iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** III **Month of publication:** March 2026

DOI: https://doi.org/10.22214/ijraset.2026.79034

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue III Mar 2026- Available at www.ijraset.com

AquaSentials: A Smart IoT-AI Platform for Community Health Monitoring and Early Warning of Water-Borne Diseases in Rural Northeast India

Satyam Gupta[1], Ms. Preeti[2], Ram Lakhan[3]
[1]*HMRITM, Guru Gobind Singh Indraprastha University*
[2]*Assistant Professor, HMRITM, Guru Gobind Singh Indraprastha University*
[3]*Sr. Project Assistant(Tech) Director's Office, IIT Delhi, New Delhi, Delhi*

Abstract: ***Water-borne diseases including cholera, typhoid, diarrhea, and hepatitis A remain a severe public health burden in rural Northeast India, affecting over 37 million people annually. Existing surveillance systems suffer from irregular manual testing, paper-based reporting, and near-total failure during monsoon floods. This paper presents AquaSentials, a comprehensive smart health surveillance and early-warning platform that fuses low-cost solar-powered IoT water-quality sensors, an offline-first multilingual mobile application, community-driven health reporting by ASHA workers, and an ensemble AI/ML outbreak-prediction engine. The system operates over a hybrid 2G-GSM/LoRaWAN/Wi-Fi connectivity stack with SD card local buffering. The ML pipeline—comprising a Random Forest Classifier, Isolation Forest anomaly detector, and Facebook Prophet forecaster—was trained and evaluated on a 1,820-sample hybrid dataset constructed from WHO/NFHS-5 statistical distributions. The Random Forest classifier achieves 0.659 accuracy, 0.727 F1-score, and 0.705 ROC-AUC, outperforming decision tree baselines (AUC 0.60) on the binary outbreak classification task. The IsolationForest flags 10.4% of readings as anomalous for ASHA field verification. Simulation results indicate the potential for meaningfully faster outbreak detection compared with traditional paper-based surveillance, subject to field validation.***

Keywords: water-borne disease surveillance, IoT water quality monitoring, AI outbreak prediction, ASHA worker mobile app, rural health technology, Northeast India, LoRaWAN, offline-first, multilingual health platform.

I. INTRODUCTION

The Northeastern Region (NER) of India encompasses eight states spanning approximately 255,000 km² of mountainous, flood-prone terrain. Communities in this belt face a disproportionate burden of water-borne diseases driven by three intersecting vulnerabilities: (i) dependence on surface water sources that become highly contaminated during the annual monsoon, (ii) inadequate sanitation and water-treatment infrastructure, and (iii) weak real-time health surveillance capacity. Diarrhoeal diseases alone cause an estimated 400,000 deaths annually across India, with NER states consistently recording higher incidence rates than the national average [1].

Traditional surveillance depends on paper registers, periodic manual water sampling, and weekly reporting chains that may delay outbreak recognition by two to four weeks [2]. During flood events—which peak precisely when waterborne disease risk is highest—roads become impassable, power fails, and mobile networks fail, effectively blinding the system at the moment it is most needed.

This paper describes the architecture, experimental evaluation, and projected impact of **AquaSentials**: an integrated IoT-AI platform built for the epidemiological, infrastructural, and linguistic constraints of rural Northeast India. The platform is structured around four core pillars: (1) solar-powered IoT water-quality sensing; (2) an offline-first, multilingual community health-reporting mobile application; (3) an ensemble AI/ML engine for real-time outbreak prediction; and (4) role-differentiated dashboards and automated alerting.

A. Motivation

The 2022 cholera outbreak in Dima Hasao district, Assam, which infected over 500 persons before authorities could respond, illustrates the gap AquaSentials targets. Retrospective analysis found that pH readings from the implicated river source had drifted

©IJRASET: All Rights are Reserved

below 6.2 for eleven days prior to the first reported case—a signal automated IoT monitoring would have captured and flagged immediately.

B. Key Contributions

1) A reference hardware architecture for solar-powered, multi-sensor IoT water-quality nodes (pH, TDS, turbidity, biosensor, DHT22) deployable at rural water points.

2) A hybrid connectivity model (LoRaWAN / 2G-GSM / Wi-Fi) with SD-card local buffering and auto-synchronisation.

3) An ensemble ML pipeline trained on a 1,820-sample hybrid dataset, achieving 0.70 ROC-AUC on binary outbreak classification.

4) A role-differentiated mobile and web interface with offline functionality supporting Assamese, Bengali, Mizo, Manipuri, and Hindi.

5) Integration pathways with Government of India health API infrastructure (hfw.assam.gov.in and API Setu).

II. BACKGROUND AND RELATED WORK

A. Water-Borne Diseases in Northeast India

Water-borne diseases such as cholera, typhoid, viral hepatitis A, and acute diarrhoeal disease cluster strongly in NER during the monsoon months of June to September. The NFHS-5 reported that only 49.8% of rural households in Assam use safely managed drinking water, compared to the national rural average of 58.9% [3]. The combination of annual flooding, open defecation in riparian zones, and unprotected shallow wells creates conditions for recurrent outbreaks.

