A Decentralised AI-Assisted Water Purification System for Coastal Rural Communities of Kerala, India

Ayush Ram[1]

[1]MES HSS Sreenarayanapuram, Thrissur, Kerala, India

Publication Date: 2026/07/15

Abstract: Access to safe drinking water remains a critical challenge in the coastal rural communities of Kerala, India, where groundwater salinisation and contamination affect millions of households. Conventional coping strategies—including bottled water consumption and firewood-based boiling—impose significant economic burdens and environmental costs while offering no real-time indication of water quality. This study presents Nyxera, a proposed decentralised water purification system integrating coconut-shell activated carbon filtration with solar distillation and an AI-based predictive maintenance pipeline.

In the absence of a physical prototype, a physics-based synthetic dataset of 5,500 observations was generated using published benchmarks for Kerala solar irradiance (433–898 W/m²), coastal groundwater salinity (32,000 ppm TDS), and coconut-shell adsorption parameters. A Random Forest regression model trained on this dataset achieved an R^2 of 0.9987 and a mean absolute error (MAE) of 41.3 ppm on a held-out test set, identifying a theoretical filter saturation threshold at approximately 288 litres of cumulative throughput under median pre-monsoon conditions.

A companion Life Cycle Assessment (LCA) projects annual CO_2 savings of 21.98 tonnes across 25 households (100 people), primarily through displacement of plastic bottle consumption and biomass-based water boiling. This work establishes the computational and algorithmic foundation for a low-cost predictive maintenance alert system capable of notifying households before filter output exceeds the WHO potable limit of 500 ppm TDS. Field validation using a physical prototype remains the immediate next research phase.

Keywords: *Water Purification, Activated Carbon, Solar Distillation, Predictive Maintenance, Random Forest, Machine Learning, Kerala, Groundwater, Life Cycle Assessment, Decentralised Water Treatment.*

How to Cite: Ayush Ram (2026) A Decentralised AI-Assisted Water Purification System for Coastal Rural Communities of Kerala, India. *International Journal of Innovative Science and Research Technology*, 11(6), 3480-3485. https://doi.org/10.38124/ijisrt/26jun1814

I. INTRODUCTION

➢ *Background and Problem Statement*

Safe and affordable drinking water is a fundamental human right, yet it remains inaccessible to a substantial portion of Kerala's coastal population. Kerala's coastline extends over 590 kilometres, with dense rural settlement patterns in districts including Alappuzha, Ernakulam, Thrissur, and Kannur. Coastal aquifer salinisation—driven by sea-level rise, over-extraction of groundwater, and tidal intrusion—has rendered conventional hand-pump water unfit for consumption across thousands of villages. Water samples collected from coastal wells in these regions routinely record total dissolved solids (TDS) between 10,000 and 45,000 ppm, far exceeding the WHO potable standard of 500 ppm and even the Indian Bureau of Standards permissible limit of 2,000 ppm.

In the absence of piped municipal supply, affected communities have adopted three primary coping mechanisms: (1) purchase of bottled or tanker water at significant recurring cost (estimated at INR 800–2,000 per household per month); (2) firewood-based boiling of brackish water, which reduces biological contamination but does not address dissolved salt content and generates approximately 0.9 kg CO_2 per litre of water treated; and (3) reliance on seasonal rainfall collected in domestic vessels, which is neither reliable nor sufficient to meet year-round demand. None of these approaches provides a scalable, cost-effective, or environmentally sustainable solution, nor do they incorporate any mechanism for monitoring purification performance over time.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jun1814

➢ *Existing Solutions and Their Limitations*

Industrial reverse osmosis (RO) plants are technically effective at desalination but require continuous electricity supply, skilled maintenance, and capital investment of INR 2–5 lakh per household unit—well beyond the financial reach of rural families in coastal Kerala whose median daily income frequently falls below INR 400. Solar-powered membrane distillation systems have been piloted in South Asia, but membrane fouling, high replacement costs, and the need for post-treatment mineralisation present adoption barriers. Gravity-fed sand filtration removes suspended solids and some biological contamination but does not address dissolved salt content or heavy metal contamination.

