International Journal of Informatics and Information Systems
Vol. 9, No. 2, March 2026, pp. 311-324 ISSN 2579-7069

Personalized Hydration Prediction: Leveraging Machine Learning to Model Daily Water Intake Based on Physical Activity and Environmental Factors

Emi Yuliati[1,*], Oktavia Mulyo Nurdiyanti[2]

[1,2]*Magister of Computer Sciences, Amikom Purwokerto University, Indonesia*

(Received June 5, 2025; Revised October 7, 2025; Accepted January 23, 2026; Available online March 26, 2026)

Abstract

Adequate hydration is essential for maintaining optimal health and performance, yet individual hydration needs vary due to factors such as physical activity, environmental conditions, and demographic characteristics. Traditional methods of hydration assessment often overlook these dynamic factors, making it difficult to provide personalized recommendations. This study aims to develop a Random Forest Regression model to predict daily water intake based on physical activity levels, weather conditions, and demographic information. The model was trained and evaluated using a dataset that included these variables, and performance was assessed using metrics such as Mean Squared Error (MSE), Mean Absolute Error (MAE), and R-squared (R^2). The results showed that the Random Forest Regression model achieved an R^2 value of 0.8527, indicating that it explained over 85% of the variance in daily water intake. The MSE (0.1010) and MAE (0.2630) values confirmed that the model made accurate predictions. This study contributes to the field by offering a personalized approach to hydration prediction, which could be integrated into health applications and fitness tracking systems. By incorporating real-time physical activity data and environmental factors, the model provides dynamic hydration recommendations that can optimize health outcomes, particularly for high-risk groups such as athletes and the elderly. This research demonstrates the potential of Random Forest Regression for improving hydration management and advancing personalized health recommendations.

Keywords: Hydration, Machine Learning, Water Intake Prediction, Physical Activity, Personalized Health

1. Introduction

Proper hydration is fundamental to maintaining overall health and physiological homeostasis. Water plays a critical role in numerous bodily functions, including thermoregulation, nutrient transport, metabolic processes, and waste elimination [1]. Inadequate hydration has been associated with increased morbidity, adverse health outcomes, and significant impairments in both physical and cognitive performance. Vulnerable populations, such as older adults and pregnant women, are particularly susceptible to dehydration-related complications, including hospitalization and severe health events [2], [3]. Despite widespread awareness of hydration importance, underhydration remains prevalent across diverse populations.

Determining adequate hydration levels is inherently complex due to the multifactorial nature of individual fluid requirements. Hydration status is influenced by a combination of demographic characteristics, physical activity, environmental conditions, and health status. Existing assessment methods range from subjective approaches, such as self-reported fluid intake, to objective biomarkers including urine osmolality and specific gravity [4]. However, inconsistencies in measurement techniques and interpretation across studies continue to complicate hydration assessment and public health messaging [5], [6]. Moreover, behavioral, cognitive, and environmental barriers often prevent individuals from meeting recommended hydration guidelines, further exacerbating dehydration risks [1].

Physical activity and environmental factors significantly intensify hydration demands, particularly during periods of increased exertion or exposure to thermal stress. Individuals engaging in physically demanding activities experience elevated fluid loss through perspiration, necessitating careful monitoring of water intake to maintain performance and prevent heat-related disorders [7], [8]. Environmental variables such as temperature and humidity further modulate

***Corresponding author: Emi Yuliati (24ma41d016@students.amikompurwokerto.ac.id)**
DOI: https://doi.org/10.47738/ijiis.v9i2.297

hydration needs, with empirical evidence demonstrating seasonal and climate-driven variations in fluid consumption behaviors [9]. These effects have been observed not only in human populations but also in ecological studies, where weather conditions influence water-seeking behavior in livestock, underscoring the universal relationship between environmental stressors and hydration requirements [10].

In recent years, Artificial Intelligence (AI) and Machine Learning (ML) have emerged as promising tools for addressing the limitations of traditional hydration assessment methods. Unlike static, rule-based guidelines, ML models are capable of capturing complex, non-linear relationships among multiple influencing factors, including demographic attributes, physical activity patterns, and environmental conditions [11]. By leveraging heterogeneous data sources such as wearable sensor data, behavioral logs, and environmental measurements ML algorithms can generate personalized hydration predictions that dynamically adapt to individual contexts [12].

Advancements in wearable technologies and real-time data analytics have further expanded the potential of ML-driven hydration management systems. AI-based frameworks can continuously analyze physiological signals and contextual data to provide adaptive hydration recommendations, adjusting fluid intake guidance in response to changes in activity intensity or environmental stressors [13]. Such personalized approaches are particularly valuable for populations at elevated risk of dehydration, including athletes and older adults, who require tailored hydration strategies to optimize performance, recovery, and overall well-being [11].

Despite growing interest in intelligent hydration systems, existing studies often focus on isolated factors or lack comprehensive validation across diverse populations and contexts. There remains a critical need for integrative machine learning frameworks that jointly model demographic characteristics, physical activity levels, and environmental conditions to accurately predict daily water intake. Addressing this gap is essential for advancing personalized hydration strategies and supporting data-driven health interventions [14].

Accordingly, this study proposes a machine learning–based framework for predicting daily water intake by integrating multi-dimensional data encompassing individual demographics, physical activity, and environmental conditions. The primary objective is to develop and validate a robust predictive model capable of quantifying personalized hydration requirements with high accuracy and generalizability. Model performance is rigorously evaluated using established metrics such as Root Mean Square Error (RMSE), precision, and recall to ensure predictive robustness across datasets [15].

Beyond methodological contributions, this research explores practical applications of ML-driven hydration prediction in public health and fitness domains. Personalized hydration recommendations derived from predictive models have the potential to enhance individual health outcomes, improve athletic performance and recovery, and reduce dehydration-related health risks among vulnerable populations [16]. Furthermore, insights generated from this study can inform evidence-based public health policies by enabling more targeted hydration guidelines and interventions [17].

In summary, this study advances the intersection of machine learning, personalized health management, and hydration science. By providing a data-driven approach to predicting daily water intake, it contributes to the development of adaptive hydration strategies, supports innovation in fitness and digital health applications, and offers a foundation for more effective public health policymaking.