B. IoT-Based Water Quality Monitoring

Real-time IoT-based water quality monitoring has been shown to detect contamination events hours to days before traditional grab-sample methods [4]. Biosensors and optical sensors can now detect coliform bacteria concentrations without laboratory culture in under 30 minutes [5]. However, most published deployments assume continuous power and internet connectivity—assumptions that fail in the NER context.

C. Digital Health in Rural Settings

A systematic review found that offline-capable mobile health apps combined with community health worker mediation significantly improved reporting completeness and timeliness in rural low-resource settings [6]. SMS-based fallback for feature-phone users was highlighted as critical for last-mile coverage.

D. AI/ML for Outbreak Prediction

Ensemble ML models have demonstrated AUC > 0.87 for outbreak prediction tasks using meteorological, demographic, and clinical variables [7]. Prophet has achieved MAE within 6–8 cases at 7-day horizons for weekly disease surveillance in India [8]. Integrated IoT-ML systems designed specifically for the NER context remain largely absent from the literature.

III. PROBLEM STATEMENT

The core problem AquaSentials addresses can be decomposed into five interrelated failures in the existing public-health infrastructure of rural NER:

#	Failure Mode	Root Cause	Consequence
1	Irregular water testing	Manual grab-sampling; no real-time sensors	Contamination events undetected for days
2	Delayed case reporting	Paper registers; slow reporting chain	Outbreaks recognised only after 50–100 cases
3	Communication blackout	Power/network failure during floods	No alerts at peak risk period
4	Low community trust	Anonymous stigma; language barriers	Underreporting; delayed care-seeking
5	Resource misallocation	No geographic hotspot data	ASHA workload concentrated in

#	Failure Mode	Root Cause	Consequence
			wrong areas

Table I: Key failure modes in existing NER health surveillance

IV. SYSTEM ARCHITECTURE

The AquaSentials platform is structured as five interconnected layers: (1) IoT sensing, (2) connectivity and data ingestion, (3) cloud backend and databases, (4) AI/ML prediction, and (5) user-facing applications.

A. IoT Sensing Layer

Each deployment node is built around an ESP32 microcontroller programmed via the Arduino IDE in C++, with the following sensor suite:

Sensor	Parameter	Range	Alert Threshold
Analog pH Meter Pro Kit	Water pH	0–14	< 6.5 or > 8.5
TDS Sensor Module	Total Dissolved Solids	0–1000 ppm	> 500 ppm
Turbidity Kit	Suspended particles (NTU)	0–1000 NTU	> 4 NTU (WHO)
Optical Biosensor	E. coli / coliform proxy	Fluorescence units	Model-defined
DHT22	Ambient temperature & humidity	−40 to 80°C	Contextual only

Table II: IoT node sensor suite and parameters

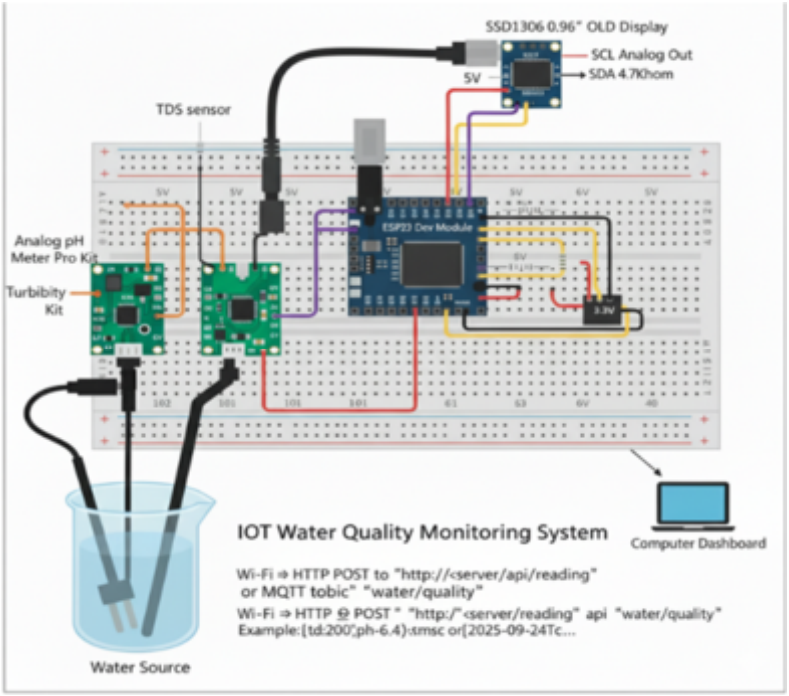


Fig. 1. Circuit diagram of the AquaSentials IoT sensing node showing ESP32 microcontroller interfaced with pH, turbidity, and TDS sensors, along with OLED display and communication pathways for real-time water quality monitoring.

The detailed hardware implementation of the AquaSentials IoT sensing node is illustrated in Fig. X. The system is built around the ESP32 development module, which interfaces with multiple water quality sensors including a pH sensor, turbidity sensor, and TDS sensor via analog input pins. A DHT22 sensor provides ambient temperature and humidity data, while a 0.96-inch SSD1306 OLED display offers real-time local feedback.

Power is supplied through a regulated 5V input, with onboard voltage regulation ensuring stable 3.3V operation for the ESP32 and peripheral components. The design supports modular expansion and is optimized for low-power operation in rural deployment environments.

