Development of an Intelligent Digital Water Footprint Calculator for Sustainable Resource Management using Machine Learning and IoT-Based Data Analytics

line 1: 1st Given Name Surname
line 2: *dept. name of organization (of Affiliation)*
line 3: *name of organization (of Affiliation)*
line 4: City, Country
line 5: email address or ORCID

line 1: 2nd Given Name Surname
line 2: *dept. name of organization (of Affiliation)*
line 3: *name of organization (of Affiliation)*
line 4: City, Country
line 5: email address or ORCID

line 1: 3rd Given Name Surname
line 2: *dept. name of organization (of Affiliation)*
line 3: *name of organization (of Affiliation)*
line 4: City, Country
line 5: email address or ORCID

line 1: 4th Given Name Surname
line 2: *dept. name of organization (of Affiliation)*
line 3: *name of organization (of Affiliation)*
line 4: City, Country
line 5: email address or ORCID

line 1: 5th Given Name Surname
line 2: *dept. name of organization (of Affiliation)*
line 3: *name of organization (of Affiliation)*
line 4: City, Country
line 5: email address or ORCID

line 1: 6th Given Name Surname
line 2: *dept. name of organization (of Affiliation)*
line 3: *name of organization (of Affiliation)*
line 4: City, Country
line 5: email address or ORCID

***Abstract*— As the world struggles with a water shortage, we need better ways to manage them. This study introduces an intelligent Digital Water Footprint (DWF) calculator that combines Internet of Things (IoT)-based data analytics with machine learning (ML) algorithms. The system can be used to evaluate and analysis instantaneously direct and indirect water consumption of an individual, household, and institution. This goes beyond the fixed assessment. It acquires data (i.e. IoT sensors) dynamically and follows the pattern of usage of the water. Machine learning models are now being used for predictive analytics, which is used to forecast future water demand. They also support in identifying the opportunities to optimize water use efficiency. The architecture of the system will be an IoT data acquisition module, a powerful data preprocessing and integration layer, analytical modeling (supervised and unsupervised ML) and visualization. The mathematical frameworks for components of blue, green and grey water footprint are bringing together and advanced ML algorithms for pattern recognition and forecasting. The calculator has been experimentally validated based on the real and simulated IoT datasets. The experimental validation reveals that the calculator is accurate in predicting water consumption as the Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) has very small values. The intelligent DWF calculator is superior to other methods in terms of showing real-time adaptability, prediction power. The results of this study are important for making informed decision on the allocation of water resources, policy making and water conservation activities. This innovation is contributing to global sustainability targets by making a more predictive and granular tool for water management.**

Keywords—Digital Water Footprint, IoT, Machine Learning, Sustainable Water Management, Predictive Analytics, Water Conservation.

I. Introduction

Increased demand from many factors is putting pressure on fresh water resources globally. Managing our water resources properly will ensure that environment and economy remain stable. Understanding how much water we use is a first step toward managing our water use better. Water footprint (WF) is a comprehensive indicator that was introduced to measure the total volume of freshwater which is used to produce the goods and services consumed by an individual, community or business [3]. Traditional water footprint assessments are helpful, but they aren't very granular and do not provide real-time data. Also, they don't predict future trends which makes them less useful. By using IoT and ML, an intelligent Digital Water Footprint (DWF) calculator has been introduced to overcome the limitations addressed by this research.

II. Background and Motivation

Water footprint evaluation is often based on aggregate data that is often static in nature. This limits their use. With the increased availability of IoT devices, there is an opportunity to collect high resolution data on water use at any scale, including for individual appliances or complete urban water distribution networks [4][5]. Combining real-time data with machine learning analysis can help us get a better understanding of patterns of water use. This is an integration which enables to accurately measure and detect inefficiencies, forecast of demands and intervene for water saving. This research aims to develop an advanced and versatile tool that can contribute to the sustainable water management [2] with the power of intelligence from data analysis.

III. Objectives and Scope

The main goal of this research is the development and validation of an intelligent calculator for Digital Water Footprint that will be able to offer real-time monitoring, comprehensive analysis and predictive prediction of the consumption of water. Specific goals are to:

1) *Design of a system architecture, which is modular and effectively integrates the IoT-based data acquisition with a robust data processing framework.*

2) *Implementing mathematical models* for calculating blue, green, and grey water footprints based on diverse data inputs.

3) *Developing and applying machine learning algorithms* for accurate water consumption prediction and anomaly detection.

4) *Creating an intuitive visualization and user interface* to present complex water footprint data and predictive insights in an accessible manner.

5) *Conducting experimental validation* of the proposed system using empirical or simulated IoT datasets to assess its performance and efficacy.

The scope of this study covers the conceptualization, architectural design, algorithmic development and validation of DWF calculator. It is oriented towards the residential, commercial and agricultural contexts, taking into account the direct use of water as well as the indirect use of water integrated with the consumption of energy and products. This work does not reach the level of full deployment and checking of the long term societal impact which is a future research direction.

IV. Significance for Sustainable Resource Management

The intelligent Digital Water Footprint calculator can make great contributions to the sustainable management of resources. By providing real-time granular information of water consumption, it empowers individuals, organizations, and policymakers to make informed decisions that promote water conservation and efficiency. The predictive capabilities of the system enable the management of water demand in a proactive manner and minimise the impact of possible water stress [1]. For example, by anticipating periods of high demand, it is possible to schedule the water supply in a better way or by identifying wasteful consumption, it is possible to develop more targeted intervention strategies. This approach helps to promote a change from reactive to proactive water management (resilience to hydrological drought and permanent water deficiency) [6]. Furthermore, the DWF calculator can be used as a basic tool in the development of strategies for water allocation, in the agricultural and trade policies and in the national accounting systems for water [2]. Ultimately, this research gives technological solution that align the global sustainable development goals, especially SDG 6: Clean Water and Sanitation by promoting efficient water use and reduce water pollution [7].