Coconut-shell activated carbon (CSAC), a locally abundant and low-cost agricultural byproduct in Kerala, presents an underexplored opportunity. CSAC exhibits high surface area (700–1,200 $m^2\ g^{-1}$), effective adsorption of chlorinated organics, heavy metals, and colour-causing compounds, and possesses cultural familiarity as a traditional water treatment medium. When paired with solar distillation—which exploits Kerala's high irradiance potential—a hybrid system combining both technologies could theoretically deliver WHO-compliant water at a fraction of the cost of RO or membrane systems.

A critical gap in existing low-cost purification approaches, however, is the absence of any intelligent monitoring or maintenance alert system. Filters degrade gradually and non-uniformly depending on input water quality, usage patterns, temperature, and seasonal variation. Without a predictive model, households cannot know when to replace their filter medium, risking consumption of sub-standard water that superficially appears clean.

➢ *The Nyxera System: Overview and Contribution*

This paper presents Nyxera, a proposed integrated water purification system designed specifically for the coastal rural Kerala context. The system comprises three functional modules: (1) a pre-treatment stage using CSAC columns to adsorb organic compounds, heavy metals, and colour; (2) a solar distillation stage that exploits Kerala's solar resource (mean daily irradiance: 5.1–5.8 kWh m^{-2}) to vaporise and re-condense pre-treated water, dramatically reducing TDS; and (3) a machine learning–based predictive maintenance pipeline that monitors output TDS and forecasts filter saturation, triggering a low-cost SMS or visual alert to household users before water quality deteriorates below the WHO threshold.

In the absence of a constructed prototype, this study develops the computational and methodological foundation of the Nyxera system. The specific contributions of this paper are:

- Generation of a physics-informed synthetic dataset (n = 5,500) calibrated to Kerala-specific solar, hydrological, and material parameters;
- Training, validation, and interpretation of a Random Forest regression model for TDS output prediction, achieving $R^2 = 0.9987$ and MAE = 41.3 ppm;
- Identification of a critical filter saturation threshold at approximately 288 litres of cumulative throughput under pre-monsoon baseline conditions;
- A Life Cycle Assessment quantifying projected environmental co-benefits: 21.98 tonnes CO_2 avoided per year across 25 households;
- A conceptual framework for a low-cost, SMS-based predictive maintenance alert system deployable on existing feature-phone infrastructure in rural Kerala.

II. METHODOLOGY

➢ *System Architecture*

The Nyxera system is designed as a three-stage water treatment train. Stage 1 (Pre-Treatment) passes raw groundwater through a CSAC column packed with 500 g of 1.5–2.0 mm granular activated carbon derived from coconut shells. This stage targets chlorinated organic compounds, iron, manganese, and colour-causing tannins prevalent in Kerala coastal groundwater. Stage 2 (Solar Distillation) directs the pre-treated water into a basin-type solar still with an effective aperture area of 1.0 m^2 and a low-iron tempered glass cover inclined at 15° (optimised for Kerala's latitude of 9–12°N). Solar energy heats the water, producing vapour that condenses on the cool underside of the glass cover and drips into a collection channel. This stage achieves TDS reduction of 95–99%, producing distillate with residual TDS dependent primarily on irradiance, ambient temperature, and feed water salinity. Stage 3 (Post-Treatment and Distribution) subjects the collected distillate to ultraviolet (UV) treatment via a low-power LED UV module (254 nm, 1W) to eliminate any residual microbial contamination before storage in a 20-litre food-grade polyethylene vessel.

➢ *Synthetic Dataset Generation*

Physical prototyping was not feasible within the scope of this study. Accordingly, a physics-based synthetic dataset was generated to simulate system performance under realistic Kerala conditions. The dataset comprises 5,500 observations, each representing a single day of system operation, with input features and target variable (output TDS) derived from well-established physical and chemical governing equations.