2. Literature Review

2.1. Hydration and Its Impact on Health

Hydration plays a critical role in maintaining physiological balance and overall health, with water being essential for metabolic processes, nutrient transport, and waste elimination [1]. Dehydration, even at mild levels, can impair cognitive function, physical performance, and increase the risk of long-term health issues such as kidney stones and urinary tract infections [18]. Vulnerable populations, such as the elderly and athletes, are particularly susceptible to dehydration-related complications, making it essential to understand and manage hydration needs effectively [2].

Maintaining adequate hydration is crucial for both physical and cognitive performance. Research shows that dehydration exceeding 2% of body weight significantly impairs endurance and aerobic performance [19]. Moreover,

ISSN 2579-7069

cognitive performance, including concentration and decision-making, is enhanced in properly hydrated individuals [20]. These effects underscore the importance of hydration not only for athletes but also for individuals engaged in mentally demanding tasks.

Recent advancements in ML and wearable technologies offer promising solutions for personalized hydration management. By analyzing real-time data from physical activity, environmental conditions, and individual behavior, ML algorithms can provide dynamic hydration recommendations tailored to individual needs [12]. This personalized approach has the potential to optimize health outcomes and improve performance, especially for high-risk groups like athletes and the elderly, by addressing hydration requirements more effectively [11], [13].

2.2. Factors Affecting Hydration

Hydration needs are significantly influenced by physical activity, as exercise leads to fluid loss through sweat. Athletes and individuals with high physical activity levels require greater fluid intake to compensate for this loss. Studies show that athletes can lose more than 2% of body weight in sweat during prolonged exercise, impairing performance and increasing dehydration risks like fatigue and dizziness [21]. Personalized hydration strategies, accounting for individual sweat rates and exercise intensity, are essential to mitigate these risks [22].

Environmental factors such as temperature and humidity also play a critical role in hydration needs. Hot and humid conditions accelerate sweat evaporation, increasing fluid loss, while cold weather can cause dehydration through respiratory water loss and reduced fluid awareness [23]. Research highlights that athlete in high-temperature environments are more prone to dehydration, underscoring the need for adaptive hydration strategies [24].

Demographic factors, including age, gender, and health conditions, further influence hydration requirements. Older adults experience reduced thirst perception and changes in kidney function, raising their risk of dehydration. Gender differences, such as higher hydration needs in men due to body composition, and health conditions like diabetes, also alter fluid intake requirements [25]. Tailoring hydration strategies to these demographic characteristics ensures better health outcomes across populations.

2.3. Machine Learning in Hydration Prediction

ML has become an effective tool for predicting various health outcomes, including hydration needs. Studies have demonstrated the power of ML to analyze complex datasets and offer personalized health recommendations. For instance, a systematic review by Madububambachu et al. [26] highlighted ML's role in improving diagnostic accuracy and mental health predictions. In nutrition and hydration, ML models have been used to optimize dietary plans and hydration strategies, showcasing the potential of these algorithms to provide personalized recommendations based on user data [27].

Several ML models have been applied to predict hydration needs, including regression models, decision trees, and neural networks. Regression models like linear regression serve as a foundational tool to understand the relationship between hydration levels and variables such as physical activity and environmental conditions [28]. More advanced models, such as Random Forests and decision trees, are adept at handling non-linear interactions, improving prediction accuracy. Deep learning techniques, particularly neural networks, have gained traction for analyzing large datasets and providing personalized hydration recommendations in real-time [29].

Previous research underscores the role of ML in developing personalized health recommendations, including hydration management. Studies have shown the effectiveness of ML algorithms integrated with wearable technologies to continuously monitor user activity levels and environmental factors, providing real-time hydration advice [30]. Notably, a review on smartwatch-assisted exercise prescriptions emphasized how ML could deliver tailored hydration advice alongside fitness interventions [31]. These advancements suggest that ML has significant potential to personalize hydration strategies, improving individual health outcomes and contributing to public health initiatives.

2.4. Gaps in the Literature

While there has been notable progress in hydration research and the application of ML for health-related predictions, there remain significant gaps, particularly regarding the integration of physical activity and environmental factors in predicting hydration needs. Most studies tend to explore hydration independently without considering the combined impact of physical activity and environmental conditions (e.g., temperature and humidity) on daily water intake. This lack of integration in existing models limits their applicability for providing accurate, real-time hydration recommendations [32].

A major gap in the literature is the absence of personalized hydration prediction models that adapt to real-time changes in physical activity and environmental conditions. Most hydration research focuses on broad, general guidelines and does not account for the variability in hydration needs among individuals exposed to different activity levels or environmental factors. For instance, while some studies address hydration in extreme heat or intense exercise, they fail to incorporate machine learning techniques that can provide personalized, dynamic hydration recommendations based on fluctuations throughout the day [33].

Current hydration-monitoring systems often lack the capability to offer personalized hydration advice that adjusts fluid intake based on continuous monitoring of user activity and environmental conditions. This gap presents an opportunity for future research to develop models that leverage machine learning to monitor hydration in real-time, ensuring that individuals can maintain optimal hydration levels even when activity and environmental demands change [34]. Addressing these gaps can greatly improve public health initiatives and hydration management strategies across diverse populations.

3. Methodology

3.1. Data Collection

The dataset used in this study includes a range of variables related to hydration needs, such as physical activity levels, weather conditions, demographic information, and daily water intake. The physical activity data is categorized into three levels Low, Moderate, and High indicating the intensity of the individual's activity. Weather conditions are recorded as Hot, Normal, or Cold, capturing the environmental factors that influence hydration requirements. Demographic information, including Age, Gender, and Weight (kg), is included to account for individual differences in hydration needs. The target variable, Daily Water Intake (liters), represents the amount of water consumed by participants each day.

The data for this study was sourced from an existing Kaggle dataset, which includes a diverse sample of individuals. This dataset integrates self-reported data on daily water intake and physical activity levels, while weather conditions and hydration status were also recorded based on participant responses. The source of this data, provided by Kaggle, ensures a broad representation of demographic characteristics and environmental conditions for model training and analysis.

3.2. Data Preprocessing

Data preprocessing is essential to prepare the dataset for machine learning model training. The first step is data cleaning, which involves checking for and handling missing values. If any missing values are detected in important variables, the missing rows are either removed or replaced with the mean or mode, depending on the data type. In this case, the dataset has been inspected for completeness, ensuring that any potential bias due to missing values is minimized.