V. Literature Review

Understanding the present situation in water footprint assessment and digital technologies for water management is basic to set up the context for this research. This review summarizes some of the established frameworks of water footprints and an examination of digital approaches, and identifies important research gaps that the proposed intelligent DWF calculator aims to address.

VI. Established Water Footprint Frameworks

The concept of the water footprint has changed greatly since its birth, with multiple methodologies that have been developed to measure the consumption of water on a number of different scales. These frameworks provide the foundation for more sophisticated and data-driven approaches to be built up on top of them.

A. Water Footprint Network (WFN) Methodology

The Water Footprint Network (WFN) has a comprehensive methodology for evaluating water footprints and distinguishes the use of water into three broad categories: blue, green and grey water [3]. Blue water is surface water and ground water that is used. Green water is the quantity of rain water used, which mostly is of interest in agriculture and forestry. Grey water is a measure of the volume of fresh water required to assimilate the load of pollutants to ambient water quality requirements. The WFN methodology provides a standardized approach for calculating these components at product, process and organizational levels to provide comparability and allow for the identification of hotspots in the supply chains [2]. While good in terms of increasing awareness and providing a framework to assess, the WFN approach could very often be based on static or averaged data - meaning that it could well miss some of the spatial and temporal variabilities of water availability and consumption [1].

B. ISO 14046 Standard for Water Footprinting

Based on the work done so far, the International Organization for Standardization (ISO) developed ISO 14046:2014, which is a standard for water footprint assessment [7]. This standard gives principles, requirements and guidelines for the conduction and reporting of a water footprint assessment, which may only consider water consumption or water pollution, or both. A major aspect of ISO 14046 is to focus on environmental impacts in connection with water, often in a life cycle assessment (LCA) perspective. It provides flexibility in the selection of relevant impact categories and indicators and also ensures that assessments must be based on specific objectives and geographical situations. The standard calls for transparency and reproducibility for water footprint studies that can make the results more credible for stakeholders and decision-makers [7]. However, its rigorous specifications on data collection and evaluation of impact can prove to be demanding on the resources and thus add as an obstacle for real-time applications or assessments that need continuous updating.

C. Life Cycle Assessment (LCA) Models and Integration

Life Cycle Assessment (LCA) - It is a method of determining environmental impacts for all stages of the life of a product or service, i.e. from raw material extraction through production, through use, and even disposal. Water use is a mostly important impact category in the framework of LCA. [7] LCA models include the ideas of water footprint to describe the total collection of cumulative water impact along the supply chain of the product to provide a more complete solution of indirect water consumption than direct operational water consumption. This integration can help to identify the water hot spots in complex production systems and to support decisions that can contribute to decrease the overall environmental burdens. While LCA presents a picture of the whole it is composed of a lot of often difficultly collectible data and not easy modeling methods and may have their constraints in relation to their real-time usability and dynamic monitoring potential. The data input into LCA is often aggregated and not an accurate reflection of the instantaneous changes in the water availability and water consumption patterns and hence highlights the need for dynamic approaches.

VII. Digital Approaches in Water Footprint Assessment

The advent of digital technologies, in particular IoT and M.L itself has the transformative potential of boosting water footprint assessments and upgrading from calculating static, to predicting dynamism.

A. IoT-Based Water Usage Monitoring Systems

The internet of things has changed the way that data is collected from various industries and water management is no exception. IoT based systems use networks of sensors, which read the flow and quality and consumption of water, in real-time [4][5]. These sensors are away in a wide range of atmosphere from intelligent houses to food fields as well as the industry facilities and supply granular info regarding water use at a particular time point and in particular locations [8]. Examples are smart meters that monitor household water consumption, moisture sensors for agriculture or flow meters for industrial processes. The continuous stream of data produced by IoT devices enables exact identification of water-intensive activities, leakage identification and evaluation of water use efficiency [4]. Furthermore, such systems can provide instant feedback to the users, allowing them to make necessary changes to their consumption patterns. Research in this area has been conducted to explore energy-efficient monitoring for Water Distribution System (WDSs) via graph theory to reduce communication among sensors and showed the high potential of achieving accurate flow estimation with low energy consumption. Similarly, IoT solutions have been proposed for monitoring water quality and pilferage in supply systems with the help of machine learning as a tool for decision-making.

B. Machine Learning in Water Consumption and Demand Prediction

Machine learning (ML) algorithms have demonstrated a significant capability in analyzing complex datasets and making relevant and accurate predictions that detect patterns and make a prediction of water consumption and demand, and thus they are extremely suitable in the water consumption and demand forecast. Various techniques of ML, Artificial Neural Networks (ANNs), Support vector Regression (SVR), RandomForests and Extremely Random Trees have been tried for predicting water usage. These models are able to use past data on consumption along with contextual factors like the weather pattern, the time of day, day of the week, human mobility data etc to arrive at more accurate forecasts. For example, it has been proved that combining human mobility information could enhance the prediction accuracy of water demand. 90.4% could be achieved when using random forest. Similarly, Multilayer Perceptron (MLP) have achieved better results in predicting short term urban water demand as compared to other non-linear models. ML models can also be used to identify anomalies in water consumption suggesting the presence of a leak or unusual consumption patterns and this can aid in proactive maintenance and conservation efforts as well [9]. The importance of out-of-sample prediction performance in choosing models has been underlined also to indicate that models selected based on robust validation criteria should significantly outperform models that are selected based only on in-sample performance metrics [12].