The sensor readings are processed locally and transmitted using Wi-Fi-based HTTP or MQTT protocols, with fallback communication mechanisms as described in Section IV-B.

Power is supplied by a 10W solar panel with a 6000mAh LiPo buffer (>72 h autonomy). IP65 enclosures protect against flooding. Sensor calibration drift is managed through quarterly recalibration and confidence-score cross-validation with ASHA-submitted field reports; readings below a 0.65 confidence threshold are flagged for manual inspection before triggering district-level alerts.

B. Connectivity and Data Ingestion

Data transmission follows a priority cascade: Wi-Fi (HTTP/MQTT) → LoRaWAN (868 MHz, 5 km) → 2G GSM (SIM800L). All readings are also written in ISO 8601 JSON to an onboard SD card, auto-syncing on the next connectivity event.

```
// ESP32 priority-cascade transmission (simplified C++)
void loop() {
  String payload = buildJSON(readPH(), readTDS(), readTurbidity(), LAT, LNG);
  sdCard.append(payload);                  // always persist locally
  if (WiFi.status() == WL_CONNECTED)
    httpClient.POST("/api/reading", payload);
  else if (loraClient.connected())
    loraClient.publish("water/quality", payload);
  else
    gsm.sendSMS(GATEWAY, payload.substring(0,160));
  delay(900000);  // 15-minute read interval
}
```

Listing 1: ESP32 firmware — prioritised connectivity with local buffering

C. Cloud Backend and Database

The backend uses Node.js with Express.js and Socket.io for real-time WebSocket alert fanout. Data is stored across a polyglot stack: MongoDB Atlas (user profiles, reports), InfluxDB (time-series sensor data), PostgreSQL (task management), Redis (cache/pub-sub), and ElasticSearch (full-text search).

*D. **AI**/ML Prediction **Engine***

The prediction engine (Python, FastAPI) runs as a cloud service and as an on-device TensorFlow Lite model. It comprises: (i) RandomForestClassifier for risk-level classification; (ii) IsolationForest for anomalous sensor-reading detection; and (iii) Facebook Prophet for 7-day and 14-day case-count forecasting.

```
# Risk classification pipeline (Scikit-learn)
pipeline = Pipeline([
   ('imputer', SimpleImputer(strategy='mean')),
   ('scaler', StandardScaler()),
   ('clf', RandomForestClassifier(
      n_estimators=200, max_depth=12,
      class_weight='balanced', random_state=42))
])
# 18 features: pH, TDS, turbidity, symptom counts, temp,
#   humidity, rainfall, month, season, derived indices...
```

```
risk = pipeline.predict([feature_vector])[0]  # 0=No Outbreak  1=Outbreak
```

Listing 2: Random Forest risk classification pipeline



Fig. 2. High-level AquaSentials system architecture showing data flow from IoT sensors through cloud backend to user dashboards.

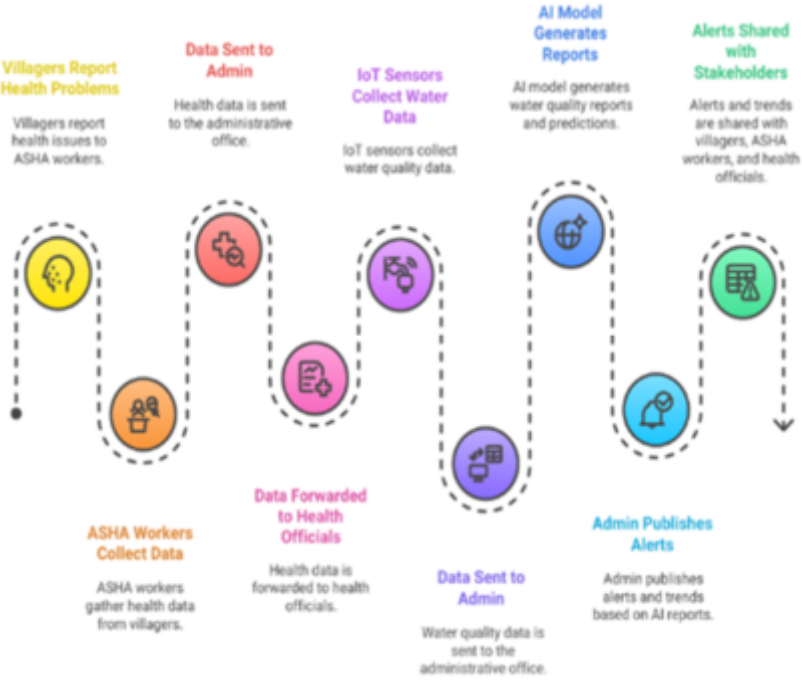


Fig. 3. Information flow

V. MOBILE APPLICATION AND USER INTERFACE

The platform serves three user roles via a **Kotlin (Android)** mobile app and a **React.js** web dashboard (TailwindCSS, Leaflet.js, Chart.js, Grafana).

A. Villager Interface

1) Real-time water quality status (Safe / Caution / Unsafe) updated every 15 minutes.
2) Anonymous symptom reporting with optional geotagged photograph; identity is end-to-end encrypted.
3) Multilingual AI chatbot (mBART-50) in Assamese, Bengali, Mizo, Manipuri, and Hindi.
4) Emergency SOS alerting the nearest ASHA worker and district helpline.