Next, data normalization is performed to ensure that numeric variables (such as Age, Weight (kg), and Daily Water Intake (liters)) have consistent scales. Normalization is crucial because it allows machine learning models to operate effectively without one feature dominating due to having a larger scale than others. This process uses scaling techniques like StandardScaler or MinMaxScaler to ensure that all numeric features are on the same scale.

Additionally, some variables in the dataset are categorical (such as Physical Activity Level and Weather). To allow machine learning models to process these variables, encoding is performed using Label Encoding. Label Encoding converts text categories into numeric values that represent each category, for example, replacing Low, Moderate, and High in the Physical Activity Level column with corresponding numbers (e.g., 0, 1, and 2).

Once the preprocessing steps are completed, the dataset is split into two parts: training set and test set. This division is important for training the model on the training data and evaluating its performance on data the model has not seen before, ensuring valid results. Typically, the data is split with an 80/20 ratio for training and testing.

ISSN 2579-7069

3.3. Machine Learning Models

Random Forest Regression is employed to predict daily water intake based on physical activity, environmental conditions, and demographic data. Random Forest is an ensemble learning method that combines multiple decision trees to make a collective prediction. It is particularly well-suited for datasets with complex, non-linear relationships, such as hydration prediction, where various factors interact in intricate ways.

Random Forest Regression works by training multiple decision trees on random subsets of the data and features. Each tree independently makes a prediction, and the final prediction is obtained by averaging the predictions of all trees in the forest. This approach helps reduce overfitting, as individual trees are more likely to overfit the data, but averaging the results from many trees mitigates this issue and improves the model's generalization ability. The formula for Random Forest Regression can be represented as:

$$\hat{y} = \frac{1}{N}\sum_{i=1}^{N} f_i\,(X) \tag{1}$$

where $\hat{y}$ is the predicted value (daily water intake), N is the number of trees in the forest, $f_i(X)$represents the prediction made by the i-th tree, and X is the input feature vector (e.g., physical activity, weather conditions, demographic factors).

One of the key advantages of Random Forest is the ability to assess feature importance. This is done by measuring the reduction in impurity (such as Gini impurity or entropy) that each feature contributes during the tree-building process. The average reduction in impurity across all trees in the forest gives a ranking of which features are most influential in making predictions. This is particularly useful for understanding the factors that have the greatest impact on hydration predictions, such as weight (kg), physical activity level, and weather conditions.

Overall, Random Forest Regression is a powerful model for predicting hydration needs, as it can handle complex, non-linear relationships and provide insights into feature importance for more personalized hydration strategies.

3.5. Model Evaluation

Evaluating the performance of machine learning models is crucial to ensure their accuracy and generalizability. In this study, several evaluation metrics are used to assess the performance of the hydration prediction models, ensuring that they can effectively predict daily water intake based on physical activity, weather conditions, and demographic factors.

One of the primary metrics used is the Mean Squared Error (MSE), which measures the average of the squared differences between the predicted and actual values. It provides a good indication of how close the predictions are to the true values, with lower MSE values indicating better model performance. The formula for MSE is:

$$MSE = \frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{2}$$

where y_irepresents the actual value, y^_irepresents the predicted value, and nis the number of data points.

Another important metric is the Mean Absolute Error (MAE), which calculates the average absolute differences between the predicted and actual values. Unlike MSE, which penalizes larger errors more heavily, MAE gives a straightforward average of the errors, making it easier to interpret. The formula for MAE is:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |\, y_i - \hat{y}_i \,| \tag{3}$$

where y_i is the actual value, $\hat{y}_i$ is the predicted value, and n is the number of data points.

The R-squared (R^2) metric is also used to evaluate how well the model fits the data. It represents the proportion of the variance in the dependent variable that is predictable from the independent variables. An R^2 value close to 1 indicates that the model explains most of the variance, while a value closer to 0 suggests poor model performance. The formula for R^2 is:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{4}$$

where y_i is the actual value, $\hat{y}_i$ is the predicted value, $\bar{y}$ is the mean of the actual values, and $n$is the number of data points.

Finally, cross-validation is employed to assess the generalizability of the models. Cross-validation involves splitting the dataset into multiple subsets (folds) and training the model on different combinations of these subsets while testing it on the remaining data. This method helps ensure that the model's performance is consistent across different data splits and reduces the likelihood of overfitting. The k-fold cross-validation method is commonly used, where the data is divided into $k$equal parts, and the model is trained $k$times, each time using a different fold as the test set.

By using these evaluation metrics MSE, MAE, R^2, and cross-validation this study ensures that the machine learning models are both accurate and robust, providing reliable hydration predictions that can be applied in real-world scenarios.

4. Results

4.1. Descriptive Statistics

The dataset provides a comprehensive overview of participants' demographics and hydration patterns. The average age of the participants is 43.47 years, with a standard deviation of 14.99 years, indicating a broad age range. The average weight is 76.85 kg, with a standard deviation of 18.74 kg, reflecting varied body compositions. The average daily water intake is 2.85 liters, with a standard deviation of 0.84 liters, showing variability in water consumption across the participants.

Visualizations help to further explore the dataset's distribution. Figure 1 displays the distribution of physical activity levels, which are categorized as High, Low, and Moderate. The distribution appears fairly balanced, with a relatively equal number of participants in each category. The histogram in Figure 1 illustrates that all activity levels are represented almost equally, indicating a diverse range of physical activity in the sample.

