants each day.

The mean age of the participants was 32.5 years (SD = 6.0). The sample included three female workers (10.0%), with the majority identifying as male. More than half of the participants reported being married (56.7%). By education, the majority had completed high school or less (80%), while 20% had attained some college education or higher. English proficiency varied, with 20% reporting intermediate English and others reporting basic or limited proficiency.

Workers' agricultural experience was diverse: 10% had less than 1 year of experience, while the majority had multiple years of experience. Most participants (73.3%) worked many weeks during the current agricultural season, and more than half (56.7%) worked 5 days in the week of data collection. Participants represented several work-task types, including irrigation and loading work (each 6.7%), with additional tasks accounting for the remaining workforce, including ~37% tobacco workers, who constituted most participants.

Regarding occupational safety practices, 30% reported having completed work-mandated hydration training, while 43.3% stated they drink water every 15–20 min as recommended. Most workers reported not using double-layered knit clothing (70.0%) or overalls (86.7%). The majority (60%) reported that the distance to the nearest bathroom was 3 min or less. The mean baseline heart rate was 74.5 beats per minute (SD = 12.7).

3.3 | Environmental and Participants' Physiological Changes

The dataset included 30 outdoor farmworkers, each with 9 hourly records collected in July, for a total of 270 hourly observations across all participants. The physiological data summary for all 30 participants is shown in Supporting Table S3.

Environmental heat exposure, measured by WBGT, remained consistently high throughout the three monitored days, with only minor daily variations. Median hourly WBGT was 28.2°C on July 15, 27.8°C on July 16, and 28.5°C on July 17, with overall ranges spanning approximately 23°C–33°C. The WBGT box plot by day (Figure 1) showed compact distributions within each day and consistently elevated medians near 30°C, indicating continuous exposure to high environmental heat. July 17 displayed the highest minimum WBGT, reflecting more sustained warm conditions throughout that day compared to the previous 2 days.

The measured physiological parameters and calculated mPSI for all the participants are shown in Figure 2. The core body temperature ranged from 36.39°C to 39.29°C (mean: 37.38°C, SD: 0.55, $n = 16{,}067$). According to the National Institute for Occupational Safety and Health (NIOSH), core temperature during hot work should remain below 38.0°C. In this dataset, 12.6% of minute-level core temperature observations exceeded this recommended limit. Among participants who exceeded 38.0°C, time spent above the threshold ranged from 17 to 449 min, with a median of 71 min. For the higher threshold of 38.5°C, affected participants spent 11 to 403 min above the limit, with a median duration of 42 min, indicating substantial variability in heat strain across participants. Skin temperature showed greater variability than core temperature, ranging from 25.46°C to 37.18°C (mean: 34.49°C, SD: 1.22, $n = 16{,}200$). Heart rate ranged from 40 to 156 bpm (mean: 91.86 bpm, SD: 15.97, $n = 15{,}660$). According to the American Conference of Governmental and Industrial Hygienists (ACGIH), heart rate should remain below 180 bpm for workers with normal cardiac function to reduce the risk of heat-related illness. In this dataset, 0% of heart rate observations exceeded this guideline. The modified Physiological Strain Index (mPSI) ranged from 0.36 to 7.55 (mean: 2.78, SD: 1.29, $n = 16{,}067$). Compared to this scale, 8.3% of the mPSI observations exceeded 5, a threshold commonly associated with elevated physiological strain in the mPSI based on the literature [48, 49].

Sweat rate varied from 196.25 to 1305.40 mL/h (mean: 308.26 ml/h, SD: 190.99, $n = 10{,}227$). Using the NIOSH (2016) standard that sweat loss should stay below 1.2 kg/h (~1200 mL/h), 4.2% of sweat-rate observations exceeded this recommended threshold. Total accumulated fluid loss ranged from 10.60 to 4826 mL (mean: 1295.97 mL, SD: 885.29, $n = 9067$). Sodium loss rate ranged from 29.00 to 1019.35 mg/h (mean: 227.13 mg/h, SD: 135.24, $n = 9761$). Total sodium loss over the work shift, defined as the final cumulative value for each participant, ranged from 844 to 3150 mg (mean: 1787.5 mg, SD: 719.6 mg). Replacement of sodium lost in sweat is

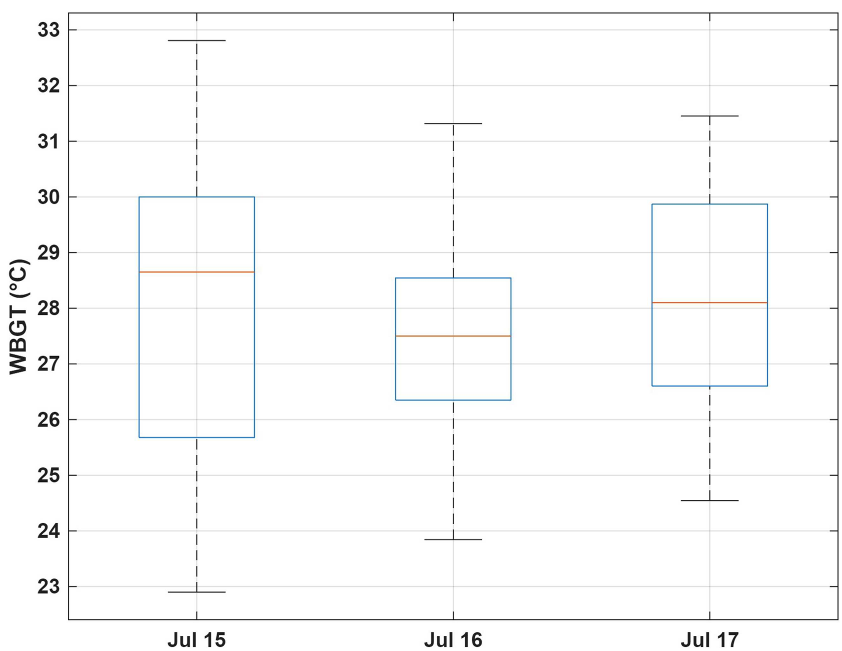

FIGURE 1 | Box-and-whisker plot illustrates the WBGT variability during the workday for each day. The data was based on 1-min mean values.

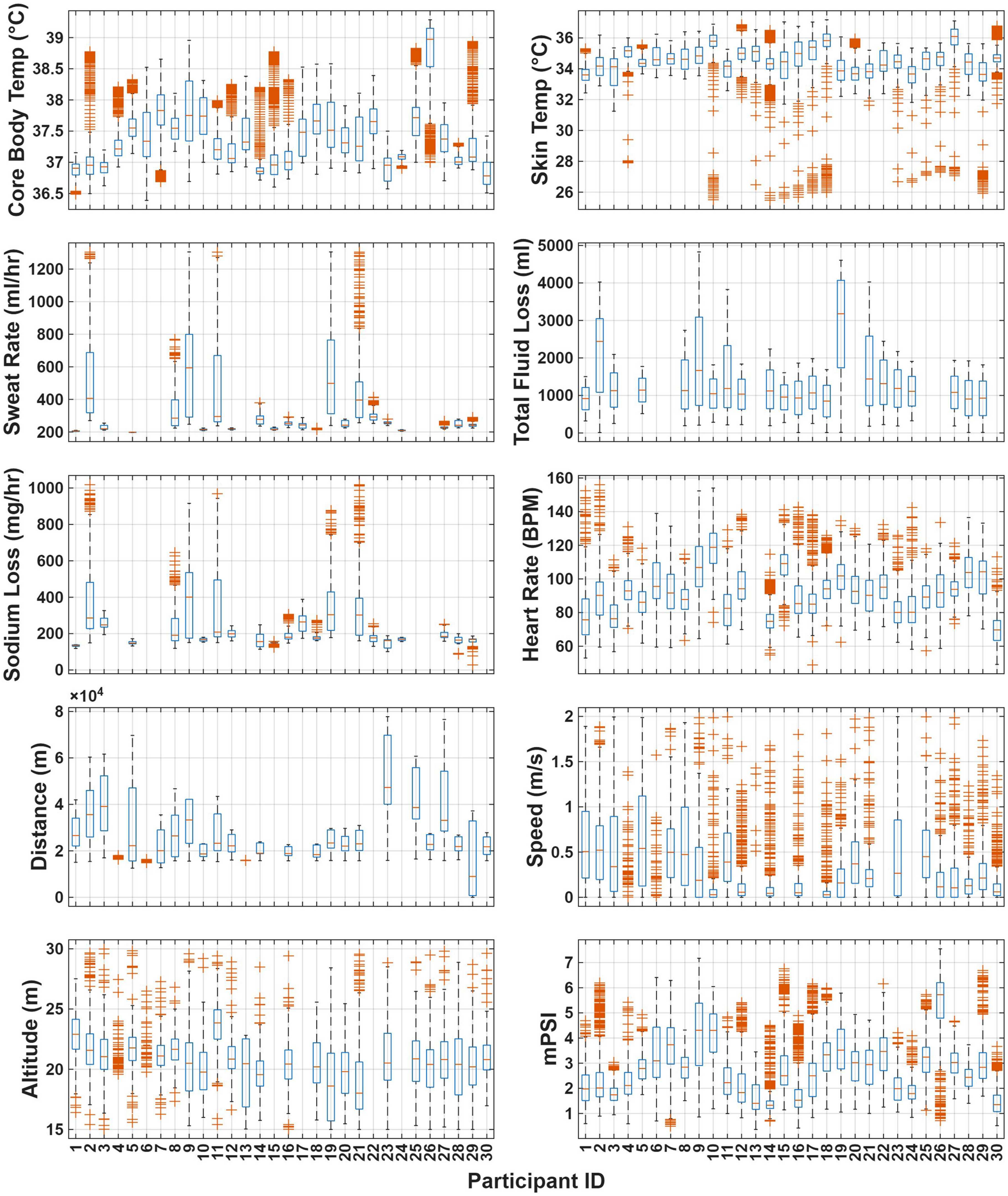

FIGURE 2 | Box-and-whisker plots illustrating the variability for the key identified physiological factors across all participants. The data was based on 1-min mean values.

recommended when the duration of exercise is longer than 2 h, when loss is greater than > 3–4 g per shift [50, 51].

Accumulated distance traveled ranged from 0 to 77,770 m (mean: 26,400 m, SD: 13,056, $n = 16{,}040$). These mobility metrics reflect total participant movement during monitoring periods, including both on-foot activity and vehicle travel between work locations. Enhanced speed ranged from 0 to 1.99 m/s (mean: 0.38 m/s, SD: 0.45, $n = 13{,}317$). The enhanced altitude ranged from 15.0 to 30.0 meters (mean: 20.96 m, SD: 2.41, $n = 12{,}445$). This metric captured changes in vertical movement or terrain exposure during work.