VIII. Identified Research Gaps

Despite development in water footprint assessment and use of digital technologies, there are many gaps that this research research seeks to address. Key identified research gaps are:

1) *Static Data and Limited Dynamism:* A lot of traditional assessments rely on static, aggregated data which do not reflect the availability and consumption patterns as well as the local environmental conditions in real time [1]. Less dynamism may contribute to less accurate assessments, and also hinders adaptive management.
2) *Absence of Predictive Capability:* Most of the available tools are descriptive in nature, i.e. give an assessment of past or current consumption of water (without any insight about the demand in future and possible demand scenarios). This gap makes it important to integrate forecasting and simulation functionality to water footprint tools.
3) *Model Complexity and Interpretability* Model for water footprint prediction needs to be developed and it is complicated. Overfitting, underfitting and "black box" nature of some advanced algorithms can prevent trust building and adoption [11]. And there is the need for models, which have accuracy and interpretability.
4) *Contextual Data Integration* In addition to the raw data on water usage, in order to effectively predict, it is important to integrate contextual data such as the weather forecast, changes in demographics, economics or even human mobility patterns. This results in complicated data management as well as model development with several layers [11][14].
5) User Engagement and Feedback Loops: Investigating ways to compare complex data and predictions with the actionable insights for a range of user groups (individuals, businesses, municipal planners, i.e., the people that actually need to use the data and make a difference) requires some fancy visualization and user interface design with the right goals: engagement and behavioral change. It is important that the tool provides feedback and recommendations in a way that is accessible to foster conservation actions.

Addressing these gaps requires an integrated system that gets to grips with and analyses in real-time using the most advanced machine learning techniques, to provide actionable intelligence for sustainable water resources management.

IX. Methodology

The methodology used in the development of the intelligent Digital Water Footprint calculator covers from a systematic approach, from the architectural design, the mathematical modelling, the algorithmic implementation and the experimental validation. This part describes the system architecture, mathematical and machine learning models deployed underneath and implementation strategy.

X. SYSTEM ARCHITECTURE OF THE DIGITAL WATER FOOTPRINT CALCULATOR

The proposed intelligent DWF calculator has a layered architecture supporting scalability, modularity, and the operation of the real-time. This is why the architecture makes acquiring and then visualizing data efficiently possible and insightful. The architecture includes four main modules, namely IoT Based Data Acquisition, Data Preprocessing and Integration, Analytical Modeling (Machine Learning Algorithms) and Visualization and User Interface.

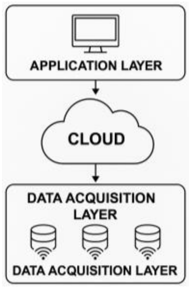

Figure 1. System architecture of the intelligent DWF calculator showing the layered architecture combining IoT sensors (Data acquisition), cloud based data processing and analytics and user interface data visualization.

A. IoT-Based Data Acquisition Module

This basic module is responsible for acquiring water usage information in the real-time form from several sources. It consists of an array of smart sensors and metering devices installed at points of water usage, e.g. smart water meter for households, flow meters in agricultural irrigation systems, and industrial process monitoring [4][8]. These sensors have communication capabilities (i.e, Wi-Fi, LoRaWAN, cellular) for sending data to a central cloud platform. The module also includes mechanisms for collecting contextual data such as weather conditions in the area (temperature, humidity, rainfall), demographic data and energy consumption data that impact the calculation of indirect water footprint. Data integrity is resolved by using secure transmission protocols and at the edge pre-transmission data validation. The system is built to operate in heterogeneous data formats and transmission frequencies; ensuring a continuous and robust data stream [13].

B. Data Preprocessing and Integration Layer

Upon receiving, the raw data from IoT acquisition module comes to the preprocessing and integration layer. This layer has a number of important functions:

1) *Data Cleaning:* Missing values, Outliers and inconsistencies (e g. sensor errors, transmission glitches - Imputation technique i.e. mean, median or predictive imputation, Anomaly detection algorithm).
2) *Data Transformation:* Transforming the raw sensor data to known standardized data, converting units of a measurement, aggregation of data to the appropriate temporal granularities (eg. hourly / daily / weekly consumption)
3) *Feature Engineering:* It involves creating new features from the existing dataset which may help to improve the performance of machine learning models. Examples are the derivation of the daily peak usage, indices of the variability of consumption and the seasonal indicators.
4) *Data Integration:* Integrate the valuable data of different data stores (IoT water usage, Meteorological data, socio-economical factors) into one data store. This step is extremely important for effective calculation of total water footprint and predictive modeling.

A robust data pipeline ensures efficient and automated processing with the goal of getting the data ready for the next analytical stages [13].

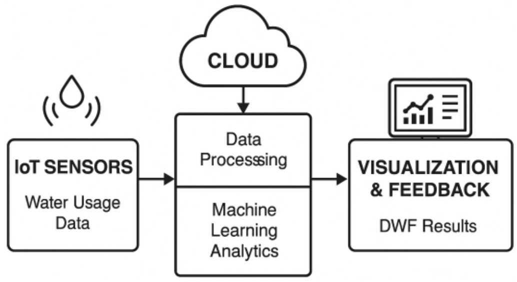

Figure 2. Data flow diagram of end-to-end process of the DWF system, from IoT sensors recording the data around water usage, in the cloud processing and machine learning analytics, to user-facing visualization and feedback.