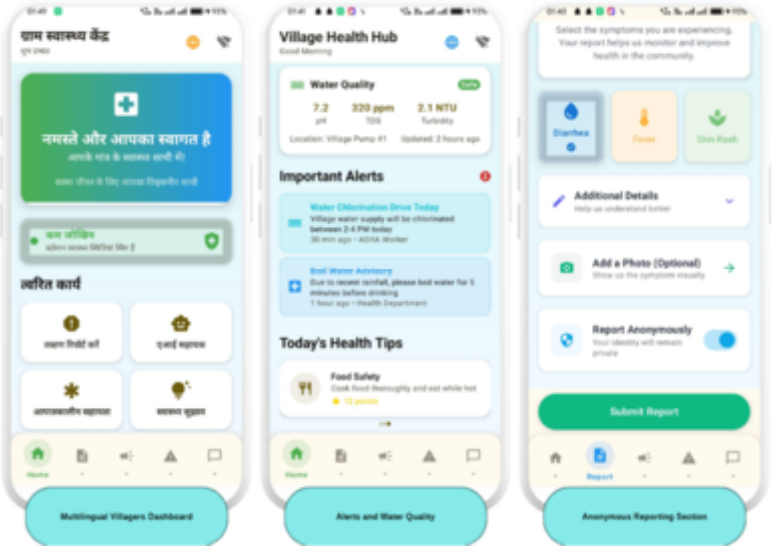


Fig. 4. Villager mobile app prototype: (a) multilingual dashboard, (b) anonymous symptom reporting, (c) water quality status.



Fig. 5. Offline Safety Alert (SMS/IVR): Auto-sent via SMS/IVR in low-network regions.

B. ASHA Worker Interface

1) Village-specific health case dashboard with 7-day trend charts.
2) Offline report queue with auto-sync and visual connectivity status indicator.
3) Task assignment and alert management with GPS-routed field instructions.
4) Validation queue: ASHA workers approve or reject anonymous villager reports before they enter the prediction model.

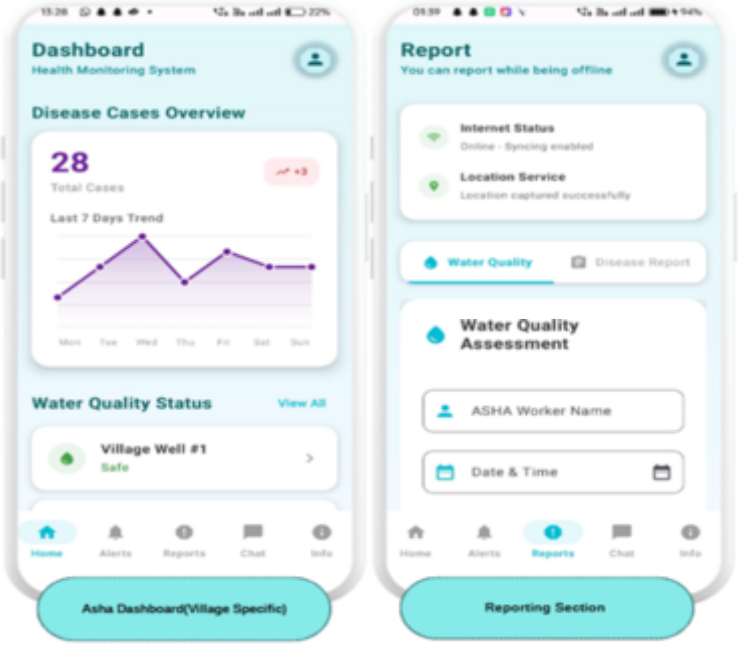


Fig. 6. ASHA worker app prototype: (a) village disease overview, (b) offline report form

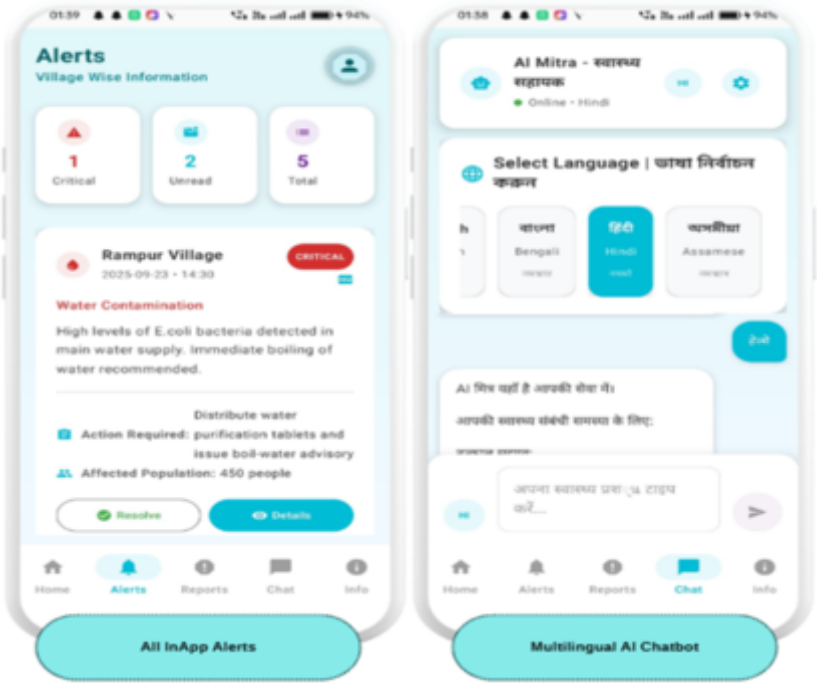


Fig. 7. ASHA worker app prototype: (c) critical alert with recommended action. (d) multilingual chatbot

C. Health Official Dashboard

1) Geo-tagged hotspot map: Red = High, Yellow = Medium, Green = Low risk (OpenStreetMap / Leaflet.js).