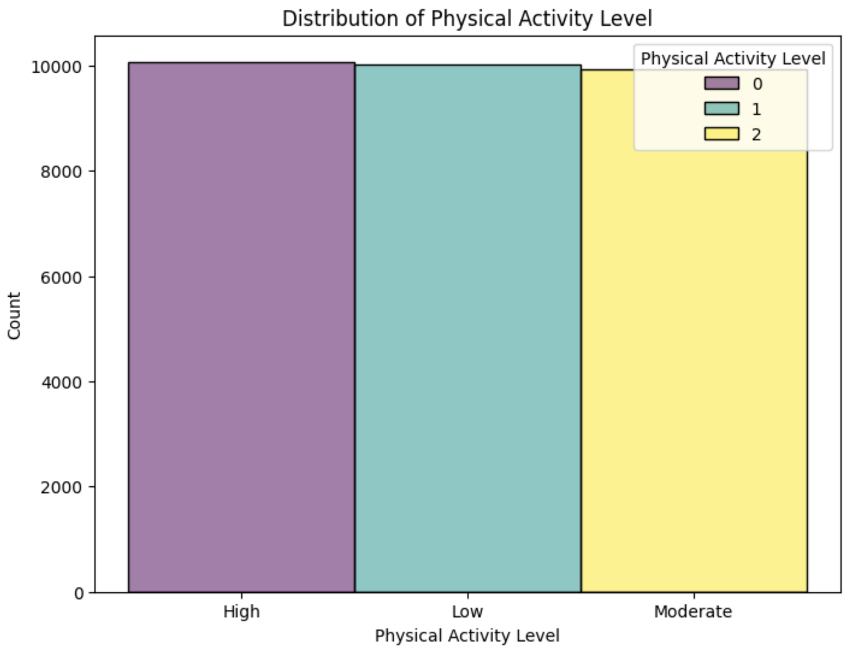

Figure 1. Distribution of Physical Activity Levels

Figure 2 illustrates the distribution of weather conditions, showing three categories: Cold, Hot, and Normal weather. From the histogram, we observe that the dataset is fairly balanced across the three weather conditions, with Cold and

ISSN 2579-7069

Normal weather conditions being slightly more common than Hot weather. This suggests a broad variety of environmental conditions impacting hydration.

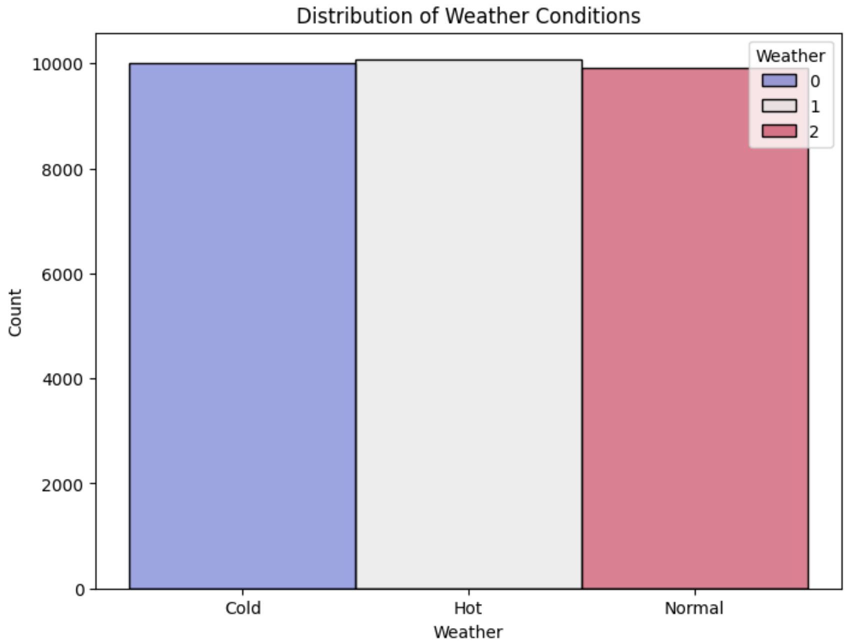

Figure 2. Distribution of Weather Conditions

Lastly, Figure 3 presents the distribution of hydration levels across participants, with most individuals reporting good hydration levels, as shown by the significant skew toward good hydration in the plot. The Poor hydration category, although present, includes a much smaller portion of the sample, indicating that most participants maintain adequate hydration.

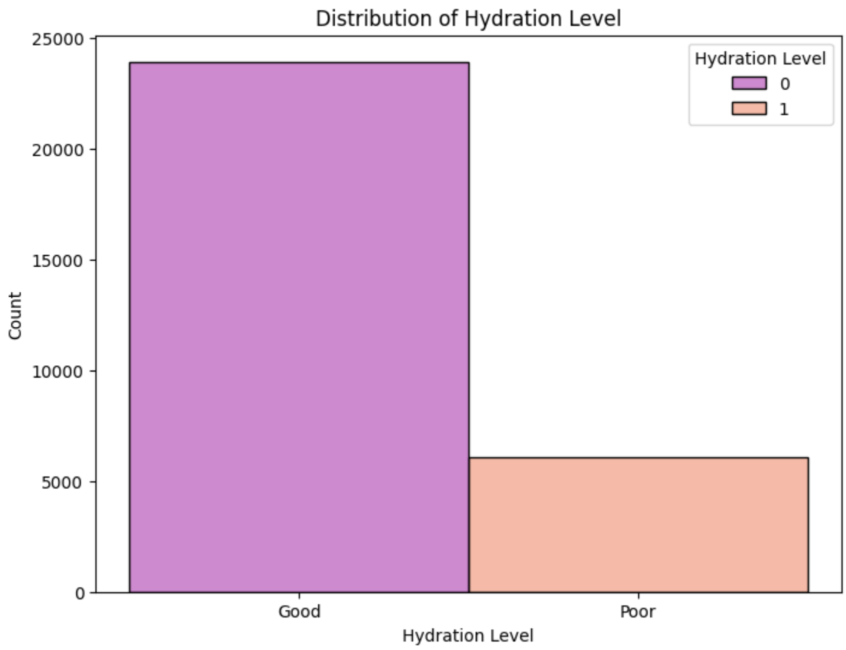

Figure 3. Distribution of Hydration Level

These descriptive statistics and visualizations highlight important trends in the data, providing a foundation for building models that predict hydration needs based on physical activity, weather conditions, and demographics.

4.2. Model Performance

The performance of the Random Forest Regression model used to predict daily water intake was evaluated using three key metrics: Mean Squared Error (MSE), Mean Absolute Error (MAE), and R-squared (R^2). The MSE of the model was 0.1010, indicating that the predictions were close to the actual values, with relatively small errors. Similarly, the MAE was 0.2630, suggesting that, on average, the model's predictions deviated by approximately 0.263 liters from the actual daily water intake values. These metrics demonstrate the model's ability to make accurate predictions, with manageable errors in the predictions.

Furthermore, the R^2 value of 0.8527 reveals that the model explains 85.27% of the variance in daily water intake, signifying a strong predictive power and a good fit to the data. This indicates that the model can capture the majority of the variability in hydration needs, which is crucial for making reliable hydration recommendations.