3.4 | AIHA App Risk Analysis Comparison With mPSI

Heat stress risk levels based on the hourly app WBGT were compared to the corresponding heat strain risk levels based on the hourly mean mPSI. The percentage of each risk level category assigned to the two data sets of WBGT from the app by workload (moderate and heavy) and to the calculated two data sets of mPSI by CBT used (mean or maximum CBT) are shown in Figure 3. The most common risk level was "no/little/minimal risk" for mPSI calculated using mean CBT (mPSI-meanCBT, 45.2%) and "low risk" for mPSI calculated using max CBT (mPSI-maxCBT, 46.7%). Both mPSI data sets had very small

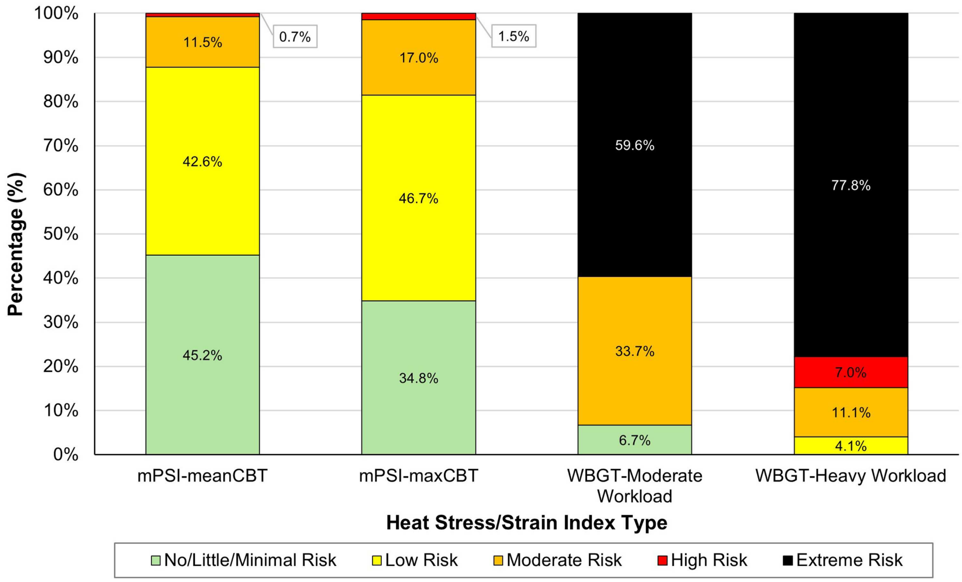

FIGURE 3 | Percentage of assigned heat stress or strain risk level categories ($n = 270$) for app-based WBGT and sensor-based mPSI.

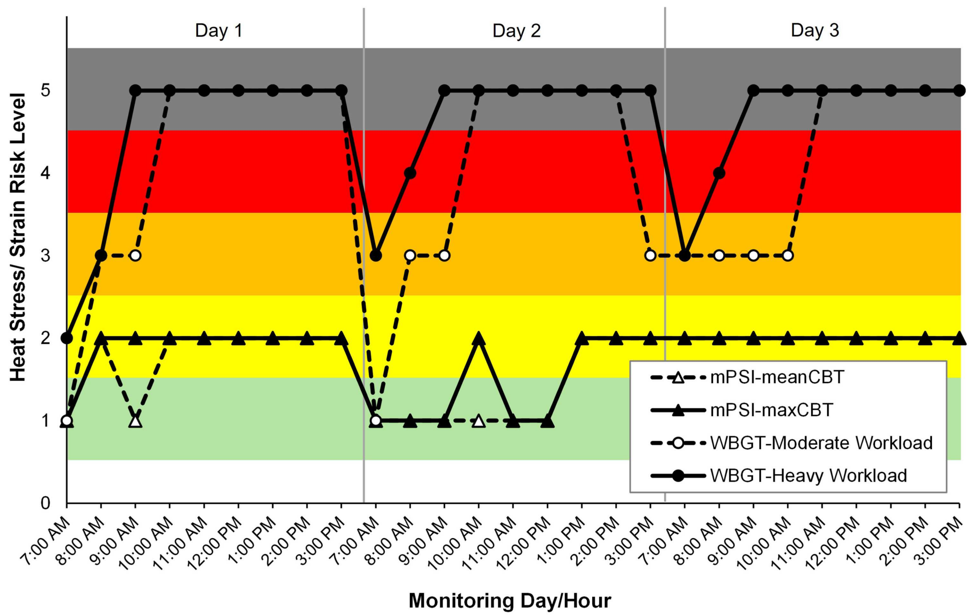

FIGURE 4 | Hourly heat stress or strain risk level by heat stress or strain index type. The numerical risk level corresponds to the qualitative risk level: 1 – no/little/minimal; 2 – low; 3 – moderate; 4 – high; 5 – extreme.

percentages of "high risk" level categories (11.5% and 17.0% for mPSI-meanCBT and mPSI-maxCBT, respectively) and had 0% assigned to "extreme risk". In contrast, the most common risk level was "extreme risk" for both WBGT data sets, regardless of the workload assumption (59.6% and 77.8% for moderate and heavy workloads, respectively), while the least common risk levels were "low risk" for WBGT under moderate workload (0%) and "no/little/minimal risk" for WBGT under heavy workload (0%). By workload, there are higher percentages of the more severe risk levels under the heavy workload assumption (77.8% for extreme risk and 7.0% for high risk) than under the moderate workload assumption (59.6% for extreme risk and 0% for high risk).

The hourly heat stress or strain risk levels assigned to each heat stress/strain index (mPSI and WBGT) for all monitoring days are shown in Figure 4. The qualitative risk levels of no/little/minimal, low, moderate, high, and extreme risks were assigned to numerical risk levels from 1 to 5, respectively. For mPSI, hourly qualitative risk levels were derived from the hourly mean mPSI for each monitoring day. The hourly heat stress risk levels based on WBGT, regardless of the workload assumption, ranged from 1 (no/little/minimal) to 5 (extreme) and were found to be consistently higher than the hourly heat strain risk levels based on mPSI, regardless of the CBTi used, which ranged from 1 (no/little/minimal) to 2 (low). For example, on monitoring day 2 at 12 noon, the WBGT-based risk level was at the most severe risk (i.e., extreme) while the mPSI-based risk level was at the lowest risk (i.e., no/little/minimal).

3.5 | Geospatial Analysis

Minutely heat strain risk data exhibited spatial patterns and were more variable than the hourly mean mPSI trends (Figure 5). Mean mPSI was greatest when participants were in a field (2.9 ± 1.7), followed by when participants were traveling along a road (2.5 ± 1.5) or at the entrance to the farm (2.4 ± 1.0). Overall, mPSI ranged from 0.5 to 7.55 (Figure 5A), with the maximum occurring in a field (Figure 5A). Mean and median mPSI values suggest that most participants were not at risk for heat strain/stress (Figure 5A). However, 57% of mPSI observations were ≥ 2, indicating that participants experienced at least low heat strain/stress for most of their workday. Furthermore, some participants experienced moderate to high risk during their workday, particularly when they were working in a field, building, or traveling along a road (Figure 5A). There were 3 occurrences where one individual experienced high heat stress risk (mPSI range: 7.0–7.55), which occurred in a field at

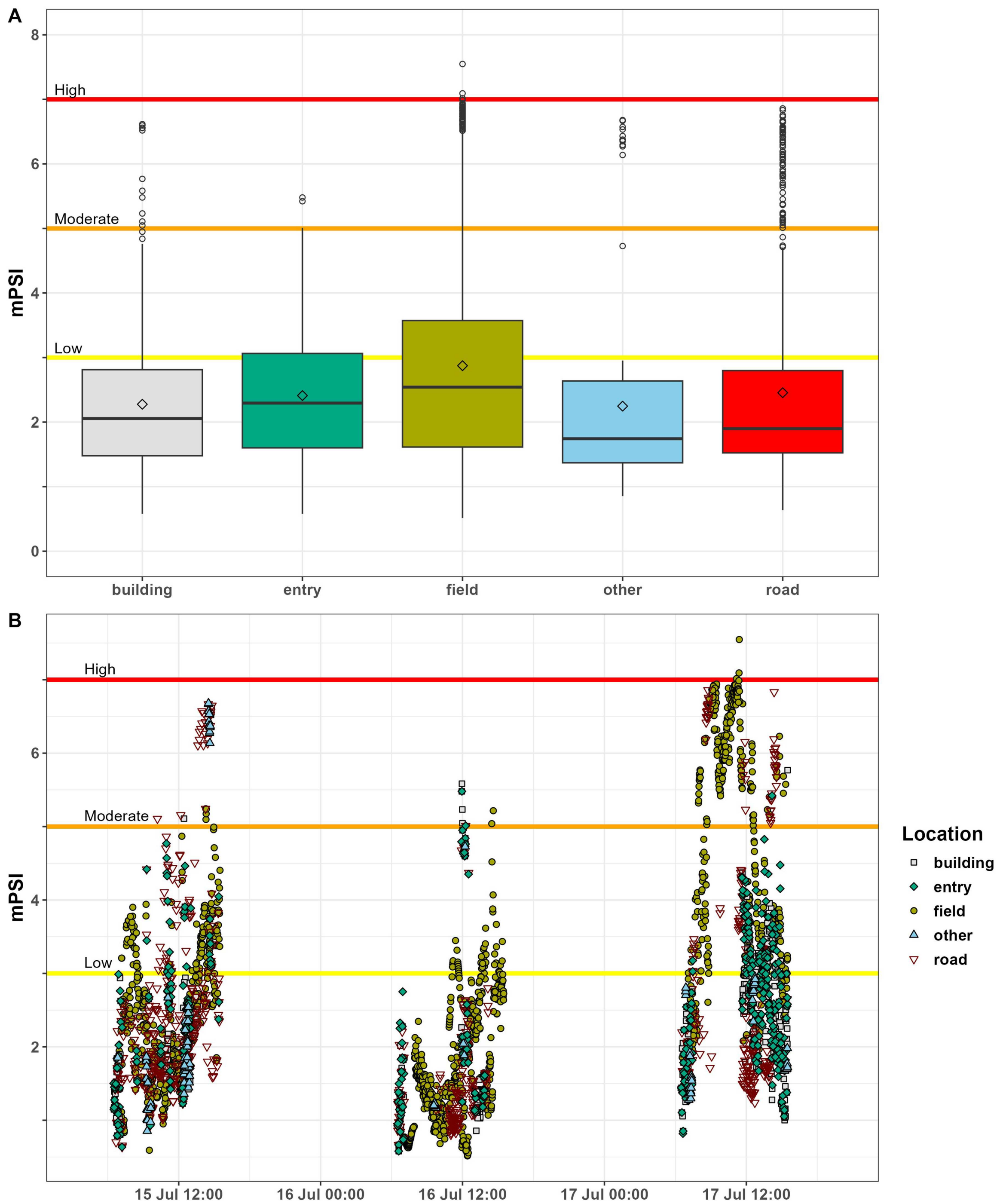

FIGURE 5 | (A) Boxplot of participant mPSI relative to location on farm. (B) Time series of mPSI categorized by location.

approximately 11:15 AM on 17 July (Figure 5B). From 15 to 17 July, there were 316 instances where participants experienced moderate heat stress (mPSI range: 5.0–6.9) and most of these occurrences were observed in a field (70%) or while on a road (23%). These data show that individuals' heat strain/stress risk level can vary considerably throughout the day and depending on their location.

3.6 | Correlation and Model Prediction

3.6.1 | *Correlation Between Study Variables*

Participant-level linear regressions were performed between each predictor and mPSI across five smoothing conditions (NO_MA, MA15, MA30, MA60, and MA90). For each variable and participant, R^2 values were computed, and median values were compared between the NO_MA and MA90 windows to quantify the impact of smoothing. Box and Whiskers plots were created to show the R^2 variation among participants for different moving averages, as shown in Figure 6. The results showed substantial variability in how moving-average smoothing influenced the strength of association between predictors and mPSI. Under the unsmoothed NO_MA condition, heart rate exhibited the strongest association with mPSI (median $R^2 = 0.60$), followed by total fluid loss (0.35) and skin temperature (0.21). Movement-related variables and WBGT showed weaker relationships, with median R^2 values generally below 0.15. When smoothing was applied, the magnitude and direction of change in median R^2 varied widely across variables.