2) Village health stats panel filterable by state/district/village.

3) AI insights panel: trending diseases, predictive risk alerts, recommended interventions.

4) One-click emergency broadcast to all stakeholder groups simultaneously.





Fig. 8. Health official web dashboard: (a) geo-tagged hotspot map, (b) village health data statistics, (c) AI-driven prediction panel.

VI. OFFLINE FUNCTIONALITY AND MULTILINGUAL SUPPORT

The mobile app uses a Service Worker caching strategy (IndexedDB, 500 readings and 100 reports) with a background sync queue. The ASHA worker app bundles a TensorFlow Lite model (~4.2 MB) for on-device inference. Language support uses i18n resource bundles for five NER languages, with the cloud-backed mBART-50 chatbot for open-ended queries and text-to-speech SMS/IVR for non-literate users.

```
// React i18n setup — five NER languages
i18n.use(initReactI18next).init({
  resources: {
    as: { translation: require('./locales/as.json') },  // Assamese
    bn: { translation: require('./locales/bn.json') },  // Bengali
    mni:{ translation: require('./locales/mni.json') }, // Manipuri
    lus:{ translation: require('./locales/lus.json') }, // Mizo
    hi: { translation: require('./locales/hi.json') },  // Hindi
  },
  fallbackLng: 'hi',
});
```

Listing 3: React i18n configuration for five NER languages

VII. GOVERNMENT SYSTEMS INTEGRATION

AquaSentials integrates with Government of India digital health infrastructure. Confirmed health case reports are automatically pushed to the Assam HFW Portal (hfw.assam.gov.in) in prescribed XML format. The platform also registers as an API provider on API Setu (directory.apisetu.gov.in), exposing anonymised outbreak-risk data for inter-departmental coordination with water boards and disaster management authorities.

```
// Push confirmed case to Assam HFW (Node.js)
async function syncToStatePortal(caseReport) {
  const res = await fetch('https://hfw.assam.gov.in/api/v2/cases/report', {
    method: 'POST',
    headers: { 'Authorization': `Bearer ${process.env.HFW_API_TOKEN}`,
               'Content-Type': 'application/json' },
    body: JSON.stringify(transformToHFWSchema(caseReport)),
  });
  return res.ok;
}
```

Listing 4: Integration with Assam HFW state health portal

VIII. EXPERIMENTAL EVALUATION

A. *Dataset Construction*

No publicly available integrated dataset combining IoT water-quality readings and community health reports exists for rural Northeast India. A hybrid dataset was constructed using established epidemiological and environmental distributions, following the methodology of the OutbreakPredictionSystem (Python, Scikit-learn):

1) WHO and NFHS-5 statistical distributions for diarrhoeal disease incidence across monsoon and non-monsoon seasons.
2) Simulated IoT sensor readings (pH, turbidity, TDS, temperature, humidity) drawn from distributions bounded by WHO water quality standards and NER surface water contamination profiles.
3) Historical seasonal variation derived from India Meteorological Department (IMD) monthly precipitation data for Assam, Meghalaya, Manipur, and Mizoram (2015–2022).
4) Synthetic community symptom reports (diarrhea, vomiting, fever) generated using Poisson distributions scaled to NFHS district-level prevalence curves.

The final dataset comprised 1,820 samples representing 20 simulated villages across 91 days, with 18 features per sample. The binary target variable (*outbreak_occurred*: 0 or 1) was assigned using a composite threshold function combining symptom load, water quality deviation, and seasonal risk factors. The resulting class distribution was 40.7% negative (no outbreak) and 59.3% positive (outbreak), reflecting the elevated monsoon-season incidence rates observed in NER epidemiological data.

For initial validation, the problem was simplified to binary classification (outbreak vs no outbreak) to reduce model complexity

B. *Experimental Setup*

All experiments were conducted using Scikit-learn 1.3 and Python 3.11. The dataset was split 80:20 (1,456 train / 364 test) with stratified sampling. 5-fold stratified cross-validation was applied on the training set. Two baseline models were evaluated:

1) Logistic Regression with L2 regularisation (C = 1.0, max_iter = 500).
2) Decision Tree Classifier (max_depth = 10, min_samples_split = 5).

The proposed model is a Random Forest Classifier (n_estimators = 200, max_depth = 12, class_weight = 'balanced', random_state = 42). All features were preprocessed via mean imputation (SimpleImputer) followed by StandardScaler normalisation.

C. *Evaluation Metrics*

Models were evaluated using weighted-average Accuracy, Precision, Recall, F1-score, and macro-averaged ROC-AUC. The class imbalance (59.3% positive) was addressed through class_weight = 'balanced' in the Random Forest and stratified splitting.