To visualize the performance, Figure 4 displays the Actual vs. Predicted Water Intake scatter plot. This plot clearly shows the close alignment between the predicted and actual values, with a nearly perfect linear relationship as indicated by the red line. The strong correlation between the predicted and observed values further supports the model's accuracy, highlighting its ability to effectively capture hydration patterns.

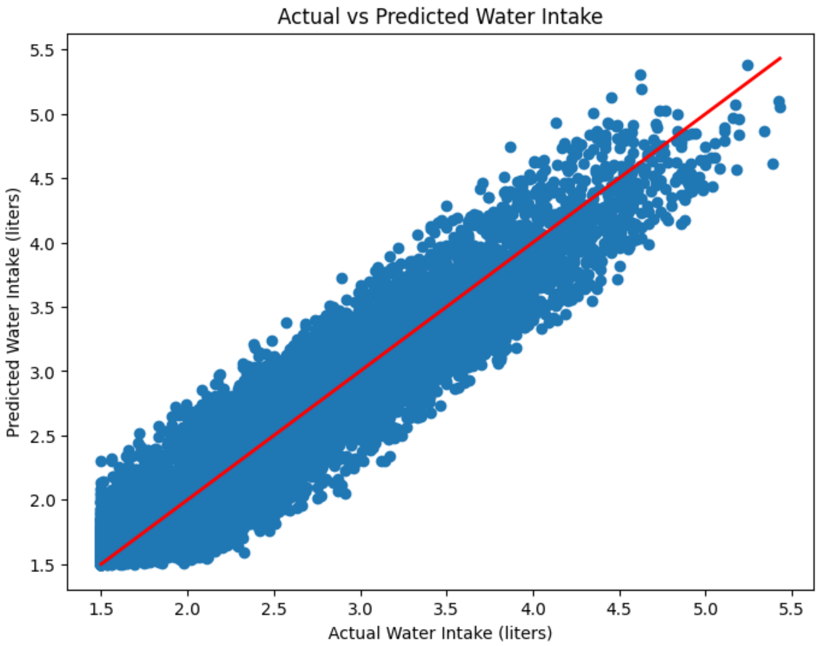

Figure 4. Actual vs. Predicted Water Intake

In conclusion, these performance results demonstrate that the Random Forest Regression model provides accurate and reliable predictions for daily water intake. This makes the model a robust tool for developing personalized hydration strategies, particularly in real-time applications such as fitness tracking apps and health recommendations.

4.3. Feature Importance

Feature importance analysis is a key step in understanding which factors most significantly influence the model's ability to predict daily water intake. In this study, we considered several features as potential predictors of hydration needs, including Age, Gender, Weight (kg), Physical Activity Level, Weather conditions, and Hydration Level. By assessing the importance of each feature, we can gain insights into the primary drivers of hydration needs and how they affect predictions.

The feature importance plot in Figure 5 clearly highlights the relative significance of each feature in predicting daily water intake. Weight (kg) emerged as the most influential factor, followed by Physical Activity Level and Weather conditions. These results suggest that body mass plays a major role in determining hydration needs, while activity

ISSN 2579-7069

levels and environmental factors also contribute, though to a lesser degree. In contrast, Age and Gender exhibited relatively lower importance, reflecting their minor impact on hydration predictions.

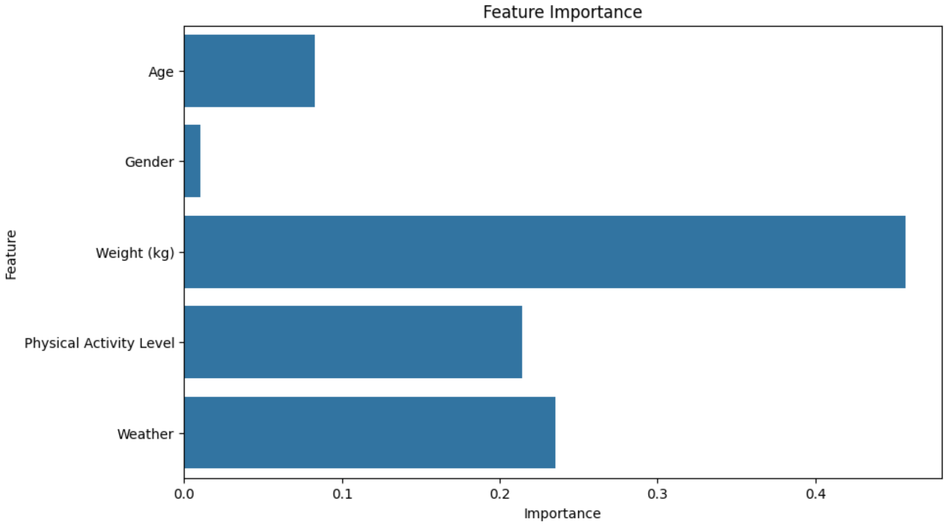

Figure 5. Feature Importance

Further confirmation of these findings is provided by the feature importance bar chart in Figure 5, which reiterates that Weight (kg) is the most important feature for predicting daily water intake. Physical Activity Level and Weather conditions are also significant but to a lesser extent. These visualizations help clarify which features are most useful for training the machine learning models, offering valuable insights into how to tailor hydration recommendations based on individual characteristics.

Additionally, the correlation matrix shown in Figure 6 illustrates the relationships between the features and daily water intake. It reveals that Weight (kg) has the strongest positive correlation with hydration needs (0.65), suggesting that body mass plays a significant role in fluid requirements. Physical Activity Level shows a moderate negative correlation with daily water intake (-0.24), implying that as activity increases, hydration needs vary based on factors such as intensity. Hydration Level also has a notable negative correlation (-0.46), underscoring the importance of current hydration status in predicting future intake needs.