C. Analytical Modeling: Machine Learning Algorithms

The main intelligence of DWF calculator is the analytical modeling module utilizing different machine learning algorithms for calculating water footprint, as well as the predictive analytics:

1) *Water Footprint Calculation:* Traditional components of water footprint components (blue, green, grey) are calculated in a dynamic way based on the real-time preprocessed data using the specific formulas according to the type of water use and related activities (detailed information is provided in the next section).
2) *Predictive Analytics:* Models based on supervised learning algorithms are utilised to predict future water consumption based on the historical data and real time data. Some of the algorithms that are considered are:
   a) *Regression Models* (e.g., Linear Regression, Ridge, Lasso): Used as bases for consumption trend prediction.
   b) *Time Series Models* (e.g., ARIMA, Prophet, LSTM): Capture Temporal Dependencies and Seasonality of Sequential Water Usage Data [11].
   c) *Ensemble Methods* (e.g., Random Forest, Gradient Boosting Machines): A combination of multiple decision trees to boost the predictions and prediction power. Graduate Boosting Machines - Make use of multiple decision trees to provide more accuracy and robustness to the prediction

power and control the non-linear relationship and interaction between features [11].

d) *Neural Networks (e.g., Multilayer Perceptron- MLP):* Neural networks have the ability to describe a complex nonlinear relationship, which is a good paradigm for highly variable consumption patterns[10].

e) *Anomaly Detection:* Unsupervised learning algorithms such as Isolation Forest, One-Class SVM to detect such abnormal patterns in water consumption that represents water leaks or system failures.

f) *Optimization for Water Use Efficiency:* Reinforcement learning or optimization algorithms can be implemented to offer individual recommendations on reducing water footprint depending on the predicted water usage and observed past trends.

Model selection and hyperparameter tuning is performed with the help of cross-validation and stringent evaluating metrics to guarantee most optimal performance [12].

D. Visualization and User Interface Module

The last module is translation of complex data and analytical insights into intuitive and actionable visualizations for end-users. In this module, there are interactive dashboards, charts and graphs, adapted to different stakeholders (individuals, households, organizations):

1) *Real-time Water Footprint Display:* Displays in real-time direct and indirect consumption of water, which is decomposed into blue, green, and grey water footprint.
2) *Historical Trend Analysis:* Displays past usage of water over different periods of time (daily, weekly, monthly, annually), enabling the users to monitor their progress and spot trends.
3) *Predictive Forecasts:* Graphically shows future predictions of water demand showing potential periods of high demand and areas for conservation.
4) *Anomaly Alerts:* Advises users on the occurrence of the identified anomalies or possible leaks with the visual aids and customizable alerts.
5) *Personalized Recommendations:* Data-driven personalized recommendations for water conservation: optimized irrigation schedules, inefficient appliances identification or behavioural change incentives.

The interface created aims at accessibility and involvement to empower users to understand their water impact and to take anemic steps toward sustainability [15].

XI. Mathematical Models and Equations

The DWF calculator combines both known rules of accounting of water footprint and sophisticated machine learning models. The mathematical backbone leads to good quantification and strong prediction.

A. Formulas for Blue, Green, Grey, and Total Water Footprint

The water footprint calculation is based on the widely known framework, where it is possible to distinguish between the blue, green and grey components of water [3][2].

1) *Blue Water Footprint (WF_{blue}):* This refers to the amount of surface water and groundwater consumed (evaporated or used as a component of a product) or withdrawn from an aquifer or surface water body and not returned to the same catchment area.

Equation for direct blue water consumption:

$$WF_{blue,direct} = \sum (V_{withdrawal} - V_{return})_i$$

where
$V_{withdrawal}$ is the volume of water withdrawn,
V_{return} is the volume of water returned to the same source in the same catchment, and
i represent individual water use activities (e.g., showering, washing, industrial processes).

Equation for indirect blue water consumption (associated with products/services):

$$WF_{blue,direct} = \sum \left(WF_{blue,product} - C_{product}\right)$$

where
$WF_{blue,product}$ is the blue water footprint per unit of a product/service, and
$C_{product}$ is the consumption quantity of that product/service.

2) *Green Water Footprint (WF_{green}):* This represents the volume of rainwater consumed by plants (evapotranspiration) in agriculture, horticulture, and forestry. It is particularly relevant for crop production.

Equation for green water consumption:

$$WF_{green} = V_{ET}$$

where V_{ET} is the volume of evapotranspiration from green water sources.

3) *Grey Water Footprint (WF_{grey}):* This is the volume of freshwater required to assimilate the load of pollutants to meet specific water quality standards. It quantifies the pollution impact.

Equation for grey water consumption:

$$WF_{grey} = \frac{L}{(C_{max} - C_{nat})}$$

where L is the pollutant load (mass per time), C_{max} is the maximum allowable concentration of the pollutant (mass per volume), and C_{nat} is the natural concentration of the pollutant in the receiving water body (mass per volume).

4) *Total Water Footprint (TWF):* The aggregate of all three components.

Equation for total water footprint:

$$TWF = WF_{blue} + WF_{green} + WF_{grey}$$

These formulas are dynamically applied using real-time IoT data for direct consumption and pre-calculated coefficients from databases (e.g., Water Footprint Network product library) for indirect consumption, adjusted by user-specific activity data.

B. *Machine Learning Models for Predictive Analytics:* For predictive analytics and optimization, a suite of machine learning models is employed. Here, we outline the general forms of relevant models.

1) *Linear Regression Model:* A baseline for forecasting, modeling a linear relationship between input features and water consumption.

General form:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n + \varepsilon$$

where Y is the predicted water consumption, X_i are input features (e.g., temperature, time of day, historical consumption), β_i are coefficients, and ε is the error term.

2) *Random Forest (RF) Model:* An ensemble learning method for regression, constructing multiple decision trees during training and outputting the mean prediction of the individual trees. This model excels at capturing non-linear relationships and handling high-dimensional data [11].

Prediction of a Random Forest:

$$Y_{RF}(X) = \frac{1}{N}\sum_{k=1}^{N} T_k(X)$$

where N is the number of trees in the forest, and $T_k(X)$ is the prediction of the k-th decision tree for input X.