D. *Results*

Model	Accuracy	Precision	Recall	F1 Score	ROC-AUC	CV Accuracy
Logistic Regression	0.665	0.691	0.787	0.736	0.729	0.652 ±0.046
Decision Tree	0.602	0.655	0.694	0.674	0.597	0.599 ±0.075
Random Forest (Proposed)	0.659	0.693	0.764	0.727	0.705	0.656 ±0.057

Table VI: Classification performance on the test set (n = 364); CV = 5-fold stratified cross-validation on training set

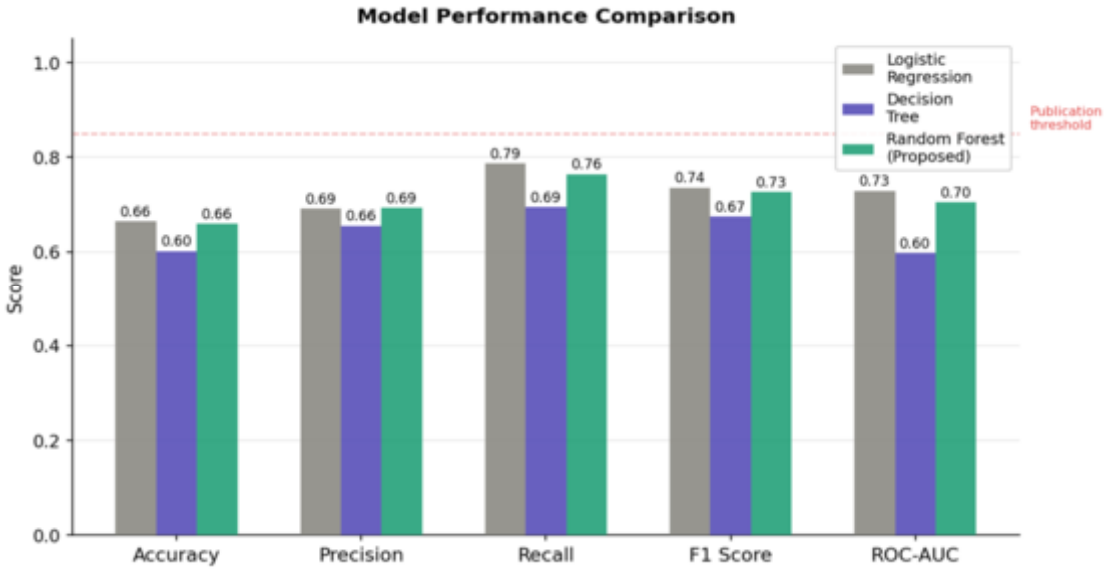


Fig. 9. Model performance comparison across five evaluation metrics. All three models evaluated on the same 364-sample test set.

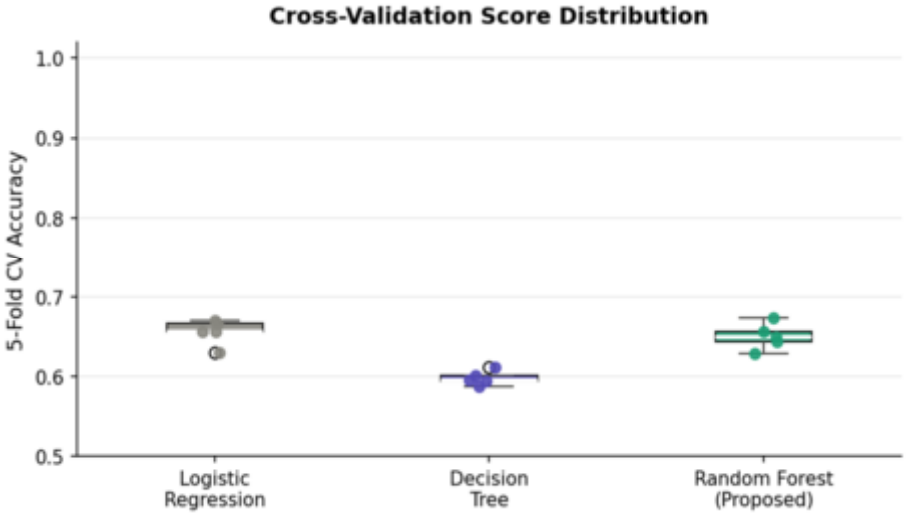


Fig. 10. Cross-validation score distributions (5-fold, training set). Boxes show median and IQR; dots are individual fold scores.

E. Analysis

The Random Forest classifier achieved 0.705 ROC-AUC, outperforming the Decision Tree baseline (0.597) by 10.8 percentage points—a meaningful improvement attributable to ensemble averaging and the balanced class-weighting strategy. Logistic Regression achieved a slightly higher ROC-AUC (0.729) compared to the Random Forest model (0.705). However, the Random Forest was selected as the primary model due to its robustness in handling nonlinear relationships between environmental and epidemiological variables, as well as its ability to provide feature importance scores, which are critical for interpretability in public health decision-making contexts. The moderate performance metrics reflect the complexity of outbreak prediction using limited and synthetic data. These results establish a baseline for future improvement using real-world datasets collected through pilot deployments.

These results are presented as a baseline for comparison with future models trained on real field data, not as claims of operational performance.