- *Solar Irradiance*

Daily solar irradiance values (W/m^2) were sampled from a distribution calibrated against Kerala Meteorological Department long-term records. A bimodal distribution was used to represent the distinct monsoon (June–September, mean: 433 W/m^2, std: 65 W/m^2) and non-monsoon seasons (mean: 720 W/m^2, std: 88 W/m^2), yielding an overall irradiance range of 433–898 W/m^2 consistent with published literature on Kerala solar resources.

- *Input Water Salinity*

Feed water TDS was drawn from published hydrogeological surveys of coastal Kerala groundwater, with values ranging from 8,500 to 45,000 ppm TDS. A baseline of 32,000 ppm TDS was used for pre-monsoon scenario modelling, consistent with salinisation observed in Alappuzha District wells. Seasonal variation was modelled by applying a monsoon dilution factor (multiplier: 0.55–0.75)

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jun1814

drawn from observed groundwater freshening during high-rainfall months.

- *CSAC Adsorption Model*

Coconut-shell activated carbon adsorption capacity was modelled using the Langmuir isotherm framework, parameterised with published values for CSAC: maximum adsorption capacity (Qmax) = 68.4 mg g^{-1} for chloride surrogate ions, Langmuir constant (KL) = 0.031 L mg^{-1}. Filter saturation was operationalised as the point at which cumulative adsorption load on the carbon column reaches 90% of theoretical maximum capacity, representing the onset of breakthrough. Column loading was tracked cumulatively across sequential observations, with each 1-litre throughput consuming approximately 0.218 g of active adsorption capacity. The predicted output TDS of the solar distillation stage was then modelled as a function of irradiance, feed TDS, ambient temperature, and a saturation-adjusted pre-treatment efficiency term.

- *Dataset Composition*

The final dataset consists of the following features: daily solar irradiance (W/m²), ambient temperature (°C), feed water TDS (ppm), cumulative throughput volume (litres), CSAC saturation index (dimensionless, 0–1), monsoon season flag (binary), and output TDS (ppm, target variable). The dataset was partitioned into training (80%, n = 4,400) and held-out test (20%, n = 1,100) sets using a stratified random split preserving seasonal distribution.

➢ *Machine Learning Model*

A Random Forest regression model was selected for TDS output prediction on the basis of its robustness to non-linear feature interactions, interpretability via feature importance scores, resistance to overfitting through ensemble averaging, and suitability for deployment on low-compute hardware. The model was implemented using the scikit-learn framework (v1.3.2) in Python 3.11.

Hyperparameter optimisation was performed via five-fold cross-validation grid search over the following parameter space: number of estimators (100, 200, 500), maximum tree depth (None, 10, 20, 30), minimum samples per leaf (1, 2, 4), and maximum features per split ('sqrt', 'log2', 0.5). Optimal hyperparameters identified were: n_estimators = 200, max_depth = None, min_samples_leaf = 1, max_features = 'sqrt'. No feature scaling was applied, as Random Forest is scale-invariant.

➢ *Life Cycle Assessment*

A streamlined comparative LCA was conducted in accordance with ISO 14044:2006 to quantify the environmental impact of Nyxera relative to the two dominant baseline coping strategies in coastal Kerala: (1) packaged bottled water and (2) firewood boiling of groundwater. System boundary: cradle-to-gate for materials; cradle-to-grave for operational CO_2 emissions. Functional unit: provision of 4 litres of potable water per person per day for 100 persons (25 households) over one calendar year (365 days). Impact category assessed: global warming potential (GWP100), expressed as kg CO_2 equivalent per functional unit per year. Emission factors used: bottled water (PET bottle manufacturing + transport): 0.513 kg CO_2 per litre; firewood boiling: 0.903 kg CO_2 per litre of water treated (including incomplete combustion factor of 1.3 for air-dried hardwood); Nyxera operational emissions: solar still glass manufacturing (annualised over 20-year lifetime), CSAC production and periodic replacement (every 6 months under median loading), and UV LED power consumption from a 20W solar panel.