Several predictors showed modest increases in median R^2 as smoothing windows increased. Skin temperature increased from 0.21 to 0.23, representing a 13.8% increase, and total fluid loss increased by 6.3%. Other variables showed larger percentage increases, such as sweat rate (+ 52.9%), distance (+ 26.7%), and enhanced altitude (+ 70.6%). The largest relative increase was observed for enhanced speed, which increased from a very low NO_MA median (0.01) to 0.10 under MA90, corresponding to a 1430% increase; however, despite the large percentage change, its absolute predictive value remained low. Not all variables benefited from smoothing. Heart rate, despite being the strongest predictor in the NO_MA condition, showed a 24.2% decrease in median R^2 after smoothing ($0.60 \rightarrow 0.46$). Sodium loss rate also showed a modest decrease of 5.3% in its median R^2 from NO_MA to MA90. These decreases reflect cases in which smoothing removed short-term variability that was informative in explaining fluctuations in mPSI.

3.6.2 | *Multi-Linear Regression With Backward Elimination to Predict mPSI*

The Box-and-Whiskers plot of R^2 and RMSE for different moving averages is shown in Figure 7. A total of ten multi-linear regression models were developed for each participant. Five models included hydration variables from the NIX sensors, and five excluded them. Within each set, the models corresponded to the five smoothing windows: NO_MA, MA15, MA30, MA60, and MA90. Model performance was evaluated using R^2 and RMSE from the holdout testing sets. When NIX hydration variables were excluded, the models demonstrated moderate to strong predictive performance across participants. Under the NO_MA window, the mean R^2 was approximately 0.68 and the mean RMSE was 0.48. Applying smoothing changed the model behavior across the windows. The MA15 window resulted in a slight reduction in R^2 and a modest decrease in RMSE to 0.48. The MA30 window improved model fit, with R^2 increasing and RMSE decreasing to 0.43. The MA60 window provided the greatest overall improvement, where the mean RMSE decreased to 0.39, the lowest error observed for this model type, and the mean R^2 reached one of the highest values in the dataset at approximately 0.75. [Correction added on 20 June 2026, after first online publication: In the preceding sentence, the RMSE values was revised from 0.3976 to 0.39 to adjust the decimal value.] The MA90 window showed greater variability across participants, reflected in higher RMSE values at 0.43 and less consistent R^2 behavior. Based on the combination of higher R^2 and lower RMSE, the MA60 window produced the best performance for models without hydration variables.

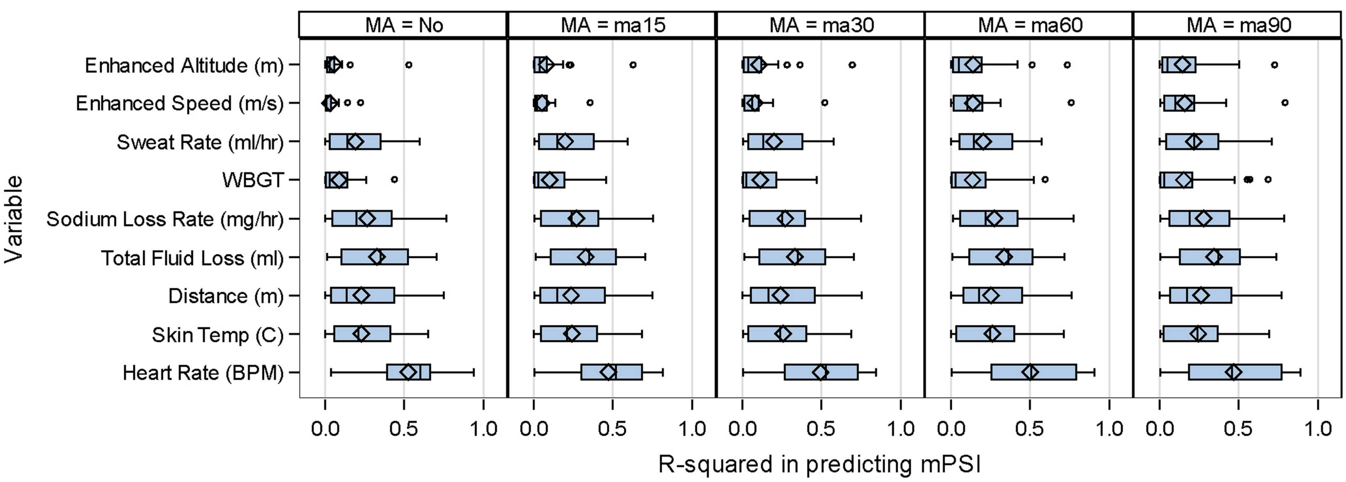

FIGURE 6 | Box-and-whiskers R^2 range for all participants for all the identified physiological variables in the study in predicting mPSI. The data was based on 1-min mean values. MA## = moving average of physiological data over the past ## minutes, while 'No' = current minute mean values. The box and whiskers show the minimum, 1st quartile, median, 3rd quartile, and maximum, while the diamond is the mean.

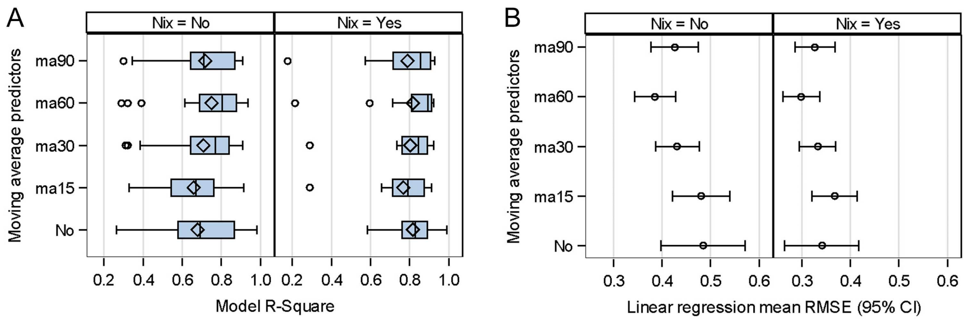

FIGURE 7 | Multiple linear regression model performance for predicting mPSI with and without Nix variables. (A) Model R^2 distributions across participants for each smoothing window. (B) Linear regression mean root mean square error (RMSE) with confidence intervals for each smoothing window. MA## = moving average of physiological data over the past ## minutes, while "No" = current minute mean values. Smaller RMSE values and higher R^2 values indicate better prediction. For Panel A, the box and whiskers show the minimum, first quartile, median, third quartile, and maximum, while the diamond indicates the mean.

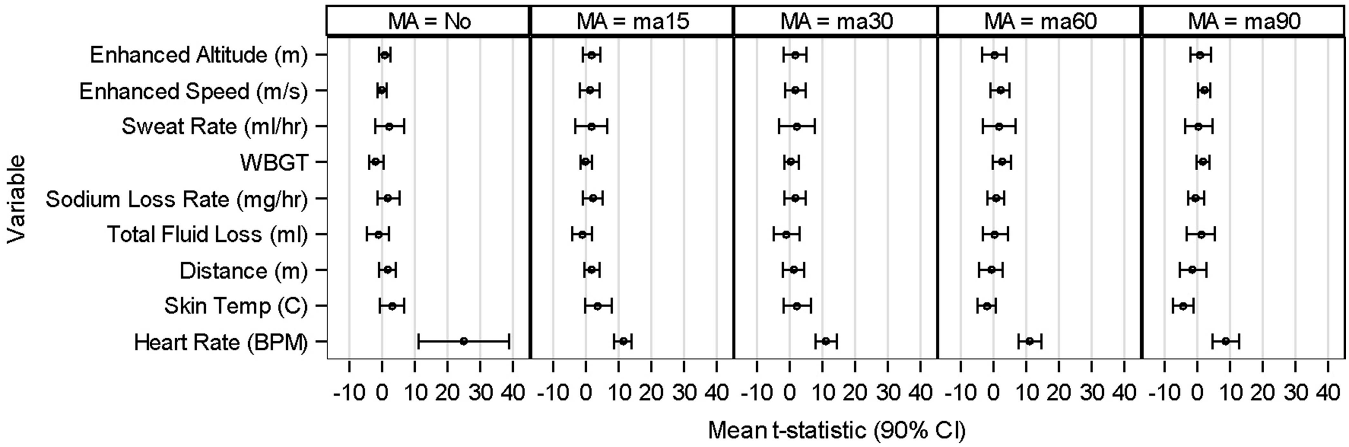

FIGURE 8 | Mean *t*-statistics (90% CI) of predictors in each multiple linear regression model predicting mPSI. MA## = moving average of physiological data over the past ## minutes, while 'No' = current minute mean values. *t*-statistics away from zero show stronger predictive effects.

The models that included hydration variables followed a similar pattern, although they consistently achieved lower RMSE values across all smoothing windows. Under the NO_MA condition, the mean RMSE for models with NIX variables was 0.34, which was already considerably lower than that of the models without NIX. Smoothing initially reduced performance at the MA15 window, where RMSE increased to 0.37, and R^2 did not improve. At MA30, RMSE decreased to 0.33 and R^2 increased relative to MA15. The MA60 window again produced the best results, with RMSE decreasing to 0.30, which was the lowest error across all models and smoothing windows, while R^2 reached its highest values for the NIX model set. At MA90, RMSE rose slightly to 0.33 and R^2 became more variable. Across all windows, the inclusion of hydration variables lowered RMSE by approximately 20% to 30% relative to the corresponding models without hydration variables, although R^2 values remained similar between the two model types.

The mean (90%) *t*-statistics for the predictors shown in Figure 8 confirmed the strong predictability of heart rate and its moving averages to mPSI across all participants. However, the predictability of other predictors varied considerably, underscoring the need for participant-level models rather than a population-level approach.

Under the unsmoothed NO_MA condition, heart rate was by far the most important predictor, with a mean *t*-value of 25.04 (90% CI: 11.21–38.88), indicating a very strong and stable positive association. All other predictors, including skin temperature, hydration metrics, and movement variables, showed much weaker associations, with confidence intervals crossing zero, indicating they did not make a significant independent contribution.

This hierarchy of predictor importance remained consistent across the 15-, 30-, and 60-min smoothing windows. In all three conditions, heart rate remained the dominant predictor with mean *t*-values stable between 11.12 and 11.31 (CIs approx. 7.7–14.8). Meanwhile, skin temperature, sodium loss rate, sweat rate, and movement variables consistently displayed weaker associations with confidence intervals spanning zero, confirming their lack of statistical reliability in these models. A notable shift occurred only at the 90-min window (MA90). While heart rate remained the strongest predictor (mean *t*-value 8.69), enhanced speed emerged as a statistically significant but modest predictor with a mean *t*-value of 2.08 (CI: 0.31–3.85). Additionally, skin temperature became significantly negative at MA90 (mean *t*-value −4.29, CI −7.41 to −1.17), suggesting that, with very long smoothing windows, higher skin temperature was associated with lower mPSI in the linear models.

For both models with and without the hydration predictors from the NIX sensors, their mean (90% CI) RMSEs, as shown in Figure 7 reached the lowest in the holdout testing set when the

60-min moving average predictors were used. On the other hand, the median model R^2's reached the highest when 60-min moving average predictors were used. A predictive model can combine different predictors and their moving averages in its predictions. However, due to collinearity among the predictors and the risk of over-parameterization in linear regression (as we had only 1 day of data per participant), we defer consideration of mixing the predictors and their moving averages for prediction to the next section, using machine learning methods.