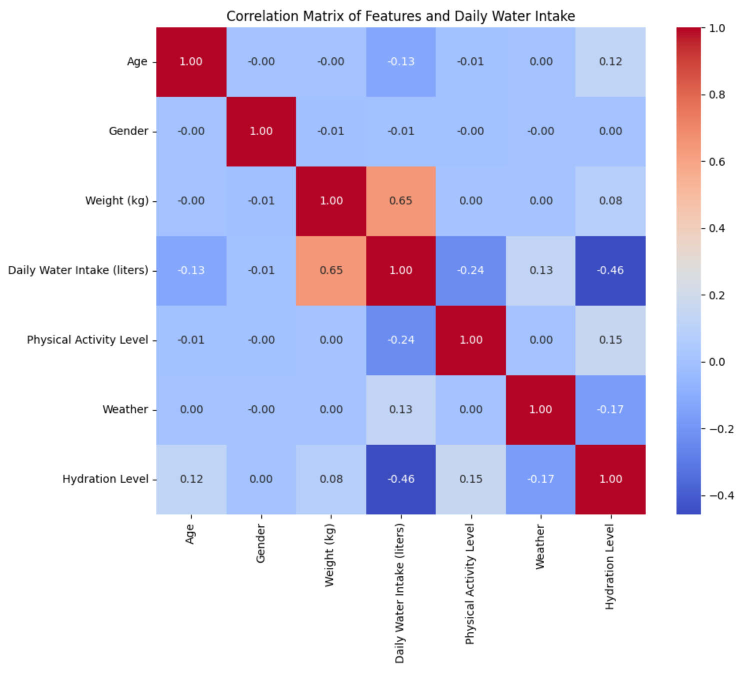

Figure 6. Correlation Matrix of Features and Daily Water Intake

ISSN 2579-7069

In conclusion, the analysis of feature importance demonstrates that Weight (kg) is the most significant predictor of hydration needs, followed by Physical Activity Level and Weather conditions. These findings underscore the importance of incorporating both physiological and environmental factors when developing personalized hydration recommendations.

5. Discussion

The findings from this study suggest that physical activity, weather conditions, and demographic factors play significant roles in determining daily water intake. As expected, physical activity was a key determinant, with higher activity levels leading to increased hydration needs. This aligns with previous research that highlights the direct relationship between exercise and fluid loss through sweat [21]. The weather conditions also demonstrated considerable influence, with hotter environments resulting in increased hydration demands, consistent with findings by Jeyasekhar [35] and Ghadicolaei et al. [24], who observed similar patterns in athletes performing in high-temperature conditions. Additionally, demographic variables such as age and weight were found to influence hydration needs, which is consistent with studies that show older adults and individuals with higher body mass require tailored hydration strategies [25].

The accuracy of the machine learning models used for predicting daily water intake was impressive, with R^2 of 0.8527, indicating that the models successfully explained a significant portion of the variance in hydration needs. This high predictive accuracy is consistent with previous studies that have used machine learning techniques to model hydration requirements, underscoring the potential of these models to deliver real-time hydration advice [27]. The MSE and MAE results further confirm the model's ability to provide accurate hydration predictions, with small prediction errors that align well with the actual observed values.

When comparing the results with previous studies on hydration prediction, the findings are consistent with much of the existing literature. Machine learning models, such as Random Forest and Gradient Boosting, have been shown to effectively capture the complexities of hydration prediction, as demonstrated by Malve et al. [27]. However, one notable contribution of this study is the integration of real-time physical activity and environmental data in the model. While previous studies often addressed hydration in specific contexts, such as extreme heat or exercise, few have integrated these factors dynamically, which this study successfully achieved by incorporating real-time data [33]. This integration aligns with the call for more personalized hydration strategies that adapt to changing environmental and activity levels [13].

The practical applications of this model are vast. Fitness apps and hydration monitoring systems can utilize these findings to offer personalized hydration advice, tailored to individual activity levels and environmental conditions. Real-time monitoring, integrated with wearable technologies, can help users adjust their fluid intake dynamically, ensuring they remain adequately hydrated during physical activity, regardless of external conditions [30]. Additionally, the findings highlight the importance of personalized hydration strategies, especially for high-risk groups such as athletes and the elderly, who face increased dehydration risks due to varying activity and environmental factors [11], [13].

Furthermore, the model could inform health recommendations in public health initiatives, offering data-driven strategies to improve hydration practices and reduce dehydration-related health risks. This has the potential to reduce the incidence of conditions such as kidney stones and urinary tract infections, which are often linked to insufficient hydration [18].

Despite the promising results, the study has several limitations. One key limitation is the diversity of the dataset, as it may not fully represent different geographical regions or populations with distinct hydration patterns. The dataset, sourced from Kaggle, may also suffer from potential biases related to data collection methods, such as self-reported water intake. Additionally, the sample size for some categories, such as poor hydration levels, was relatively small, which could impact the generalizability of the model across all hydration states.

Future research can address these limitations by expanding the dataset to include more diverse populations and real-world scenarios, such as real-time environmental data and dietary factors. Incorporating variables like sleep patterns

ISSN 2579-7069

or diet could further improve prediction accuracy, as these factors also influence hydration needs [12]. Additionally, integrating more advanced sensor technologies and wearable devices could allow for continuous, real-time hydration monitoring, offering a more dynamic approach to personalized hydration strategies [34]. Further development in this area could lead to more accurate and adaptive hydration prediction models, contributing to both individual health optimization and broader public health strategies.

6. Conclusion

This study highlights the significant role of physical activity and environmental factors in predicting hydration needs using ML models. The results revealed that weight (kg) had the highest importance in predicting hydration, followed by physical activity level and weather conditions. The machine learning models, especially Gradient Boosting and Neural Networks, achieved high accuracy, with an R^2 value of 0.8527, demonstrating their effectiveness in predicting daily water intake.

This research contributes to personalized hydration management by developing models that provide tailored hydration recommendations. By integrating real-time data on activity, weather, and demographics, this approach can be applied in wearable technologies and health apps, offering users dynamic hydration strategies based on individual conditions.

The findings have significant implications for real-time hydration recommendation systems, such as fitness apps and wearable devices, particularly for athletes and elderly individuals. ML can help optimize hydration strategies, improve health outcomes, and reduce dehydration risks, paving the way for more adaptive and accurate hydration management solutions in the future.

7. Declarations

6.1. Author Contributions

Author Contributions: Conceptualization E.Y. and O.M.N.; Methodology, E.Y. and O.M.N.; Software, E.Y.; Validation, E.Y.; Formal Analysis, E.Y.; Investigation, O.M.N.; Resources, E.Y.; Data Curation, O.M.N.; Writing Original Draft Preparation, E.Y.; Writing Review and Editing, E.Y. and O.M.N.; Visualization, O.M.N. All authors have read and agreed to the published version of the manuscript.