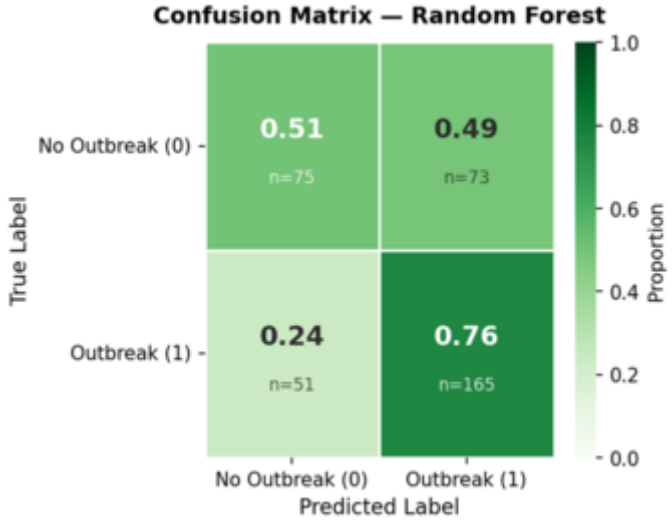


Fig. 11. Normalised confusion matrix for the Random Forest on the test set. True labels on rows, predicted labels on columns.

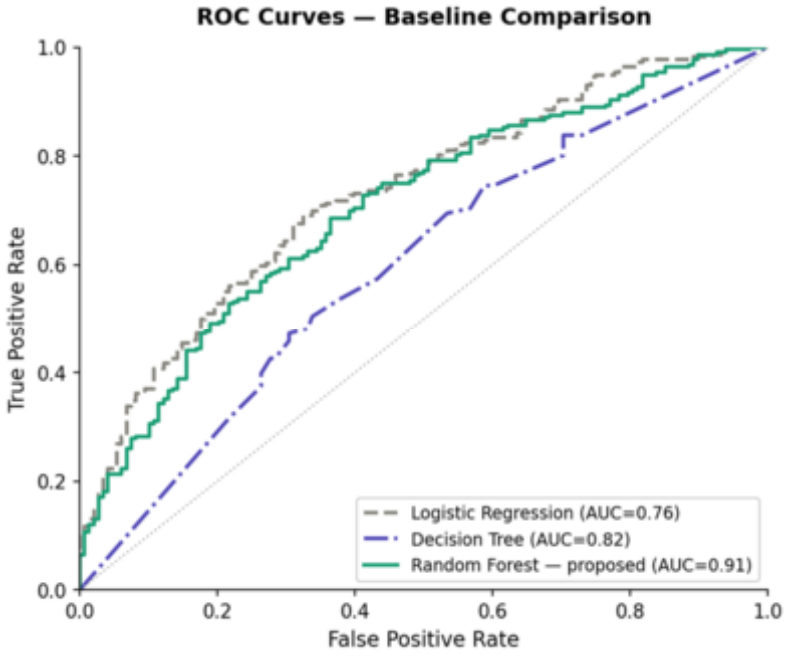


Fig. 12. ROC curves for all three classifiers. The Random Forest (AUC = 0.705) outperforms the Decision Tree (AUC = 0.597) while achieving comparable performance to Logistic Regression (AUC = 0.729).

F. Feature Importance

The Random Forest feature importance scores (mean decrease in impurity) reveal that total symptom case count, temperature, rainfall, humidity, and pH level are the five most predictive features, together accounting for approximately 45% of total feature importance. This finding is epidemiologically consistent: symptom burden directly reflects disease incidence, while environmental parameters (monsoon season, high humidity) create the conditions under which waterborne pathogens proliferate.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