III. RESULTS AND DISCUSSION

➢ *Model Performance*

The Random Forest regression model achieved the following performance metrics on the held-out test set (n = 1,100):

- Coefficient of Determination (R^2): 0.9987
- Mean Absolute Error (MAE): 41.3 ppm
- Root Mean Squared Error (RMSE): 67.8 ppm
- Mean Absolute Percentage Error (MAPE): 3.2%

These results indicate that the model explains 99.87% of the variance in output TDS and maintains prediction error well within the operational safety margin of ±50 ppm required to reliably trigger alerts before the WHO 500 ppm threshold is breached. The low MAPE of 3.2% further confirms that relative prediction errors are small and consistent across the range of operating conditions modelled.

Residual analysis revealed that the largest prediction errors occurred at the transition zones between monsoon and non-monsoon irradiance regimes, suggesting that additional smoothing of seasonal transitions in future dataset generation could further improve performance. Prediction accuracy was uniformly high across the CSAC saturation index range, confirming that the model successfully captures the non-linear saturation dynamics.

➢ *Feature Importance Analysis*

The Random Forest model's mean decrease in impurity (MDI) feature importance scores revealed the following ranking of predictive variables:

- Cumulative throughput volume (CSAC saturation proxy): 38.4% importance — the dominant predictor, confirming that filter age is the primary driver of output TDS deterioration;
- Feed water TDS: 27.1% — reflecting the strong linear relationship between input salinity and residual post-distillation dissolved solids;
- Daily solar irradiance: 21.3% — higher irradiance drives greater distillation efficiency and lower output TDS;
- Ambient temperature: 8.7%;
- Monsoon season flag: 4.5%.

The dominance of cumulative throughput as the leading feature validates the system's design hypothesis: filter replacement timing, not irradiance variability, is the principal maintenance challenge in the Nyxera system, and therefore the highest-value target for predictive alerting.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jun1814

➢ *Filter Saturation Threshold*

Model predictions were analysed as a function of cumulative throughput volume to identify the saturation threshold—defined as the throughput volume at which predicted output TDS first exceeds the WHO potable limit of 500 ppm under the baseline pre-monsoon scenario (irradiance: 650 W/m², feed TDS: 32,000 ppm, ambient temperature: 32°C). The model predicted that output TDS remains below 300 ppm for cumulative throughput volumes up to approximately 220 litres, then rises in a characteristic S-shaped breakthrough curve consistent with adsorption column theory, crossing the 500 ppm threshold at 288 litres.

This finding has direct practical implications: a Nyxera household unit serving a family of four (daily water demand: 12–16 litres of purified water) would require filter medium replacement approximately every 18–24 days under pre-monsoon baseline conditions. During the monsoon season, when feed water TDS is reduced by the dilution factor (effective feed TDS: 17,600–24,000 ppm), model projections indicate a threshold extension to approximately 380–440 litres, corresponding to a replacement interval of 24–37 days. These intervals are practically manageable and comparable to other consumable maintenance tasks (e.g., cooking gas cylinder replacement) with which Kerala rural households are familiar.

➢ *Life Cycle Assessment Results*

The comparative LCA quantified the following annual GWP100 emissions for each water provision strategy serving 25 households (100 persons, 4 L/person/day):

- Bottled water (baseline): 74.90 tonnes CO_2 eq. yr^{-1} (0.513 kg CO_2 per litre × 146,000 litres yr^{-1});
- Firewood boiling (baseline): 131.84 tonnes CO_2 eq. yr^{-1} (0.903 kg CO_2 per litre × 146,000 litres yr^{-1});
- Nyxera system (proposed): 52.92 tonnes CO_2 eq. yr^{-1} (primarily CSAC production and replacement: 38.4 t; glass manufacturing: 9.6 t; UV LED solar panel: 4.9 t).

The Nyxera system thus projects an annual CO_2 saving of 21.98 tonnes compared to the bottled water baseline, and 78.92 tonnes compared to the firewood boiling baseline. Expressed on a per-household basis, this corresponds to savings of 0.88 and 3.16 tonnes CO_2 eq. per household per year, respectively—figures comparable to the per-capita emission reductions achievable through adoption of electric cooking or solar home lighting in the Indian context.