3.6.3 | Machine Learning

The performance of all RF and GBR configurations using R^2 (bars) and RMSE (line) is shown in Figure 9. Across all RF models, R^2 values ranged from approximately 0.84–0.96, with the highest R^2 observed for the 60-min moving average model (0.962), which performed slightly better than the model combining all predictors (0.960). The RF models using individual moving-average windows showed strong performance, although the unsmoothed (NO_MA) model produced the lowest R^2 among the RF configurations. RMSE values for the RF models ranged from roughly 0.25 to 0.51, with the lowest RMSE observed for the 60-min moving average and combined models.

GBR produced consistently high R^2 values as well, ranging from approximately 0.79 to 0.93. The GBR models that incorporated 60-min, 90-min, and all moving-average predictors achieved the highest R^2 values within this group. RMSE values for GBR models ranged from roughly 0.35 to 0.59, with the lowest RMSE occurring in the configurations that included longer moving-average windows or all predictors combined. Overall, both model families demonstrated strong predictive performance, with the best configurations for each model type achieving R^2 values above 0.93 and RMSE values below 0.35.

Feature importance rankings for both RF and GBR models are shown in Figure 10. Across all models and all moving-average configurations, heart rate consistently emerged as the strongest predictor of mPSI. In the RF models, heart rate accounted for the largest proportion of feature importance in every configuration, with relative contributions ranging from approximately 0.35 to 0.49. GBR demonstrated the same pattern, with heart rate dominating feature importance across all window lengths. Following heart rate, distance traveled, and skin temperature were generally the next most influential predictors. In unsmoothed models, distance contributed between approximately 0.10 and 0.14, but its importance dropped significantly in smoothed models (generally below 0.10). Skin temperature contributed approximately 0.10 to 0.14 consistently across most windows.

Environmental exposure measured by WBGT showed moderate importance in both model types, with values typically between 0.03 and 0.10. Hydration-related variables, including sweat rate, total fluid loss, and sodium loss rate, contributed to a smaller but non-negligible importance (generally below 0.05). Movement-related variables such as enhanced speed and enhanced altitude ranked lowest in both RF and GBR models. The inclusion of moving-average predictors changed the magnitude but not the hierarchy of importance. Heart rate moving-average features consistently retained the highest importance among all moving-average features, followed by skin temperature. Overall, the feature-importance results demonstrate clear and consistent patterns across all models, with cardiovascular indicators showing the strongest relationship with mPSI.

The feature importance rankings for the RF model that included all predictors and all moving-average windows (RF_All) and the GBR model with the same configuration (GBR_All) are shown in Figure 11. In both models, heart rate remained the dominant predictor of mPSI. The 60-min heart rate moving average (heart_rate_BPM_ma60) exhibited the highest importance in both RF_All and GBR_All, followed by the 30-min moving average and the unsmoothed (current) heart rate. These three heart-rate features accounted for the largest proportion of predictive contribution among all variables.

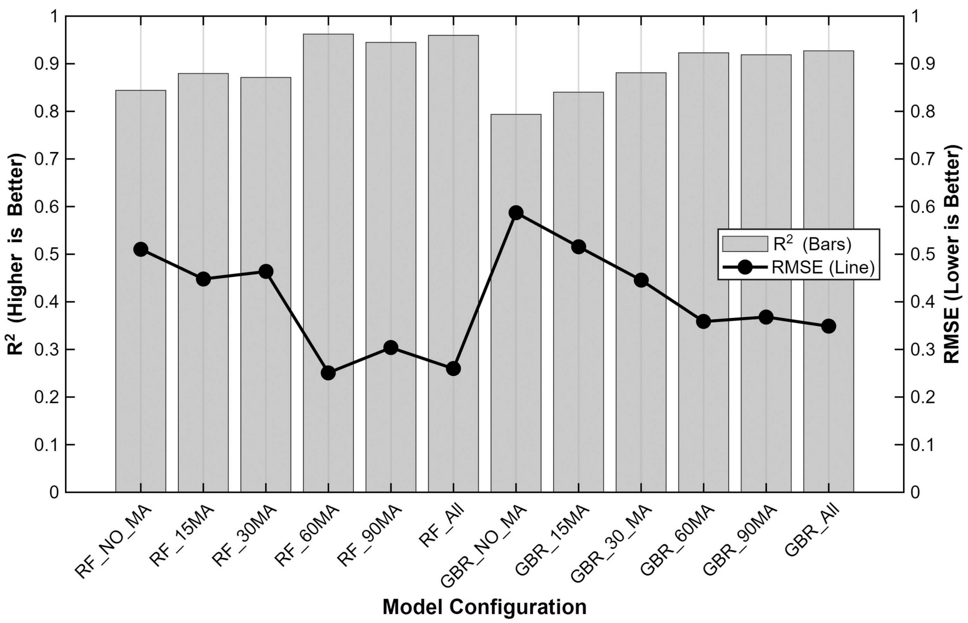

FIGURE 9 | The R^2 (y1 axis) and RMSE (y2 axis) values for all machine learning models.

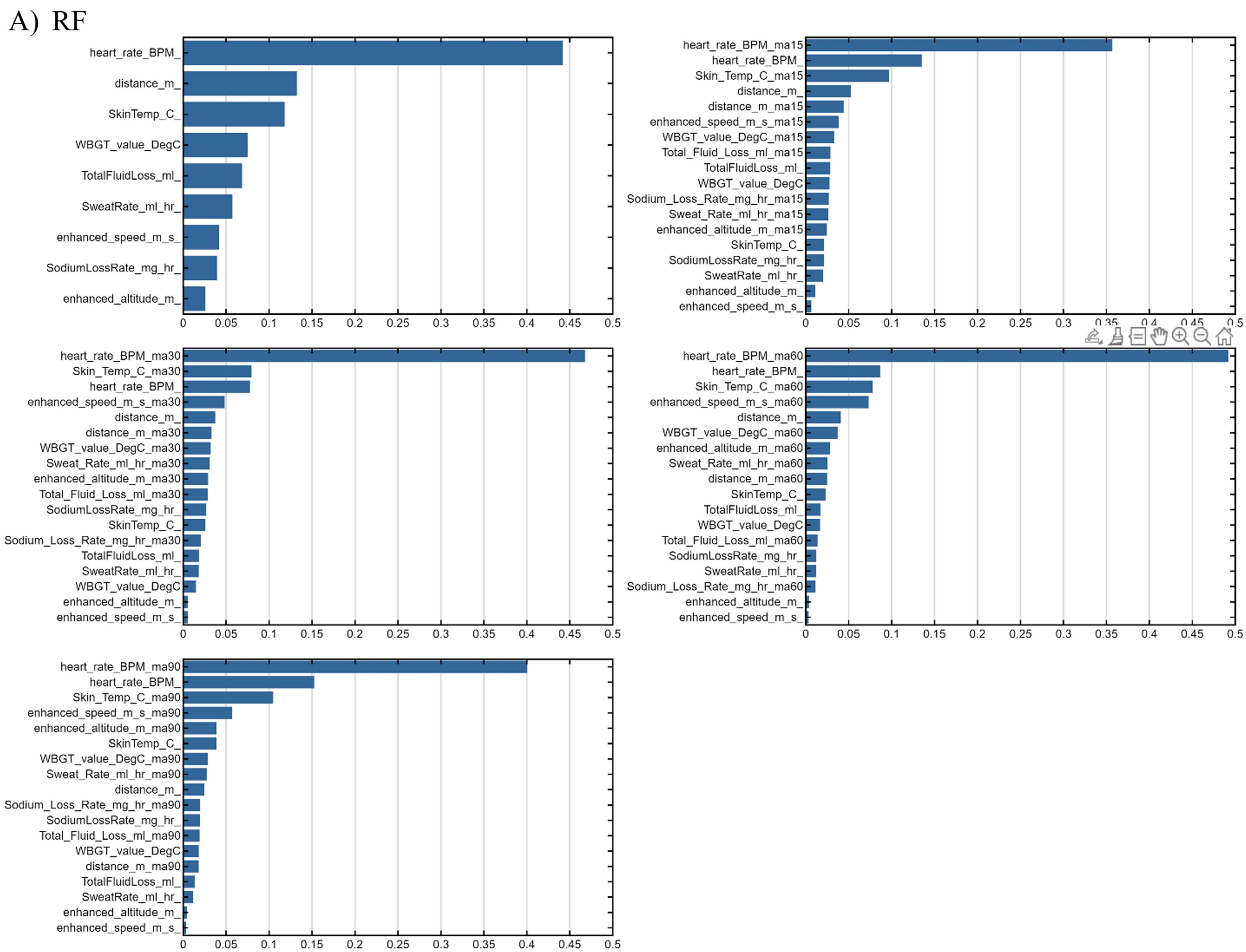

FIGURE 10 | Feature importance rankings for (A) RF and (B) GBR models for different moving average times. The higher value is better from top to bottom.

Skin temperature moving averages were the next most influential group of predictors. The 90-min and 60-min moving averages of skin temperature consistently ranked among the top features in both models. Distance moved and enhanced speed moving averages showed moderate importance, although they ranked below the core physiological variables. Environmental exposure (WBGT) and hydration-related predictors (sweat rate, total fluid loss, sodium loss rate) had relatively low importance compared to heart rate and skin temperature.

The feature-importance patterns were generally consistent between RF and GBR. Both models emphasized the role of longer moving-average windows, particularly the 60-min window, which consistently appeared among the top predictors across physiological, environmental, and movement variables.

3.6.4 | Overall Regression Comparison

The predictive performance of simple linear regression (SLR), multi-linear regression (MLR), and machine learning (ML) models was compared using R^2 and RMSE. The SLR analyses demonstrated modest predictive capability when predictors were evaluated individually. Across participants, median R^2 values ranged from 0.03 for WBGT to 0.21 for skin temperature, with a maximum of 0.60 for heart rate. No SLR model exceeded approximately 0.65, establishing a baseline for evaluating improvements from more advanced modeling approaches.

The MLR models produced substantially stronger performance. The best configuration occurred at the 60-min smoothing window, where the model with hydration variables achieved an R^2 of 0.82 and an RMSE of 0.30. In contrast, the model without hydration variables produced an R^2 of 0.75 and an RMSE of 0.39. Relative to the strongest SLR predictor ($R^2 = 0.60$), the best MLR model increased R^2 by approximately 37 percent. RMSE similarly improved, with reductions of roughly 35–40 percent compared with error levels expected from single-predictor models.