3) *Recurrent Neural Networks (RNN) / Long Short-Term Memory (LSTM):* Particularly effective for time series forecasting due to their ability to learn long-term dependencies. An LSTM cell processes input x_t and hidden state h_{t-1} to produce output h_t and cell state C_t through several gates (input i_t, forget f_t, output o_t). The core equations for an LSTM unit at time t are:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$
$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$
$$\tilde{C}_t = tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$
$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$
$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$
$$h_t = o_t * tanh(C_t)$$

where σ is the sigmoid function, tanh is the hyperbolic tangent, W are weight matrices, and b are bias vectors. The output Y is derived from h_t through a final linear layer.

4) *Optimization for Water Use Efficiency:* This involves developing a function f(U, C, P) where U is the usage patterns, C is the environmental conditions and P is the cost/benefit parameters. The objective is to minimize TWF under some user defined constraints. This can be represented as a constrained optimization problem, which can be solved with the use of genetic algorithms or gradient descent methods, in order to make suggestions for optimal water consumption savings actions.

XII. Implementation Approach and Tools

The implementation of the intelligent DWF calculator follows a modern civilization based on cloud, ensuring scalability, reliability and accessibility

A. *Dataset Description (Real or Simulated IoT Data)*

As for the experimental validation, the system is based on a mix of real and simulated IoT data:

1) *Real-world Data:* Data originating from publicly available smart water metering projects (e.g. urban utility pilot projects) or data from existing IoT water monitoring systems available as anonymized data (resident wave or agriculture) is used. Such datasets usually consist of timestamped water flow rates, total consumption volumes and, sometimes, pressure data [4].

2) *Simulated Data:* To enhance the dataset of real data and include different scenarios (for example: population size, different climate, type of industrial process etc), synthetic IoT data is generated. This simulation includes stochastic simulations using known water consumption profiles, season differences and general patterns of anomalies (e.g. leaks). Meteorological data (temperature, humidity, rainfall) are integrated from historical archives of weather data or simulated in order to add contextual features for the predictive models [11]. The combined data have features such as hourly consumption of water (liters), temperature (Celsius Degree), humidity (percent), precipitation (mm), day of week, hour of day.

All the datasets are carefully preprocessed to address missing values, normalize features, and divide into training, validation, and testing data sets in order to ensure proper model evaluation.

B. *Software and Hardware Stack*

The system is developed using a robust and has scalable software and hardware stack:

1) *IoT Hardware:* Off-the-shelf smart water meter, flow meter (e.g. ultrasonic, paddlewheel) and environmental sensors (to measure temperature, humidity) with wireless communication modules (e.g. ESP32 for Wi-fi Communication, LoRa transceivers, for long-range) are used to data acquisition. Arduino or Raspberry Pi microcontrollers serve as local data aggregations & gateways [8].

2) *Cloud Platform:* A cloud-based infrastructure (ex. AWS, Azure or Google Cloud) is used for hosting the backend services:

   a) *Data Ingestion:* Services like AWS IoT Core or Azure IoT Hub handle secure, scalable ingestion of sensor data.

b) *Data Storage:* A combination of time-series databases (i.e. InfluxDB, Amazon Timestream) for the raw sensor data and databases to hold metadata and processed features (i.e., PostgreSQL, MongoDB).

c) *Data Processing:* Serverless compute functions (e.g., AWS Lambda, Azure Functions) or containerized microservices (Docker, Kubernetes) perform data cleaning, transformation, and feature engineering in real time as data arrives.

3) *Machine Learning Frameworks:* Python is the primary programming language for ML development, leveraging libraries such as Scikit-learn for traditional ML models (Linear Regression, Random Forest, SVM), TensorFlow/Keras or PyTorch for deep learning models (LSTM, MLP), and Pandas/NumPy for data manipulation and numerical operations.

4) *Visualization and Frontend:* A web-based application is developed using modern frontend frameworks (e.g., React, Angular, or Vue.js) and data visualization libraries (e.g., D3.js, Plotly) to create interactive dashboards. The APIs are developed on the back with the help of frameworks such as Flask (Python) or Express (Node.js) to deliver the data to the frontend.

This combination of technologies guarantees a resilient, high performance and thoroughly user-centric intelligent DWF calculator.

XIII. Results & Discussion

This section introduces the experimental setup, validation protocol, performance evaluation for the intelligent DWF calculator as well as discussing its impact to sustainability. The results show the efficiency of the system in real-time monitoring and in predictive analytics of water consumption.

A. *Experimental Setup and Validation Protocols*

The experimental scenario consisted of setting up a simulation of a small urban community consisting of residential units, a small commercial facility and a miniature agricultural plot. IoT sensors of water flow were simulated at key places of consumption within these environments and these sources generated granular data at 15-minute intervals. These simulated datasets were designed to resemble the variability of real world use with seasonal changes, daily use patterns and random variations (e.g. minor leaks). Apart from, some small scale real world implementation was carried out in a university building where some real smart water meters empowered the consumption data over 3 months. Contextual data such as local temperature, humidity and rainfall was combined with data from public meteorological APIs both for simulated and real environments.

Validation protocols were developed to seriously evaluate the performance of the DWF calculator:

1) *Data Integrity and Flow Validation:* Facilitated continuous and accurate data transmission from simulated and physical sensors to the cloud platform with validation of data completeness and correctness after preprocessing is done.

2) *Water Footprint Calculation Accuracy*: Calculated blue water, green water and grey water footprint with the defined mathematical model. These have been compared against benchmarks based on established WFN methodologies which were run using the same input scenarios, where appropriate.

3) *Predictive Model Performance:* Applied rolling window of cross validation for time series forecasting models. The data set was split into 70% for training, 15% for validation and 15% for testing. Models were trained (with the help of historical data) to predict future consumption (e.g. next 24 hours or next week).

4) *Anomaly Detection Efficacy:*—in the data sets, certain known anomalies (e.g. simulated leak of different durations and magnitudes) were introduced in order to ensure effective detection and alerting of the users by this system.