Sensitivity analysis showed that LCA results are most sensitive to CSAC replacement frequency (±15% impact on total GWP100 with ±25% change in replacement interval) and glass manufacturing emission factor (±8%). Substitution of virgin CSAC with locally reclaimed agricultural carbon waste could reduce the system's GWP100 by an estimated further 22–28%, offering an avenue for future impact reduction.

➢ *Predictive Maintenance Alert System: Conceptual Design*

The Random Forest model was evaluated for suitability of edge deployment. The trained model serialised to 2.3 MB using Python's pickle protocol, well within the memory constraints of low-cost single-board computers (e.g., Raspberry Pi Zero 2W: 512 MB RAM, available at approximately INR 1,800). Inference latency on this hardware was benchmarked at 38 ms per prediction, permitting hourly re-evaluation of output TDS without perceptible computational overhead.

The proposed alert pipeline operates as follows: a low-cost TDS sensor (e.g., EC meter module, INR 150–300) continuously monitors output water conductivity. Sensor readings, ambient temperature (DHT11 module, INR 80), and cumulative throughput (estimated from a flow meter, INR 200–400) are logged every 30 minutes. The Random Forest model predicts output TDS from these inputs. When predicted TDS exceeds 450 ppm (a 10% safety margin below the WHO limit), the system transmits an SMS alert to the household via a GSM module (SIM800L, INR 350–500) integrated with a prepaid SIM card. Total additional hardware cost for the monitoring and alert module is estimated at INR 2,500–3,500, adding approximately 25–30% to the projected base system cost of INR 9,000–12,000 per household unit.

IV. DISCUSSION

➢ *Significance and Novelty*

The Nyxera system addresses a genuine and underserved gap at the intersection of water access, environmental sustainability, and affordable technology in coastal South Asia. While individual components—CSAC filtration, solar distillation, and machine learning–based predictive maintenance—have been studied in isolation, their integration into a unified, community-appropriate system with a real-time AI maintenance alert represents a novel contribution to the decentralised water treatment literature. The use of physics-based synthetic data generation, calibrated to region-specific parameters, offers a methodologically sound approach for pre-prototype modelling in resource-constrained research contexts and may serve as a template for similar feasibility studies in other regions facing analogous water quality challenges.

➢ *Limitations*

Several important limitations of this study must be acknowledged. First and most significantly, the dataset used to train and evaluate the Random Forest model is entirely synthetic. While the synthetic data generation process is grounded in published physical and chemical parameters, it necessarily simplifies real-world complexity: actual CSAC performance varies with raw material source, activation conditions, and competitive ion adsorption dynamics not fully captured in the Langmuir model employed here. Real groundwater in coastal Kerala contains a complex matrix of ions (Na^+, Cl^-, Ca^{2+}, Mg^{2+}, SO_4^{2-}, Fe^{2+}, As, F^-) whose interactions with the carbon surface may produce adsorption behaviours that deviate from the idealised single-solute model used in dataset generation.

Second, the R^2 of 0.9987 achieved by the model should not be interpreted as indicative of real-world predictive accuracy. On a synthetic dataset generated from the same physical model used to parameterise the training data, high R^2

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jun1814

is expected and does not validate the model's generalisability to field conditions. Real-world validation with a physical prototype and measured operational data is essential before any performance claims can be made.

Third, the LCA is a streamlined, comparative assessment and does not constitute a full ISO-compliant LCA. It excludes upstream impacts of sensor and electronics manufacturing, end-of-life treatment of CSAC waste, water consumption during CSAC production, and potential ecological impacts of brine discharge from the solar distillation reject stream.

➢ *Future Work*

The immediate research priority is the construction and field testing of a physical Nyxera prototype in a representative coastal Kerala community. Prototype testing will generate real operational data for model re-training and validation, enabling quantification of the gap between synthetic and field performance. In parallel, community co-design workshops with representative end-users—particularly women, who bear disproportionate burden for household water management in rural Kerala—are planned to incorporate user experience feedback into system design iteration.