The machine learning models demonstrated the highest predictive performance. The Random Forest model at the 60-min smoothing window (RF_60MA) achieved an R^2 of 0.96 and an RMSE of 0.25. Compared with the best SLR performance ($R^2 = 0.60$), this represents a 60 percent improvement in explained variance. Relative to the best MLR model ($R^2 = 0.82$), the Random Forest model increased R^2 by approximately 17 percent (or 28 percent relative to the MLR model without hydration). The

B) GBR

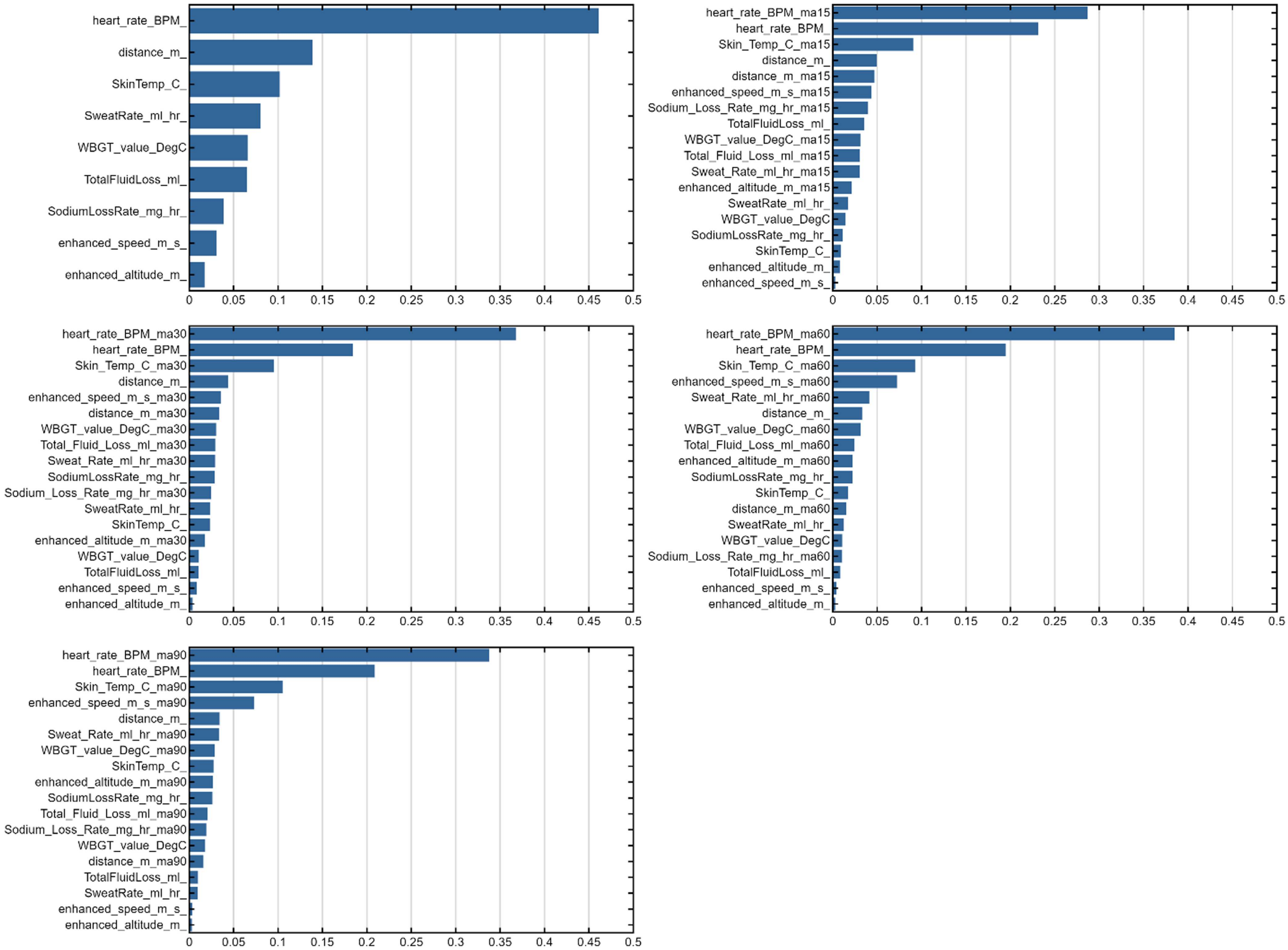

FIGURE 10 | (Continued)

Gradient Boosted Regressor at the same window also performed strongly, with an R^2 of 0.92 and an RMSE of 0.36, although both metrics were inferior to the Random Forest model.

4 | Discussion

This study evaluated the feasibility of physiological monitoring among 30 outdoor agricultural workers in eastern NC using three wearable devices. To our knowledge, this is the first study to deploy the CALERA Research sensor in the United States. Furthermore, this is the first study globally to evaluate the Nix hydration biosensor. Finally, this work is the first to combine all three devices to monitor agricultural workers. By combining all three devices in a field-based feasibility assessment, this study addresses a critical gap in validating consumer-grade wearables for occupational safety, utilizing advanced, multi-parameter technology to protect HRI-vulnerable farmworkers.

4.1 | Data Retrieval Performance

The monitoring devices revealed both strengths and challenges relevant to field-based physiological monitoring. The CALERA monitors achieved 100% data retrieval, confirming their reliability for continuous outdoor measurements. In contrast, the Nix hydration biosensors had intermittent synchronization issues with Android devices, resulting in an 80% retrieval rate. The Nix platform also offers a desktop software option that provides additional hydration metrics and allows users to download raw data. However, this feature requires a paid subscription, which may limit its broader research use. Garmin watches exhibited the second-highest reliability, with one participant inadvertently turning off data collection, thereby reducing the overall retrieval rate to 97%. Additionally, four others disabled the GPS function, which affected variables such as distance and speed. These issues underscore the need to implement device-locking or restricted-user modes in future studies to prevent participants from unintentionally interrupting measurements and to ensure the collection of complete, high-quality datasets.

4.2 | Demographics

The demographic profile of the participants partially reflects a typical agricultural workforce in the region, consisting primarily of male, married adults with varying levels of formal education and English proficiency. The cohort was younger,

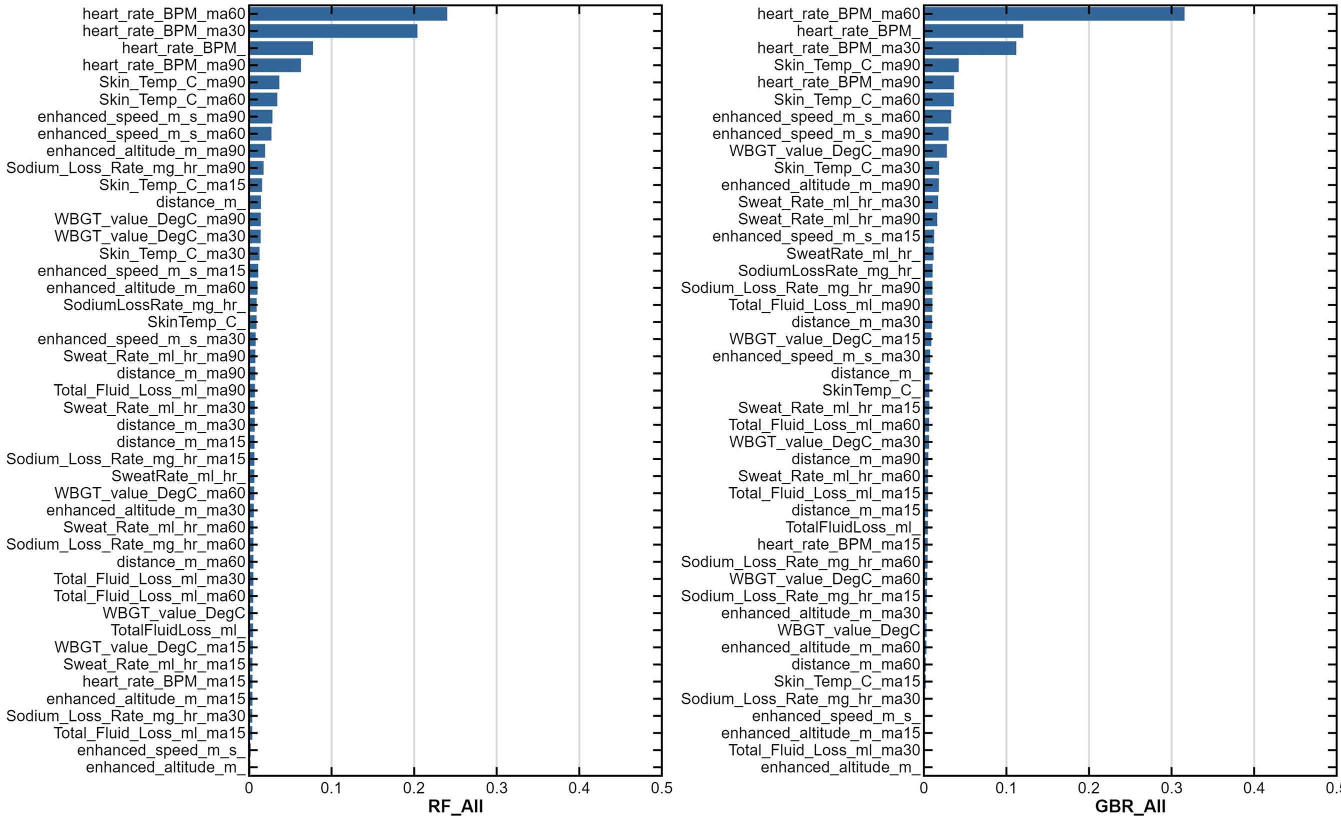

FIGURE 11 | Feature importance rankings for RF (left) and GBR (right) models for all moving averages. The higher value is better from top to bottom.

32.5 years, than a nationally representative sample of 41 years [52]. The findings also reveal generally favorable bathroom access, with most workers reporting a walking distance of 3 min or less, an important factor supporting regular hydration and minimizing intentional fluid restriction. The baseline heart rhythm and age distribution indicate a relatively healthy, working-age population, though variability in physical conditioning and experience may influence individual heat response.

4.3 | Environmental and Participants' Physiological Changes

This field-based evaluation of outdoor agricultural workers during July demonstrates that participants were consistently exposed to high environmental heat, resulting in measurable physiological responses indicative of mild to moderate heat strain. WBGT levels remained elevated throughout the monitoring period, with daily medians near 28°C and upper quartiles approaching 30°C, aligning with the ACGIH threshold limit values (TLVs) for moderate workloads. The higher minimum WBGT on July 17 suggests that heat was sustained later into the day, reducing opportunities for thermal recovery between work intervals.

The range of mPSI values observed in this study shows that participants experienced a broad spectrum of physiological strain, from minimal to high risk. The participant PSI average of 2.8 was slightly higher than that of 221 Florida farmworkers ($m = 2.5$) [39], despite our sample being majority males and younger, both protective factors for HRI. While the groups had similar job tasks, perhaps work intensity could explain the difference, as the Florida group's moderate intensity activity spanned about a third of the work shift, while the NC farmworkers were assumed to be closer to 75%. Chavez Santos et al. [53] reported mean maximum physiological strain index (PSI) values of 4.3 and 4.6 for intervention and comparison groups of agricultural workers in Washington state, respectively, noting that lower effort tasks generally corresponded to PSI values between 3 and 4. In contrast, our cohort exhibited notably higher physiological strain, characterized by a mean maximum mPSI of 5.28 and individual peaks reaching 7.55. This elevation suggests our participants experienced greater combined physical and environmental stress; however, data collection occurred monthly over 3 months, and core body temperature was estimated from one tympanic temperature reading combined with continuous heart rate monitoring, not from a continuous temperature monitor like in our study.

While approximately 8 percent of all values exceeded a commonly referenced threshold for elevated strain, core body temperature exceeded the NIOSH-recommended limit of 38 degrees Celsius in more than 12 percent of all recorded measurements. The relatively narrow range of core temperature variability among many participants suggests the workers were well adapted to HRI-high-risk work. Among Washington State pear and apple harvesters observed CBT_{gi} exceeding 38.5°C among 17% of participants [22] which is consistent with reports among agricultural workers in other US States. For example, among 221 nursery, fernery, and crop workers over three summer workdays in Florida, 16% of participants exceeded a gastrointestinal temperature of 38.5°C on at least 1 day [39].