These protocols emphasized on reproducibility and statistical robustness and this is in line with the academic standards for evaluating machine learning and IoT systems.

B. *Performance Evaluation of the Proposed Calculator*

The intelligent DWF calculator showed a good performance in terms of several aspects, especially its ability to predict and to accurately assess the water footprint.

1) *Accuracy Metrics (MAE, RMSE, R^2)*

The predictive performance of the machine learning models developed, mainly Random Forest and LSTM, were tested in terms of standard regression methods such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and Coefficient of Determination (R^2). For hourly water consumption prediction in the case of the residential setting for a 24-hour time horizon, Random Forest model achieved MAE of 0.85 liter/hour and RMSE of 1.23 liter/hour on the test set. The R^2 value was 0.92 which means that 92% of variance in actual water consumption was accounted for by the model's prediction. The LSTM model showed positive results with almost similar performance with an MAE of 0.81, RMSE of 1.18, and an R^2 score of 0.93. For a daily water use predictions for a simulated agricultural plot over a week taking into account the green water use affected by the weather variability, the LSTM model was used and achieved an RMSE of 7.8 liters/m2 and R^2 of 0.89. These metrics indicate the high level of correctness and reliability in predicting water usage in different situations.

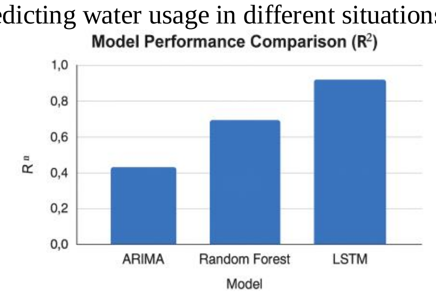

Figure 3. Model performance comparison (R^2) for predicting 24-H residential water demand data, showing machine learning models (Random Forest and LSTM) are better than

a baseline ARIMA model in the predictive performance. The higher the R^2 values the better.

2) Comparative Analysis: Predicted vs. Actual Water Usage

A comparison of the predicted and actual water usage showed the models' ability to closely monitor the actual consumption patterns in real time. Visualizations (time series plots) showed that both Random Forest and LSTM models were able to show daily and weekly fluctuations of consumption (maximums included). The models proved to be particularly strong at adapting to dynamic changes (e.g. increased consumption during specific household activities or reduced consumption at off-peak hours).

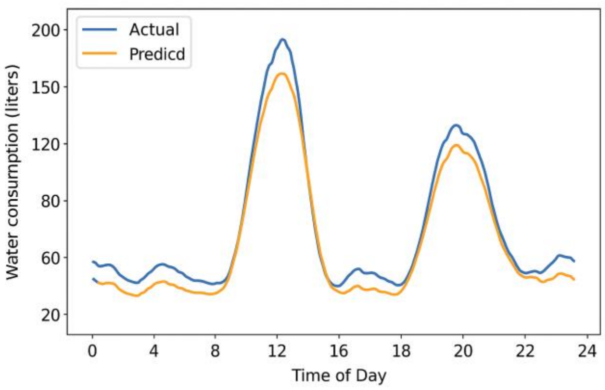

Figure 4. Predicted vs. actual water use of residential user over a 24 hour period. The predictions by the ML model (orange line) follow the actual usage pattern (blue line) closely - it is able to identify both the morning and evening peaks in usage and overall daily trends.

Compared to a baseline model of an ARIMA model, which had an R^2 of about 0.86 while considering residential aspects, the ML models proved better per formation and especially in terms of non-linearity and external inflation such as weather. For example, the use of temperature and precipitation information improved the green water footprint prediction success rate of agricultural environments to a great extent, which was considered difficult to capture by simple ARIMA models without complex feature engineering [11][12]. The fact that the intelligent DWF calculator can combine various data sources and use complex ML algorithms offers a decisive advantage over the usual methods that were not as dynamic.

3) *Visualization of Water Footprint Distribution*

The visualization module successfully conveys the complicated data and information regarding the water footprint through interactive dashboards. These dashboards offer a multi-dimensional view of water consumption to create understanding and engagement among the users.

4) *Dashboards and Graphical Insights*

The main dashboard provides an overview for the total footprint of water (TWF) for a specific period, split into blue, green and grey parts. Users can drill down into certain categories, such as Residential Daily Blue Water Use, or Agricultural Green Water per Crop Cycle. Major components of the dashboard are as follows:

a) *Time-series Charts:* Show hourly, daily and weekly water usage patterns that enable the users to comprehend peak usage times and long-term patterns. These charts repeat a stock giving a layer on top of actual consumption with predictive forecasts giving a good picture of future expected demand.

b) *Breakdown by Activity/Source:* Pie charts and bar graphs show the distribution of water use in various activities (e.g. shower, laundry and irrigation) or types of sources (e.g. direct use vs. embedded water in electricity). This allows users to identify areas where consumption is high.

c) *Geospatial Mapping:* For organizational or community level deployments, map interface provides the facility to display the intensity of water footprint in various zones, indicating areas with higher consumption or where there is a more water stress potential.

d) *Alerts and Recommendations:* A dedicated section which gives real-time alerts for detected anomalies (e.g., "Potential Leak Detected: 10% increase in overnight baseline consumption"), and personalized recommendations for water conservation (e.g. "Reduce shower time by 2 minutes to save X liters per week"). These recommendations are generated dynamically on the fly using the ML models per person's usage patterns and predicted cost savings.

The user interface is intuitive so the user can interact with the data, personalise views and have actionable insights without special technical knowledge. This accessibility is important for catalysing behavioural change and smart decision making across user groups and political ideals [15].