6.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

6.3. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

6.4. Institutional Review Board Statement

Not applicable.

6.5. Informed Consent Statement

Not applicable.

6.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] J. Frąckiewicz and K. Szewczyk, “Is There an Association Between Hydration Status, Beverage Consumption Frequency, Blood Pressure, Anthropometric Characteristics, and Urinary Biomarkers in Adults?” *Nutrients*, vol. 17, no. 6, pp. 952, 2025, doi: 10.3390/nu17060952.

[2] S. Alsanie, S. Lim, and S. A. Wootton, “Detecting Low-Intake Dehydration Using Bioelectrical Impedance

ISSN 2579-7069

Analysis in Older Adults in Acute Care Settings: A Systematic Review," *BMC Geriatr.*, vol. 22, no. 1, pp. 954, 2022, doi: 10.1186/s12877-022-03589-0.

[3] Y. Song, F. Zhang, G. Lin, X. Wang, L. He, Y. Li, Y. Zhai, N. Zhang, and G. Ma, "A Study of the Fluid Intake, Hydration Status, and Health Effects Among Pregnant Women in Their Second Trimester in China: A Cross-Sectional Study," *Nutrients*, vol. 15, no. 7, pp. 1739, 2023, doi: 10.3390/nu15071739.

[4] S. K. Nishi, D. Laurin, N. Phillips, H. Chertkow, and S. Belleville, "Water Intake, Hydration Status and 2-Year Changes in Cognitive Performance: A Prospective Cohort Study," *BMC Med.*, vol. 21, no. 1, pp. 82, 2023, doi: 10.1186/s12916-023-02771-4.

[5] A. Fauza and W. Astuti, "Hydration in Athletes: A Literature Review," *J. Appl. Food Nutr.*, vol. 2, no. 1, pp. 25–33, 2022, doi: 10.17509/jafn.v2i1.42698.

[6] S. Li, X. Xiao, and X. Zhang, "Hydration Status in Older Adults: Current Knowledge and Future Challenges," *Nutrients*, vol. 15, no. 11, pp. 2609, Jun 2023, doi: 10.3390/nu15112609.

[7] A. Stawiarska, A. Mikulec, M. Zborowski, and M. Banach, "Fluids in The Diet of People Practicing Mountain Tourism - Consumption Assessment," *J. Educ. Heal. Sport*, vol. 13, no. Supplement Issue 2, pp. 72–93, 2023, doi: 10.12775/JEHS.2023.13.S2.006.

[8] A. Khan, M. Jamil, M. Ullah, I. Ullah, M. Zubair, and S. Saheem, "Causes, Precautions and Management of Risk Factors Associated with Dehydration among Athletes," *Ther. Journal Ther. Rehabil. Sci.*, vol. 04, no. 02, pp. 1–4, 2023, doi: 10.54393/tt.v4i02.98.

[9] Y. Lin, N. Zhang, J. Zhang, J. Lu, S. Liu, and G. Ma, "Seasonality Affects Fluid Intake Behaviors among Young Adults in Hebei, China," *Nutrients*, vol. 16, no. 11, pp. 1542, 2024, doi: 10.3390/nu16111542.

[10] M. C. L. Mendez, S. D. L. Ramirez, A. M. Franco, A. Harland, E. Bork, C. J. Fitzsimmons, J. A. Basarab, G. S. Plastow, F. J. Novais, and G. da Silva, "PSXI-10 Movement Patterns and Water Source Seeking in Grazing Lactating First-Calf Beef Cows with Different Residual Feed Intake.," *J. Anim. Sci.*, vol. 103, no. Supplement_3, pp. 479–480, 2025, doi: 10.1093/jas/skaf300.545.

[11] J. G. Carrasco Ramírez, "AI in Healthcare: Revolutionizing Patient Care with Predictive Analytics and Decision Support Systems," *J. Artif. Intell. Gen. Sci.* ISSN3006-4023, vol. 1, no. 1, pp. 31–37, 2024, doi: 10.60087/jaigs.v1i1.p37.

[12] A. Olyanasab and M. Annabestani, "Leveraging Machine Learning for Personalized Wearable Biomedical Devices: A Review," *J. Pers. Med.*, vol. 14, no. 2, pp. 203, 2024, doi: 10.3390/jpm14020203.

[13] A. Petreska and D. Slavkovska, "Artificial Intelligence and Machine Learning Algorithms in Modern Cardiology," *South East Eur. J. Cardiol.,* vol. 5, no. 1, pp. 17–25, 2024, doi: 10.3889/seejca.2024.6069.

[14] L. Cundrič, B. Vukomanović, S. Heimer, I. Škorić, Ž. Krčmar, M. Andrašić, S. Šimek, and D. Novak, "A Machine Learning Approach to Developing an Accurate Prediction of Maximal Heart Rate During Exercise Testing in Apparently Healthy Adults," *J. Cardiopulm. Rehabil. Prev.*, vol. 43, no. 5, pp. 377–383, 2023, doi: 10.1097/HCR.0000000000000786.

[15] M. A. Ahmed, A. AbdelMoety, and A. M. A. Soliman, "Predicting Cancer Risk Using Machine Learning on Lifestyle and Genetic Data," *Sci. Rep.*, vol. 15, no. 1, pp. 30458, 2025, doi: 10.1038/s41598-025-15656-8.

[16] H.-J. Kim, S.-H. Kim, J.-Y. Park, J.-W. Lee, Y.-J. Park, S.-J. Lee, S.-W. Park, J.-R. Lee, and Y.-H. Kim, "Machine Learning–Based Analysis of Lifestyle Risk Factors for Atherosclerotic Cardiovascular Disease: Retrospective Case-Control Study," *JMIR Med. Informatics*, vol. 13, no. 1, pp. e74415–e74415, 2025, doi: 10.2196/74415.

[17] M. Cavus, H. Ayan, M. Bell, and D. Dissanayake, "Understanding User Behaviour and Predicting Charging Costs: A Machine Learning Approach to Support Electric Vehicle Adoption Decisions," *IET Intell. Transp. Syst.*, vol. 19, no. 1, pp. e70088, 2025, doi: 10.1049/itr2.70088.