Parsons et al. [29] found that agricultural workers spent 7.1% of their total working minutes with a core body temperature at or above 38°C. On days with severe heat (WBGT > 31.1°C), the time agricultural workers spent above this safety threshold increased drastically to 21.5% of their working minutes.

Sweat rate values varied greatly among participants, and a small subset exceeded the NIOSH guideline of 1.2 kilograms per hour. Differences in sweat rate likely reflect variation in work intensity, acclimatization, and individual sweat gland activity. These findings support the need for individualized hydration plans, particularly for heavy sweaters. Although we did not collect fluid intake quantity during the data collection, previous research suggests farmworkers drink 2–6 liters of fluid per shift [54–56]. This suggests that most workers should have been hydrated, given that total fluid loss averaged less than 1.5 liters. Post-shift dehydration is common in the studied farmworkers [54, 57, 58], and therefore, future research combining Nix and hydration status measures (i.e., urine specific gravity) is necessary.

The only published study on sweat analysis of farmworkers was with older adult Japanese farmworkers who lost 668 mL/h, determined through body weight change for sweat loss and local sweat collection with absorbent tape [59]. Sweat rate among electrical line workers in Australia was 0.44 L/440 mL/h, as estimated by pre- and post-shift weight change.

Sodium loss rate displayed substantial variation across participants. Higher losses indicate a combination of high sweat volume and higher sweat sodium concentration. Individuals with greater sodium losses may face an increased risk of muscle cramping and reduced thermoregulatory efficiency. The range of sodium losses observed in this study underscores the importance of incorporating electrolyte replacement into heat-stress management, particularly for participants with the highest sweat volumes. Hydration-related findings suggest opportunities for improvement. Only 30% of participants reported completing mandatory hydration education, and fewer than half reported adhering to recommended hydration frequency. These patterns may contribute to greater heat strain susceptibility, particularly among workers with lower English proficiency who may face barriers to understanding training materials or accessing training.

Heart rate remained below the ACGIH recommended limit of 180 beats per minute for all participants. Although the recommended limit was not exceeded, the elevated average heart rate observed throughout the work periods indicates significant cardiovascular engagement. Heart rate is a sensitive indicator of metabolic workload and thermoregulatory effort, and elevated heart rates reflect the combined strain of physical work and environmental heat exposure. Distance traveled varied widely among participants, indicating different work patterns and levels of physical activity. Enhanced speed values remained relatively low. This suggests that most participants worked at moderate movement intensity. Short periods of rapid movement can elevate metabolic heat production, but the overall movement patterns observed indicate that most tasks involve steady, moderate speed. While this level of activity still contributes to heat load, it likely did not drive extreme physiological strain on its own. Changes in altitude during work were small and unlikely to meaningfully affect metabolic workload or thermoregulatory stress. The minimal elevation changes observed are consistent with work carried out on mostly level terrain. As a result, altitude-related workload was not a major contributor to physiological strain in this setting.

4.4 | AIHA App Risk Analysis Comparison With mPSI

Findings showed that mPSI-derived heat strain risks among the agricultural workers were consistently lower than WBGT-derived heat stress risks. Two factors may explain this difference: worker acclimatization and sensor accuracy. First, the participants were likely well acclimatized to the hot environment. Approximately 90% had worked in agriculture for at least 1 year (up to more than 5 years), and all participants had worked at least ten weeks during the agricultural season when the study took place. Repeated exposure to hot conditions promotes physiological adaptations that improve the body's efficiency in coping with heat stress [60]. Acclimatized workers typically exhibit lower CBT and heart rate, and a higher sweat rate, than non-acclimatized workers [61–63]. These adaptations may have resulted in lower calculated mPSI values and, therefore, lower heat strain risk estimates when compared with WBGT-derived risk.

WBGT is considered the gold standard for assessing workplace environmental heat exposure because it incorporates environmental parameters such as ambient temperature, humidity, wind speed, and radiant heat [64]. However, WBGT does not account for individual physiological responses, which can lead to overestimation of risk, particularly when workers are acclimatized and more heat tolerant. WBGT-derived risks, which are intentionally conservative, may better represent workers who are unacclimatized, have pre-existing health conditions, take medications, or have other factors that reduce heat tolerance [60].

Second, the lower mPSI-derived risk may also be related to potential underestimation of CBT and heart rate by the wearable sensors, as reported in previous studies [19, 35].

4.5 | Geospatial Analysis

Results indicated that minutely mPSI data can vary depending on the individual and location. The elevated mPSI values among participants while working in a field were expected [39]; however, the elevated heat strain/stress risk in participants while traveling along a road was unexpected. This likely occurred due to a lag response as workers transitioned from working in a field to a vehicle to travel to the next work site. It should be noted that the workers in this study have air-conditioned work vehicles and were assumed to be functioning on the data collection day.

4.6 | Correlation and Prediction Models

4.6.1 | Correlation Between Study Variables

The analysis showed that the linear relationship between individual predictors and mPSI varies substantially by variable and by the degree of temporal smoothing applied. Under unsmoothed conditions, heart rate was the strongest predictor, with a median R^2 of 0.60, while skin temperature and other physiological measures exhibited more modest associations. These baseline differences reflect the dominant contribution of cardiovascular strain to the mPSI metric smoothing that had mixed effects across variables. Skin temperature demonstrated a modest but consistent improvement (+ 13.8%), suggesting that reducing short-term noise enhances its alignment with mPSI. Larger percentage increases were seen for sweat rate, distance, and enhanced altitude, though these increases often resulted from low baseline R^2 values and did not translate into strong absolute predictive performance. Enhanced speed showed the largest relative increase, but its overall predictive value remained low.

In contrast, smoothing reduced the predictive strength of heart rate (− 24.2%), indicating that high-frequency variability in heart rate is relevant for capturing rapid physiological strain responses. The sodium loss rate also declined slightly with smoothing, further indicating that not all predictors benefit from temporal averaging.

Overall, these findings highlight that smoothing does not uniformly improve predictor mPSI relationships. Variables differ in whether their informative content lies in short-term fluctuations or longer-term trends. Effective preprocessing for heat-strain modeling should therefore be variable-specific rather than uniformly applied, ensuring that smoothing enhances the predictive signal rather than removing physiologically meaningful information.

4.6.2 | Multi-Linear Regression

The multi-linear regression models demonstrated substantial improvements over simple linear regression by incorporating multiple physiological predictors simultaneously. While the strongest single predictor in SLR, heart rate, achieved an R^2 of approximately 0.60, the best-performing MLR model increased this value to 0.82 at the 60-min smoothing window. This represents a 37 percent improvement in explained variance and a meaningful reduction in RMSE. These results confirm that combining physiological variables provides a more accurate estimation of mPSI than evaluating predictors individually.

The t-statistics provided additional insight into the relative importance of the predictors and the internal structure of the MLR models. Across all smoothing windows, heart rate was consistently the dominant predictor. In the unsmoothed model, heart rate produced an exceptionally high mean *t*-value of 25.04, with a 95 percent confidence interval from 11.21 to 38.88. This large, stable, and unambiguously positive effect indicates that heart rate contributes the majority of the linear predictive power in the MLR framework. Even after smoothing, the effect remained strong. At MA15, MA30, and MA60, heart rate *t*-values remained between 11.12 and 11.31, with confidence intervals that did not overlap zero for any window. These findings demonstrate that heart rate is the primary physiological indicator of mPSI across all versions of the linear model.

Other predictors contributed far less. Skin temperature showed moderate *t*-values, but its confidence intervals crossed zero for most smoothing windows, indicating inconsistent or participant-dependent effects. Hydration-related variables, including sodium loss rate, sweat rate, and total fluid loss, showed weak and unstable contributions across all smoothing levels, with wide intervals overlapping zero. Movement variables, including distance, enhanced altitude, and enhanced speed, demonstrated minimal independent predictive contribution, with t-values near zero and confidence intervals that crossed zero for nearly all windows. The only exception occurred at the 90-min smoothing window, where enhanced speed showed a small but statistically significant contribution (*t*-value 2.08, CI 0.31–3.85). Even in this case, the effect was modest relative to the effect on heart rate.

Overall, these results show that the performance gains observed in the MLR models stem primarily from the inclusion of heart rate, which carries the strongest and most stable linear relationship with mPSI. Other predictors provide limited additional explanatory power in the linear model structure, and their contributions are often statistically insignificant. The improvement of MLR over SLR, therefore, reflects the inclusion of multiple correlated predictors, but the model's predictive strength is dominated by a single variable. These findings also clarify why more flexible machine learning models outperform MLR. When predictors beyond heart rate have weak or inconsistent linear associations, linear models have limited ability to extract additional signals. Ensemble tree models, which capture nonlinear interactions and complex dependencies, can leverage this structure more effectively, leading to substantial accuracy gains observed in the ML analyses.

4.6.3 | Machine Learning

Machine learning results from RF and GBR indicate that RF consistently outperformed later, both with and without the moving-average features. In general, RF offers advantages such as reduced risk of overfitting, greater robustness, and fewer hyperparameters requiring tuning. In contrast, GBR is more susceptible to overfitting because boosting models iteratively fit residuals and can more easily memorize fine-grained patterns in the training data. This study employed a holdout validation method in which training and testing samples were randomly selected. Therefore, the temporal order of the data was not preserved. For example, the training set may include records from 9:31, 9:32, and 9:50, while the testing set contains records from 9:33 and 9:34. This sampling strategy introduces temporal leakage, where short-range time-dependent patterns present in the training data also appear in the testing data. Such leakage can artificially inflate performance estimates and can affect models differently.

Because RF is an ensemble of decorrelated trees, it is generally more robust to subtle temporal leakage and less prone to memorizing sequential dependencies. GBR, on the other hand, is more sensitive to this issue because each tree attempts to correct the errors of the previous one, making the model more likely to fit noise associated with short-term temporal structure. We believe this difference in model behavior explains why RF demonstrated superior performance in our experiments.

4.6.4 | *Overall Regression Comparison*

This study compared the predictive performance of simple linear regression, multi-linear regression, and machine learning models for estimating mPSI across multiple smoothing windows and input configurations. The results demonstrated a clear and consistent progression in model performance, with substantial quantitative improvements from SLR to MLR and from MLR to ML. The simple linear regression models provided limited predictive strength when each variable was assessed individually. Median R^2 values ranged from 0.03 to 0.21 for most predictors and reached a maximum of 0.60 for heart rate, with no SLR model exceeding approximately 0.65. These results indicate that single predictors alone do not capture enough variability in the physiological strain index to support accurate prediction. This limitation of SLR served as a baseline against which improvements from MLR and ML could be measured.

The multi-linear regression models significantly improved predictive performance by incorporating multiple physiological variables simultaneously. The best MLR model, which used a 60-min smoothing window and included hydration variables, achieved an R^2 of 0.82 and an RMSE of 0.30. This represented approximately a 37 percent increase in R^2 compared with the strongest SLR model, where R^2 improved from 0.60 to 0.82. RMSE also improved meaningfully when transitioning from SLR to MLR, reflecting enhanced prediction accuracy. These results show that integrating multiple predictors better captures the multidimensional physiological processes underlying mPSI, thereby improving explanatory power and accuracy beyond what SLR can achieve. Machine learning models provided the largest performance gains, particularly the Random Forest model at the 60-min smoothing window. This model achieved an R^2 of 0.96 and an RMSE of 0.25, which represented the strongest performance across all modeling approaches. Compared with the best SLR model, this reflects a 60 percent increase in explained variance, rising from 0.60 to 0.96. Relative to the best MLR model ($R^2 = 0.82$), the Random Forest increased R^2 by an additional 17 percent. These improvements demonstrate the ability of machine learning models to capture nonlinear relationships and complex interactions among variables that linear models cannot represent. Although the Gradient Boosted Regressor also performed well, with an R^2 of 0.92 at the 60-min window, the Random Forest consistently outperformed GBR across all smoothing windows, highlighting its stability and suitability for this dataset.