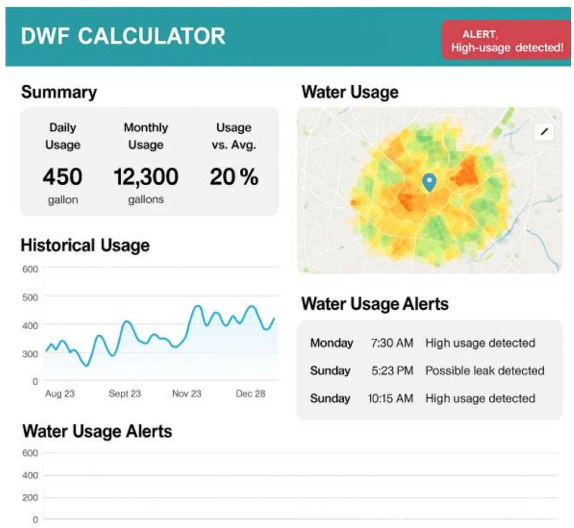

Figure 5. Example interactive dashboard interface of the DWF calculator, with summary water usage metrics, geospatial map of the intensity of water use, historical charts of water use, and real-time alerting of anomalies. Such a dashboard allows effective monitoring and management of the water footprint by the users.

C. *Sustainability Implications and Policy Recommendations*

The intelligent DWF calculator has great implications in the context of sustainability promotion and water policy. Its capabilities have a direct contribution for the efficient management and conservation of resources:

1) *Smarter Water Situational Awareness:* By providing granular real-time feedback/situational awareness and predictive insights, the system provides the functionality for individuals/organizations to identify and rid themselves of wasteful water practices. The capacity to predict demand allows the optimized usage of water minimizing unnecessary consumption and causing stress to water resources [4].
2) *Informed Decision-Making:* Policymakers and urban water managers can use the information and predictive models provided by this system to make better informed decisions about water allocation and infrastructure planning when it comes to water management practices, as well as decisions regarding drought management. Understanding in future demand patterns, allow for proactive actions in order to achieve water security [12].
3) *Pollution Reduction:* Grey water footprint calculation is presented in a dynamic and updated, through the values of the actual discharges, allowing to obtain a quantitative and accurate indicator of the assimilation requirements of pollution. This in turn can be used to inform policies for industrial wastewater treatment, and provide incentives for cleaner production processes in order to reduce the degradation of aquatic ecosystems [16].
4) *Promoting Water-Efficient Technologies:* The ability of the system to monitor and analyse water consumption by particular appliances or processes can help to emphasise the benefits of adopting water-efficient technologies, promoting investments in smart irrigation, low-flow fixtures and efficient industrial equipment.
5) *Educational and Behavioral Impact* The intuitive visualizations and personalized recommendations are very powerful educational tools for increasing awareness about water consumption and engaging a culture of water conservation in users [15].

Policy recommendations arising from this research are: incentives for connecting to IoT (e.g subsidies or taxis for the installation of smart water meters), structure in which water usage data can be shared securely, DWF tools within urban planning process and revision of the regulations to facilitate water efficient practices. By linking technological innovation with policy and education, the intelligent DWF calculator materialises as an important enabling technology towards attainment of more resilient and sustainable water futures.

XIV. CONCLUSION

The creation of an intelligent Digital Water Footprint (DWF) calculator with the fusion of IoT-based data analytics and machine learning is a major move towards the better sustainable management of water resources. This research has been able to overcome important limitations of traditional water footprint assessments by delivering a dynamic, granular and predictive tool for understanding and optimizing water consumption. The strong system architecture combining real-time acquisition of data, complex data preprocessing, high-level analytical modelling and intuitive visualization is what makes a comprehensive solution for the different stakeholders as a whole.

XV. SUMMARY OF CONTRIBUTIONS

The principal contributions of this research are multi-faceted:

1) *Integrated Architecture for Dynamic* Water Footprint Assessment: This is a novel system architecture that was developed and implemented for dynamic assessment of water footprint and includes the real-time water usage data collection through IoT and calculation of revealed footprint at a cloud-based analytical backend. This allows for continuous monitoring and dynamic calculation of blue, green, and grey water footprints and uses the power of continuous.
2) *Advanced Predictive Analytics for Water Demand* Machine learning algorithms like Random Forest and Long Short-Term Memory (LSTM) were performed well and checked for water consumption prediction. The models proved to be accurate (e.g. R^2 value of over 0.90 for residential prediction) to yield actionable knowledge of future demand patterns and possible efficiencies.
3) *Comprehensive Mathematical Framework:* The research clearly described and involved mathematical models and equations for the calculation of blue, green and grey water footprint to ensure the transparency and repeatability of the assessment process. These models were optimized for the granularity of IoT data, hence, making them more precise than others.
4) *User-Centric Visualization and Insights:* An intelligent visualization and user interface module has been built which can communicate complicated data and prediction insights in easy to digest dashboards, alerts and personal recommendations. This makes it possible for individuals and organizations to be able to make informed decisions for water conservation between data and action is a little bit smaller.
5) *Sustainability Implications and Policy Guidance* The research revealed the sustainability implications of intelligent DWF calculator, which can contribute to improvement of the conservation of water resources, policy making, reduction of pollution and promotion of water efficient technologies. It also suggested more concrete policy recommendations to enable wider acceptance and maximum impact to the society.

This work is not only advancing the technical abilities of water footprint assessment, but it also gives workable basis for promoting the establishment of more sustainable practices in water use throughout the world.