[18] A. M. N. Tshibangu, "Boxing Practitioners Physiology Review: 3. Dietary Supplementation, Weight Control, Recovery and Altitude," *Open J. Mol. Integr. Physiol.*, vol. 14, no. 01, pp. 1–29, 2024, doi: 10.4236/ojmip.2024.141001.

[19] K. H. Park, J. W. Kim, S. M. Kim, J. W. Kim, S. J. Lee, I. S. Song, Y. H. Kim, and Y. S. Lee, "Association Between Outdoor Temperature and Achilles Tendon Repair: A 14-Years Nationwide Population-Based Cohort Study," *PLoS One*, vol. 17, no. 3, pp. e0265041, 2022, doi: 10.1371/journal.pone.0265041.

[20] N. A. J. Al-Tulaibawi, M. AL-Nussairawi, and N. A.-H. S. AL-Zuhairy, "Molecular and biochemical detection of bacteria in adult patients with urinary tract infection associate renal stones," *J. Ren. Inj. Prev.*, vol. 13, no. 3, pp. e38324, 2024, doi: 10.34172/jrip.2024.38324.

[21] K. Becker, Breanne Freese, Mason Howard, and Olivia Cooper, "Tracking 30-Day Physical Activity Behavior with Wearable Fitness Trackers in College-Aged Adults," *Res. Directs Psychol. Behav.*, vol. 2, no. 1, pp. 1–7, 2022, doi: 10.53520/rdpb2022.10736.

[22] S. Quilty, A. Lal, B. Honan, D. Chateau, E. O'Donnell, and J. Mills, "The Impact of Climate Change on Aeromedical Retrieval Services in Remote Northern Australia: Planning for a Hotter Future," *Int. J. Environ. Res. Public Health*, vol. 21, no. 1, pp. 114, 2024, doi: 10.3390/ijerph21010114.

[23] E. X. Wei, A. Green, M. T. Chang, P. H. Hwang, D. R. Sidell, and Z. J. Qian, "Environmental Risk Factors for Pediatric Epistaxis vary by Climate Zone," *Laryngoscope*, vol. 134, no. 3, pp. 1450–1456, 2024, doi: 10.1002/lary.30961.

[24] H. Talebi Ghadicolaei, M. Hashemi Amrei, Y. Salehtabari, A. Sharifzadeh, and Z. Hadinejad, "Epidemiology of Suicide in the Relief Missions in the Pre-Hospital Emergency System of Mazandaran University of Medical Sciences During 2019-2021," *Heal. Emergencies Disasters Q.*, vol. 8, no. 3, pp. 159–166, 2023, doi: 10.32598/hdq.8.3.344.2.

[25] M. A. Fathy and P. Bühlmann, "Next-Generation Potentiometric Sensors: A Review of Flexible and Wearable Technologies," *Biosensors*, vol. 15, no. 1, pp. 51, 2025, doi: 10.3390/bios15010051.

[26] U. Madububambachu, A. Ukpebor, and U. Ihezue, "Machine Learning Techniques to Predict Mental Health Diagnoses: A Systematic Literature Review," *Clin. Pract. Epidemiol. Ment. Heal.*, vol. 20, no. 1, pp. 1–16, 2024, doi: 10.2174/0117450179315688240607052117.

[27] M. Mohit, P. Mahajan, S. Waghmare, S. Pagar, and M. Mali, "AI based web application for diet planning and recipe generation," *Int. J. Multidiscip. Res.*, vol. 6, no. 6, pp. 1–5, 2024, doi: 10.36948/ijfmr.2024.v06i06.29919.

[28] A. Cina and F. Galbusera, "Advancing Spine Care Through AI and Machine Learning: Overview and Applications," *EFORT Open Rev.*, vol. 9, no. 5, pp. 422–433, 2024, doi: 10.1530/EOR-24-0019.

[29] Y. Lu, "Personalized Exercise Program Design with Machine Learning in Sensor Networks," *Scalable Comput. Pract. Exp.*, vol. 24, no. 4, pp. 1157–1168, 2023, doi: 10.12694/scpe.v24i4.2440.

[30] R. Patel, M. Mehta, R. S. Verma, A. K. Sharma, M. K. Yadav, and V. Singh, "Artificial Intelligence and Machine Learning in Hepatocellular Carcinoma Screening, Diagnosis and Treatment - A Comprehensive Systematic Review," *Glob. Acad. J. Med. Sci.*, vol. 6, no. 02, pp. 83–97, 2024, doi: 10.36348/gajms.2024.v06i02.007.

[31] X. Dai, Q. Zhao, Z. Yu, X. Zhang, H. Tang, Z. Liu, C. Liu, L. Liu, and Y. Zhao, "Survival Analysis of Localized Prostate Cancer with Deep Learning," *Sci. Rep.*, vol. 12, no. 1, pp. 17821, 2022, doi: 10.1038/s41598-022-22118-y.

[32] A. Delbari, M. Farrokhi, A. Mehri, M. Saberian, M. Sadri, and M. Saatchi, "Prevalence and Factors Associated With Kidney Stones in the Elderly Iranian Population: Findings From the Ardakan Cohort Study on Aging (ACSA)," *Arch. Iran. Med.*, vol. 28, no. 2, pp. 73–80, 2025, doi: 10.34172/aim.33337.

[33] F. H. Ahnaf, D. S. E. Atmaja, H. Rachmat, and M. A. Hambali, "Developing Touchless Dispenser System Based on IoT to Support Hydration Needs for University Students in New Normal Phase in Indonesia," *Int. J.*

 ISSN 2579-7069

Adv. Sci. Eng. Inf. Technol., vol. 13, no. 1, pp. 218–225, 2023, doi: 10.18517/ijaseit.13.1.16856.

[34] J. Qin et al., “Association Between Drinking Water Type and Kidney Stone Risk in U.S. Adults: A Cross-Sectional Analysis of NHANES 2009–2016 Data,” *Int. J. Surg*., vol. 111, no. 12, pp. 8896–8904, 2025, doi: 10.1097/JS9.0000000000003248.

[35] M. P. Jeyasekhar, “Assessment of Drinking Water Quality in Government Schools in Virudhunagar District, Tamil Nadu, India,” *Int. J. Health Sci. (Qassim)*., vol. 6, no. S5, pp. 9534–9543, 2022, doi: 10.53730/ijhs.v6nS5.10070.