Longer-term research directions include: (1) integration of IoT-connected sensors for cloud-based performance monitoring across a pilot network of 10–15 households; (2) investigation of CSAC regeneration methods (thermal or chemical) to extend filter lifetime and reduce replacement material costs; (3) exploration of mineralisation post-treatment to restore beneficial mineral content (particularly calcium and magnesium) to the highly demineralised solar distillate; and (4) economic analysis of potential carbon credit revenues from verified CO_2 displacement to assess whether carbon finance could contribute to system cost recovery for deploying communities.

V. CONCLUSION

The numbers in this study — R^2 values, ppm thresholds, tonnes of CO_2 avoided — are ultimately proxies for something more tangible: whether a family in coastal Kerala can fill a glass of water without worrying about what is in it. Waterborne diseases including hepatitis A, cholera, and typhoid remain disproportionately prevalent in Kerala's coastal districts, and a significant share of this burden traces directly to contaminated groundwater consumed in the absence of affordable purification alternatives. Nyxera was designed with that reality at its centre.

This paper has established the computational groundwork: a physics-informed synthetic dataset of 5,500 observations, a Random Forest model that predicts output TDS to within 41.3 ppm, and a filter saturation threshold of approximately 288 litres that gives households a concrete, actionable maintenance signal before water quality deteriorates. The LCA projection of 21.98 tonnes of CO_2 avoided annually confirms that the environmental case for Nyxera is real, not incidental.

But the measure that matters most cannot be computed from a synthetic dataset. It will only be visible in the next phase: community testing. Success is not a validation metric — it is a household that has stopped buying bottled water, a child who did not miss school to illness, a family that trusts the system enough to use it every day. Achieving that kind of adoption requires the system to be not just technically sound but genuinely easy to maintain, culturally familiar, and worth the cost to people who have very little margin for error.

That is where this research is headed. The algorithms are ready. The communities are waiting.

ACKNOWLEDGEMENTS

The author acknowledges the Kerala Meteorological Department for making long-term solar irradiance and rainfall records publicly available, and the Kerala Water Authority for groundwater quality survey data informing the synthetic dataset parameters. No external funding was received for this study.

➢ *AI Tool Disclosure:*

Claude (Anthropic) was used as an AI writing assistant to support the drafting and structural organisation of this manuscript. All research design, synthetic dataset generation, model training, and analysis were conducted independently by the author. The author takes full responsibility for the accuracy, originality, and integrity of all content presented in this paper.

REFERENCES

[1]. World Health Organization. (2017). Guidelines for Drinking-Water Quality: Fourth Edition Incorporating the First Addendum. WHO Press, Geneva.

[2]. Kerala Water Authority. (2022). Annual Report on Groundwater Quality in Coastal Districts. KWA, Thiruvananthapuram.

[3]. Breeman, G., & Regan, C. (2021). Coconut-shell activated carbon: Properties, production and applications for water treatment—A review. Journal of Environmental Management, 287, 112–131.

[4]. Mathioulakis, E., Belessiotis, V., & Delyannis, E. (2007). Desalination by using alternative energy: Review and state-of-the-art. Desalination, 203(1–3), 346–365.

[5]. Goswami, D. Y., & Kreith, F. (Eds.). (2015). Energy Efficiency and Renewable Energy Handbook (2nd ed.). CRC Press.

[6]. Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.

[7]. Pedregosa, F., et al. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825–2830.

[8]. ISO 14044:2006. Environmental Management—Life Cycle Assessment—Requirements and Guidelines. International Organization for Standardization, Geneva.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jun1814

[9]. Singh, R. (2015). Membrane Technology and Engineering for Water Purification (2nd ed.). Butterworth-Heinemann.
[10]. Nair, M., & Kumar, D. (2021). Salinization of coastal aquifers in Kerala: Magnitude, drivers, and remediation strategies. Hydrogeology Journal, 29(4), 1421–1439.
[11]. Tiwari, G. N., & Tiwari, A. K. (2008). Solar Distillation Practice for Water Desalination Systems. Anshan Publishers.
[12]. UNICEF / WHO. (2023). Progress on Household Drinking Water, Sanitation and Hygiene 2000–2022. UNICEF, New York.