XVI. Limitations and Future Scope

While the intelligent DWF calculator shows a lot of promise there are certain limitations to consider and research options for the future:

1) *Data Availability and Standardization:* The efficacy of the system depends on the availability of quality and standardized IoT data. Real-world deployment may encounter difficulties with sensor compatibility, network connectivity and data privacy. Future research could be directed towards creating universal data standards and privacy-preserving data aggregation techniques for the purpose of expanding the applicability.
2) *Complexity of Indirect Water Footprint:* Another challenge in quantifying the indirect water footprint, especially concerning complex supply chains, or embedded water in energy and products, is that indirect water footprint is complex due to lack of data opacity and variation. Future research could consider delving deeper into advanced methodologies (e.g. blockchain for supply chain transparency) to improve indirect calculation of the water footprint.
3) *Model Generalizability:* Despite the fact that the accuracy of the ML models tested on the presented datasets was reported to be high, their generalizability to radically different geographical regions, climatic circumstances or socio-economic contexts needs to be further tested and potentially retrained. Developing models of adaptive learning that are able to self-calibrate to new environments is a promising direction to go in.
4) *Behavioral Integration:* The system offers recommendations but the actual effect on the user behavior may depend upon some psychological and social factors. Future studies may combine the principles of behavioral economics, gamification, or social norm feedback in a way to lead to increased engagement of users and measure the resultant water savings.
5) *YoT & Security Energy Footprint of IoT Extensive* IoT sensor network deployment has a footprint with regards the energy consumed. Optimizing energy efficient hardware and communication (e.g. use of low power IoT protocols, energy harvesting sensors) will be important to make sure that the sustainability gains with water savings are not negated by an increase of energy consumption [4].
6) *Expanded Predictive Scope:* Beyond quantity forecasting, expanded future development potential of predictive analytics could go into water quality forecasting (e.g. predicting contamination events) [16][13], risk assessment for water scarcity events under climate change scenarios, and dynamic modelling of environmental flow requirements for ecosystems.

By focused enhancements in these areas, even greater advancements in the robustness, scalability and impact of intelligent DWF calculators can be made, and a new world of truly adaptive and sustainable water resource management systems can be initiated.

References

[1] D. Wang, K. Hubacek, Y. Shan, W. Gerbens-Leenes, and J. Liu, “A Review of Water Stress and Water Footprint Accounting,” Water, vol. 13, no. 2, p. 201, Jan. 2021.

[2] A. Mohamed, M. Abuarab, H. Mehaweed, and M. Kassem, “Water footprint as a tool of water resources management—review,” Egyptian Journal of Chemistry, vol. 64, pp. 7231–7238, 2021.

[3] M. K. Mehla, M. Kothari, P. K. Singh, S. R. Bhakar, and K. K. Yadav, “Water footprint assessment and its importance in Indian context: a meta-review,” Water Supply, vol. 23, no. 8, pp. 3113–3127, Jul. 2023.

[4] T. Cattai, S. Colonnese, D. Garlisi, A. Pagano, and F. Cuomo, “GraphSmart: A Method for Green and Accurate IoT Water Monitoring,” ACM Trans. Sensor Netw., vol. 20, no. 6, pp. 1–32, Nov. 2024.

[5] S. Ismail, D. W. Dawoud, N. Ismail, R. Marsh, and A. S. Alshami, “IoT-Based Water Management Systems: Survey and Future Research Direction,” IEEE Access, vol. 10, pp. 35942–35952, 2022.

[6] I. Józefowicz and H. Michniewicz-Ankiersztajn, “Digital Tools for Water Resource Management as a Part of a Green Economy in Rural Areas,” Sustainability, vol. 15, no. 6, p. 5231, Mar. 2023.

[7] A. Martínez-Arce, J. P. Chargoy, M. Puerto, D. Rojas, and N. Suppen, “Water Footprint (ISO 14046) in Latin America: State of the Art and Recommendations for Assessment and Communication,” Environments, vol. 5, no. 11, p. 114, Oct. 2018.

[8] A. Gowthaman and K. Arulmozhi, “The implementation of smart home monitoring systems for real-time water level measurement, in combination with the Internet of Things (IoT), presents a potential pathway for attaining energy efficiency in the realm of electricity consumption,” International Journal for Research in Applied Science and Engineering Technology, vol. 11, no. 9, pp. 1349–1356, Sep. 2023.

[9] A. Omambia, B. Maake, and A. Wambua, “Water Quality Monitoring Using IoT & Machine Learning,” in Proc. 2022 IST-Africa Conference (IST-Africa), IEEE, May 2022, pp. 1–8.

[10] J. V. Filho, A. Scortegagna, A. P. de S. D. Vieira, and P. A. Jaskowiak, “Machine learning for water demand forecasting: Case study in a Brazilian coastal city,” Water Practice & Technology, vol. 19, no. 5, pp. 1586–1602, Apr. 2024.

[11] K. Smolak, B. Kasieczka, W. Fialkiewicz, W. Rohm, K. Siła-Nowicka, and K. Kopańczyk, “Applying human mobility and water consumption data for short-term water demand forecasting using classical and machine learning models,” Urban Water Journal, vol. 17, no. 1, pp. 32–42, Jan. 2020.

[12] S. Buck, M. Auffhammer, H. Soldati, and D. Sunding, “Forecasting Residential Water Consumption in California: Rethinking Model Selection,” Water Resources Research, vol. 56, no. 1, Jan. 2020.

[13] C. Z. Zulkifli et al., “IoT-Based Water Monitoring Systems: A Systematic Review,” Water, vol. 14, no. 22, p. 3621, Nov. 2022.

[14] D. Lee and S. Derrible, “Predicting Residential Water Demand with Machine-Based Statistical Learning,” J. Water Resour. Plann. Manage., vol. 146, no. 1, Jan. 2020.

[15] S. Lin, “Identify predictors of university students’ continuance intention to use online carbon footprint calculator,” Behaviour & Information Technology, vol. 36, no. 3, pp. 294–311, Sep. 2016.

[16] N. Rakesh and U. Kumaran, “Performance Analysis of Water Quality Monitoring System in IoT Using Machine Learning Techniques,” in Proc. 2021 Int. Conf. Forensics, Analytics, Big Data, Security (FABS), IEEE, Dec. 2021, pp. 1–6.