Together, these findings show a clear hierarchy of predictive capability. SLR provides limited insight into mPSI variability when predictors are evaluated individually. MLR improves performance substantially by modeling multiple predictors simultaneously. Machine learning models provide the best performance, achieving the highest R^2 values and lowest RMSE values, with the Random Forest model at the 60-min window emerging as the optimal configuration. These results underscore the value of ensemble learning methods for capturing the complex physiological and behavioral relationships inherent in mPSI and highlight the benefit of moderate smoothing for stabilizing the input signals and improving predictive accuracy.

4.7 | Study Limitations and Future Work

The study has several limitations that should be addressed in future work. Due to limited funding, the devices were not evaluated against reference instruments or standard methods, which restricts the ability to quantify measurement accuracy. Future work will include formal evaluations and the development of device-specific regression models to improve data accuracy. In addition, the findings may not fully represent the heat strain risk experienced by unacclimatized outdoor workers in agriculture or similar outdoor environments. These workers are more likely to experience higher PSI values and, therefore, may face heat stress risks that align more closely with risks estimated from WBGT. The prediction models in this study were validated using a holdout method for both multilinear regression and machine learning approaches. Future work should explore multi-day validation by applying model parameters derived on the first day to subsequent days to better assess temporal stability. Finally, future work will focus on designing and implementing more advanced machine learning models, including transformer-based architectures and deep neural networks, to enhance predictive performance. These approaches require substantially larger datasets, and current funding limitations have prevented the collection of sufficient data to support their development. High-resolution WBGT data were obtained from a weather station ~32 km from the farm and may not reflect the actual WBGT on the study site due to potential differences in weather conditions. Although air temperature and relative humidity may not significantly vary between the weather station and the study site, WBGT also incorporates wind speed and solar irradiance, which may differ significantly by location. The use of on-site heat stress monitors to obtain more accurate WBGT data in future studies is recommended.

Acclimatization status was not formally assessed in this study. Although demographic data indicate that the majority of participants had substantial prior agricultural experience and had been working throughout the current season, which suggests a reasonable degree of heat acclimatization, individual acclimatization levels or their specific contribution to the observed mPSI values cannot be quantified. Acclimatization is known to lower resting and exercise heart rate, reduce core body temperature during exertion, and increase sweat rate, all of which directly influence mPSI calculations. Future studies should incorporate validated acclimatization assessments (e.g., heat tolerance tests or serial physiological monitoring across the season) and, where feasible, stratify analyses by acclimatization status or time in the season to better understand its moderating role.

5 | Conclusion

This pilot study demonstrated that commercially available wearable devices can successfully capture continuous physiological and environmental data among agricultural workers performing real outdoor tasks during periods of substantial heat exposure. The CALERA core temperature sensor, the Nix Hydration Biosensor, and the Garmin smartwatch generated meaningful measurements with high retrieval rates and provided information that reflected both individual strain and environmental conditions throughout the workday. The observed physiological responses, including elevated core temperature, increased heart rate, and variable sweat and sodium loss, showed that workers experience measurable heat-related strain even when risk levels derived from WBGT may suggest more severe conditions. Machine learning models, particularly the Random Forest model using 60-min moving averages, demonstrated strong predictive capacity for estimating mPSI, which supports the use of integrated modeling approaches for real-time heat risk assessment.

These findings indicate that wearable sensor-based systems can complement traditional heat stress tools by providing personalized, continuously updated physiological information. This approach has the potential to strengthen workplace heat safety programs, guide earlier interventions, and reduce the likelihood of heat-related illness among workers in physically demanding outdoor industries. From an occupational health perspective, employers could supply these wearable sensors as a proactive safety measure. However, to prevent discriminatory employment practices, the raw physiological data should remain exclusively accessible to individual workers to privately manage their hydration and rest needs. Further research with larger and more diverse populations, multiple work tasks, and longer monitoring periods will help refine these systems and advance the development of practical, data-driven decision tools for occupational heat protection.

Author Contributions

Conceptualization: Sinan Sousan, Elizabeth Mizelle, Jo Anne Balanay. Methodology and formal analysis: Sinan Sousan, Elizabeth Mizelle, Qiang Wu, Rui Wu, Guy Iverson, Jo Anne Balanay. Funding acquisition: Sinan Sousan. Project administration and supervision: Sinan Sousan, Elizabeth Mizelle, Jo Anne Balanay. Data curation: Sinan Sousan, Elizabeth Mizelle, Dylan Hemedinger, Eliezer Loyola, Takari Hood, Monica Ramirez, Jo Anne Balanay. Data visualization: Sinan Sousan, Qiang Wu, Guy Iverson, Jo Anne Balanay. Writing – original draft: Sinan Sousan, Qiang Wu, Rui Wu, Guy Iverson, Jo Anne Balanay. Writing – review and editing: Sinan Sousan, Elizabeth Mizelle, Qiang Wu, Rui Wu, Guy Iverson, Ciprian Popoviciu, Dylan Hemedinger, Eliezer Loyola, Takari Hood, Monica Ramirez, Jo Anne Balanay.

Acknowledgments

The authors want to thank the study participants and farm personnel. Roanoke Chowan Community Health Center, Association of Mexicans in North Carolina, Inc., the North Carolina Agromedicine Institute, NC State Extension agents, NC Hispanic Federation, and Access East assisted with farm and farmworker recruitment. Thank you to the North Carolina State Climate Office for the WBGT data. Thank you to Dr. Juan Daneri for recruiting students for our research team. The project was funded by several sources, which include (1) the National Institute of Environmental Health Sciences of the National Institutes of Health under Award Number P30ES025128; (2) the NC GlaxoSmithKline Foundation grant; and (3) the North Carolina Agromedicine Institute. In addition, Part of S. Sousan's efforts were supported by the National Institute of Environmental Health Sciences of the National Institutes of Health, award number P30ES025128. Part of E. Mizelle's efforts were supported by the National Institute for Occupational Safety and Health (NIOSH) under award number K01OH012697. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the Centers for Disease Control and Prevention (CDC).

Ethics Statement

This study was approved by the University and Medical Center Institutional Review Board (UMCIRB 24-001488) at East Carolina University.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

De-identified CBT and Nix data are available upon request. Data that cannot be de-identified (e.g., GPS data from the Garmin watch) will not be available to maintain participants' anonymity.

References

1. EPA. *Environmental Protection Agency: Excessive Heat Events Guidebook. EPA 430-B-06-005* (US EPA, Office of Atmospheric Programs, 2006).

2. WHO. *Climate Change and Workplace Heat Stress: Technical Report and Guidance*. (World Health Organization, 2025).

3. "Heat Stress and Workers," National Institute for Occupational Safety and Health, Centers for Disease Control and Prevention, accessed October 2025, https://www.cdc.gov/niosh/heat-stress/about/index.html.

4. "Clinical Overview of Heat and Cardiovascular Disease," Heat Health, Centers for Disease Control and Prevention, 2024, https://www.cdc.gov/heat-health/hcp/clinical-overview/heat-and-people-with-cardiovascular-disease.html.

5. "Heat Stress: What Is It and How Is It Measured?," Copernicus Climate Change Service, 2024, https://climate.copernicus.eu/heat-stress-what-it-and-how-it-measured.

6. A. P. C. Chan and W. Yi, "Heat Stress and Its Impacts on Occupational Health and Performance," *Indoor and Built Environment* 25, no. 1 (2016): 3–5.

7. "Heat-related Illnesses," National Institute for Occupational Safety and Health, Centers for Disease Control and Prevention, accessed October 2025, https://www.cdc.gov/niosh/heat-stress/about/illnesses.html.

8. K. L. Ebi, A. Capon, P. Berry, et al., "Hot Weather and Heat Extremes: Health Risks," *Lancet* 398, no. 10301 (2021): 698–708.

9. J. M. Mix, L. Elon, V. V. Thein Mac, et al., "Physical Activity and Work Activities in Florida Agricultural Workers," *American Journal of Industrial Medicine* 62, no. 12 (2019): 1058–1067.

10. L. Benos, D. Tsaopoulos, and D. Bochtis, "A Review on Ergonomics in Agriculture. Part I: Manual Operations," *Applied Sciences* 10, no. 6 (2020): 1905.

11. "Using the Heat Index: A Guide for Employers," Occupational Safety and Health Administration, accessed on October 2025, https://www.osha.gov/heat-exposure/hazards. 2017.

12. S. K. Sankar, V. P, K. S, R. Shanmugam, L. P. Kamalakkanan, and V. Venugopal, "Effectiveness of Heat Stress Interventions Among Outdoor Workers: A Protocol Paper," *Frontiers in Public Health* 12 (2024): 1477186.

13. G. P. Bates and J. Schneider, "Hydration Status and Physiological Workload of UAE Construction Workers: A Prospective Longitudinal

Observational Study," *Journal of Occupational Medicine and Toxicology* 3, no. 1 (2008): 21.

14. J. F. Piil, J. Lundbye-Jensen, L. Christiansen, et al., "High Prevalence of Hypohydration in Occupations With Heat Stress—Perspectives for Performance in Combined Cognitive and Motor Tasks," *PLoS One* 13, no. 10 (2018): e0205321.

15. R. S. Wagoner, N. I. López-Gálvez, J. G. de Zapien, S. C. Griffin, R. A. Canales, and P. I. Beamer, "An Occupational Heat Stress and Hydration Assessment of Agricultural Workers in North Mexico," *International Journal of Environmental Research and Public Health* 17, no. 6 (2020): 2102.

16. D. S. Moran, A. Shitzer, and K. B. Pandolf, "A Physiological Strain Index to Evaluate Heat Stress," *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* 275, no. 1 (1998): R129–R134.

17. A. Eisenkraft, N. Goldstein, M. Fons, et al., "Comparing Body Temperature Measurements Using the Double Sensor Method Within a Wearable Device With Oral and Core Body Temperature Measurements Using Medical Grade Thermometers—A Short Report," *Frontiers in Physiology* 14 (2023): 2023.

18. Y. Seo, T. DiLeo, J. B. Powell, J.-H. Kim, R. J. Roberge, and A. Coca, "Comparison of Estimated Core Body Temperature Measured With the BioHarness and Rectal Temperature under Several Heat Stress Conditions," *Journal of Occupational and Environmental Hygiene* 13, no. 8 (2016): 612–620.

19. P. Helmer, S. Hottenrott, P. Rodemers, et al., "Accuracy and Systematic Biases of Heart Rate Measurements by Consumer-Grade Fitness Trackers in Postoperative Patients: Prospective Clinical Trial," *Journal of Medical Internet Research* 24, no. 12 (2022): e42359.

20. S. Gillinov, M. Etiwy, R. Wang, et al., "Variable Accuracy of Wearable Heart Rate Monitors During Aerobic Exercise," *Medicine and Science in Sports and Exercise* 49, no. 8 (2017): 1697–1703.

21. D. S. Moran, S. J. Montain, and K. B. Pandolf, "Evaluation of Different Levels of Hydration Using a New Physiological Strain Index," *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* 275, no. 3 (1998): R854–R860.

22. J. Egbert, J. Krenz, P. D. Sampson, et al., "Accuracy of an Estimated Core Temperature Algorithm for Agricultural Workers," *Archives of Environmental & Occupational Health* 77, no. 10 (2022): 809–818.

23. M. Dally, A. Dye-Robinson, Y. Li, et al., "Performance of Summary Measures of Core Body Temperature in Heat-Related Health Outcome Regression Models," *International Journal of Biometeorology* 69, no. 9 (2025): 2405–2409.

24. W. M. Januário, N. F. Lessa, A. J. O. Schittine, et al., "Validity and Reproducibility of the CALERA Research Sensor to Estimate Core Temperature at Different Intensities of a Cycling Exercise in the Heat," *Journal of Thermal Biology* 123 (2024): 103907.

25. N. Kubota, K. Okada, and Y. Yamanaka, "Circadian Phase Assessment of Core Body Temperature Using a Wearable Temperature Sensor Under the Real World," *Sleep Science* 18, no. 03 (2025): e246–e252.

26. R. Wibowo, V. Do, C. Quartucci, et al., "Effects of Heat and Personal Protective Equipment on Thermal Strain in Healthcare Workers: Part B—Application of Wearable Sensors to Observe Heat Strain Among Healthcare Workers under Controlled Conditions," *International Archives of Occupational and Environmental Health* 97, no. 1 (2024): 35–43.

27. J. D. Runkle, C. Cui, C. Fuhrmann, S. Stevens, J. Del Pinal, and M. M. Sugg, "Evaluation of Wearable Sensors for Physiologic Monitoring of Individually Experienced Temperatures in Outdoor Workers in Southeastern U.S," *Environment International* 129 (2019): 229–238.

28. "CALERA Core Body Temperature for Research," Greenteg, 2025, https://www.greenteg.com/en/research.

29. L. Parsons, P. Mishra, J. Cole, S. Lawreniuk, and L. V. Long, "Climate-Linked Heat Inequality in the Global Southern Workforce: Cambodian Workers' Economic and Health Vulnerability to High Core Temperatures in Five Occupational Sectors," *Climate and Development* 17, no. 10 (2025): 869–881.

30. K. Li, C. Cardoso, A. Moctezuma-Ramirez, A. Elgalad, and E. Perin, "Heart Rate Variability Measurement Through a Smart Wearable Device: Another Breakthrough for Personal Health Monitoring?," *International Journal of Environmental Research and Public Health* 20, no. 24 (2023): 7146.

31. N. Davis, J. Heikenfeld, C. Milla, and A. Javey, "The Challenges and Promise of Sweat Sensing," *Nature Biotechnology* 42, no. 6 (2024): 860–871.

32. S. M. Saia, S. P. Heuser, M. D. Neill, W. A. LaForce IV, J. A. McGuire, and K. D. Dello, "A Technical Overview of the North Carolina ECONet," *Journal of Atmospheric and Oceanic Technology* 40 (2023): 701–717.

33. J. Shen, K. Abdelkader, Z. Isaac, et al., "Data Missingness in Digital Phenotyping: Implications for Clinical Inference and Decision-Making," preprint, medRxiv, 2024, https://doi.org/10.1101/2024.10.03.24314808.

34. D. J. Catellier, P. J. Hannan, D. M. Murray, et al., "Imputation of Missing Data When Measuring Physical Activity by Accelerometry," *Medicine and Science in Sports and Exercise* 37, no. 11 (2005): 555–562.

35. X. Xu, G. Wu, Z. Lian, and H. Xu, "Feasibility Analysis of Applying Non-Invasive Core Body Temperature Measurement in Sleep Research," *Energy and Buildings* 303 (2024): 113827.

36. N. Verdel, T. Podlogar, U. Ciuha, H.-C. Holmberg, T. Debevec, and M. Supej, "Reliability and Validity of the CORE Sensor to Assess Core Body Temperature During Cycling Exercise," *Sensors* 21, no. 17 (2021): 5932.

37. A. Jolicoeur Desroches, C. Naulleau, T. A. Deshayes, T. Pancrate, and E. D. B. Goulet, "CORE™ Wearable Sensor: Comparison Against Gastrointestinal Temperature During Cold Water Ingestion and a 5 km Running Time-Trial," *Journal of Thermal Biology* 115 (2023): 103622.

38. M. Etiwy, Z. Akhrass, L. Gillinov, et al., "Accuracy of Wearable Heart Rate Monitors in Cardiac Rehabilitation," *Cardiovascular Diagnosis and Therapy* 9, no. 3 (2019): 262–271.

39. V. V. Mac, L. Elon, D. J. Smith, et al., "A Modified Physiological Strain Index for Workplace-Based Assessment of Heat Strain Experienced by Agricultural Workers," *American Journal of Industrial Medicine* 64, no. 4 (2021): 258–265.

40. National Weather Service. What is the Heat Index?, 2025, https://www.weather.gov/ama/heatindex.

41. B. Angol, S. Sousan, and J. A. G. Balanay, "Comparison Between WBGT App Prototype and WBGT Monitor to Assess Heat Stress Risk in an Eastern North Carolina Outdoor Setting," *Journal of Occupational and Environmental Hygiene* 22, no. 4 (2025): 274–287.

42. "Thermal (Heat) Stress - Heat Stress App." American Industrial Hygiene Association, accessed on October 2025, https://www.aiha.org/public-resources/healthierworkplaces/healthier-community-resources/thermal-heat-stress-resources-oehs-professionals.

43. D. Dillane and J. A. G. Balanay, "Comparison Between OSHA-NIOSH Heat Safety Tool App and WBGT Monitor to Assess Heat Stress Risk in Agriculture," *Journal of Occupational and Environmental Hygiene* 17, no. 4 (2020): 181–192.

44. J. Dieffenderfer, H. Goodell, S. Mills, et al., "Low-Power Wearable Systems for Continuous Monitoring of Environment and Health for Chronic Respiratory Disease," *IEEE Journal of Biomedical and Health Informatics* 20, no. 5 (2016): 1251–1264.

45. "IntelliSafe P31 Research Smartwatch," IntelliSafe, 2025, https://intellisafeanalytics.com/research/.

46. S. J. Rigatti, "Random Forest," *Journal of Insurance Medicine* 47, no. 1 (2017): 31–39.

47. J. Ye, J.-H. Chow, J. Chen, et al., "Stochastic Gradient Boosted Distributed Decision Trees," *Proceedings of the 18th ACM Conference on Information and Knowledge Management* (2009): 2061–2064.

48. M. Buller, D. Looney, A. Welles, B. Ely, W. Tharion, and R. Hoyt, "An Adaptive Physiological Strain Index That Accounts for Thermal-work Strain in High-performance Athletes and Encapsulated Workers." *Paper presented at: International Conference on the Physiology and Pharmacology of Temperature Regulation, Slovenia* (2016).

49. D. S. Moran, Y. Shapiro, A. Laor, S. Izraeli, and K. B. Pandolf, "Can Gender Differences During Exercise-Heat Stress Be Assessed by the Physiological Strain Index?," *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* 276, no. 6 (1999): R1798–R1804.

50. E. F. Coyle, "Fluid and Fuel Intake During Exercise," *Journal of Sports Sciences* 22, no. 1 (2004): 39–55.

51. S. M. Shirreffs and M. N. Sawka, "Fluid and Electrolyte Needs for Training, Competition, and Recovery," *Journal of Sports Sciences* 29, no. sup1 (2011): S39–S46.

52. W. Fung, K. Prado, A. Gold, A. Padovani, D. Carroll, and E. Finchum-Mason. 2023. "Findings From the National Agricultural Workers Survey (NAWS) 2021–2022: A Demographic and Employment Profile of United States Crop Workers (Research Report No. 17)." https://www.dol.gov/sites/dolgov/files/ETA/naws/pdfs/NAWS%20Research%20Report%2017.pdf.

53. E. Chavez Santos, J. T. Spector, J. Egbert, et al., "The Effect of the Participatory Heat Education and Awareness Tools (HEAT) Intervention on Agricultural Worker Physiological Heat Strain: Results From a Parallel, Comparison, Group Allocated Study," *BMC Public Health* 22, no. 1 (2022): 1746.

54. E. Mizelle, K. L. Larson, L. P. Bolin, and G. D. Kearney, "Fluid Intake and Hydration Status Among North Carolina Farmworkers: A Mixed Methods Study," *Workplace Health & Safety* 70, no. 12 (2022): 532–541.

55. T. Bodin, R. García-Trabanino, I. Weiss, et al., "Intervention to Reduce Heat Stress and Improve Efficiency Among Sugarcane Workers in El Salvador: Phase 1," *Occupational and Environmental Medicine* 73, no. 6 (2016): 409–416.

56. D. J. Smith, E. P. Ferranti, V. S. Hertzberg, and V. Mac, "Knowledge of Heat-Related Illness First Aid and Self-Reported Hydration and Heat-Related Illness Symptoms in Migrant Farmworkers," *Workplace Health & Safety* 69, no. 1 (2021): 15–21.

57. J. Mix, L. Elon, V. Vi Thien Mac, et al., "Hydration Status, Kidney Function, and Kidney Injury in Florida Agricultural Workers," *Journal of Occupational & Environmental Medicine* 60, no. 5 (2018): e253–e260.

58. A. J. Vega-Arroyo, D. C. Mitchell, J. R. Castro, et al., "Impacts of Weather, Work Rate, Hydration, and Clothing in Heat-Related Illness in California Farmworkers," *American Journal of Industrial Medicine* 62, no. 12 (2019): 1038–1046.

59. Y. Saito, O. Kashimura, T. Noda, and M. Sakurai, "Thermoregulatory Responses of Agricultural Workers at House Cultivation Work under Heat Environmental Conditions in Summer," *Japanese Journal of Biometeorology* 53, no. 2 (2016): 95–103.

60. B. Jacklitsch, W. J. Musolin, J. H. Kim, and N. C. Turner, eds. National Institute for Occupational Safety and Health. "Criteria for a Recommended Standard: Occupational Exposure to Heat and Hot Environments," in *Centers for Disease Control and Prevention, U.S. Department of Health and Human Services* (DHHS (NIOSH), 2016), https://www.cdc.gov/niosh/docs/2016-106/pdfs/2016-106.pdf.

61. L. E. Armstrong and J. Stoppani, "Central Nervous System Control of Heat Acclimation Adaptations: An Emerging Paradigm," *Reviews in the Neurosciences* 13, no. 3 (2002): 271–285.

62. P. L. Moseley, "Mechanisms of Heat Adaptation: Thermotolerance and Acclimatization," *Journal of Laboratory and Clinical Medicine* 123, no. 1 (1994): 48–52.

63. L. E. Armstrong and C. M. Maresh, "The Induction and Decay of Heat Acclimatisation in Trained Athletes," *Sports Medicine* 12, no. 5 (1991): 302–312.

64. M. D. Larrañaga, "Chapter 28 – Thermal Standards and Measurement Techniques," in *The Occupational Environment: Its Evaluation, Control, and Management*, ed. D. H. Anna. (American Industrial Hygiene Association, 2011).

Supporting Information

Additional supporting information can be found online in the Supporting Information section.
Supporting File: ajim70103-sup-0001-Supplemental_Material_accepted_modified.docx.