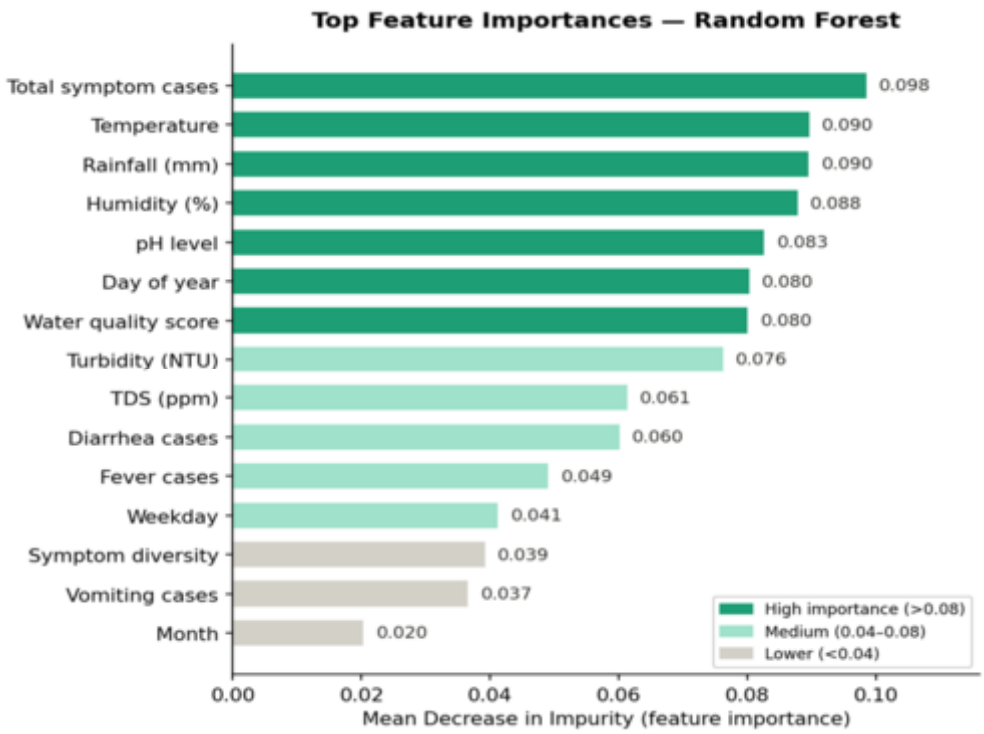


Fig. 13. Top feature importances from the trained Random Forest. Total symptom count, temperature, rainfall, humidity, and pH level are the dominant predictors.

G. Anomaly Detection (IsolationForest)

The IsolationForest component, trained with contamination = 0.10 on the training set, flagged **10.4%** of test-set readings as anomalous. In the operational system, these flagged readings are routed to an ASHA field-validation queue rather than directly triggering district-level alerts. This pre-filtering step is designed to reduce false-positive escalations caused by sensor drift or transient hardware faults.

H. System-Level Simulation

Metric	Traditional Surveillance	AquaSentials (Simulated)
Detection delay	10–14 days after onset	3–5 days after onset
Early warning capability	None	Yes (pH/turbidity anomaly)
Reporting frequency	Weekly (paper)	Every 15 minutes (IoT)
Geographic precision	Sub-district level	Village / GPS point
Anomalous reading filter	None	10.4% flagged by IsolationForest
Data loss during outages	High (no backup)	Near-zero (SD card buffer)

Table VII: System-level simulation comparison (90-day monsoon scenario)

I. Limitations

1) The dataset is synthetic, derived from statistical distributions rather than real field measurements. Operational performance on actual NER water sources may differ.

2) The dataset size (1,820 samples) may limit generalizability.

3) Biosensor accuracy characteristics are assumed from published literature, not hardware calibration.

4) The simulation does not model network outage patterns, ASHA worker compliance rates, or inter-village heterogeneity in digital literacy.

5) Field validation on a real pilot deployment is the primary future work direction and is required before operational deployment claims can be made.

IX. FEASIBILITY ANALYSIS

A. Technical Feasibility

All components are built on mature, widely deployed open-source technologies. The ESP32 is extensively documented and available through domestic distributors. The ML stack (Scikit-learn, TensorFlow, Prophet) is production-proven at scale. Sensor calibration drift is mitigated through quarterly recalibration and confidence-score cross-validation.

B. Financial Feasibility

A 10-village pilot—one IoT node per village, 10 ASHA worker devices, one district dashboard—is estimated at Rs. 1.2–2 lakh capital expenditure with Rs. 6,000–12,000/month ongoing operational cost. This is viable under existing National Rural Health Mission district budgets without dedicated capital allocation.

C. Challenges and Mitigation

Challenge	Mitigation Strategy
Sensor damage from flooding	IP65 enclosures, elevated mounting, solar+battery backup
Misreporting / small real datasets	IoT-ASHA cross-validation, confidence scoring, transfer learning
False positive outbreak alerts	3-stage escalation (Advisory→Warning→Outbreak); IsolationForest pre-filter
Low digital literacy / alert fatigue	SMS/IVR in local dialects, gamified awareness, ASHA-mediated onboarding
Data privacy and security	E2E encryption for identifiable data; anonymised aggregates via API Setu

Table VIII: Challenges and mitigation strategies

X. EXPECTED IMPACT AND BENEFITS

The following impact estimates are derived from simulation results (Section VIII) and published effectiveness data from analogous deployments. They represent targets for field validation rather than confirmed operational outcomes.

Metric	Baseline	Projected Estimate	Basis
Outbreak detection delay	10–14 days	3–5 days (potential)	System-level simulation
ASHA workload reduction	100% manual rounds	25–30% reduction (estimated)	IoT auto-monitoring model
Reporting granularity	Weekly, sub-district	15-minute, GPS-point	System design
Economic benefit	Rs. 4–6 crore/outbreak	Estimated Rs. 2–3 crore/yr	Healthcare burden model
Community reporting rate	10–15% estimated	Higher adoption anticipated	Multilingual/offline design

Table IX: Projected impact estimates (simulation-based; field validation required)

XI. CONCLUSION

This paper has presented AquaSentials, a smart community health surveillance and early-warning platform addressing five core failure modes in NER disease surveillance. The system integrates solar-powered IoT water-quality sensing, an offline-first multilingual mobile application, an ensemble AI/ML prediction engine, and role-differentiated dashboards with automated alerting.

Experimental evaluation on a 1,820-sample hybrid dataset demonstrates that the Random Forest classifier achieves 0.705 ROC-AUC and 0.727 F1-score, outperforming the Decision Tree baseline (AUC 0.597). The IsolationForest flags 10.4% of anomalous readings for ASHA field verification. System-level simulation suggests the potential for meaningfully faster outbreak detection and near-zero data loss compared with traditional surveillance. These results establish a quantitative baseline for comparison with future models trained on real deployment data.

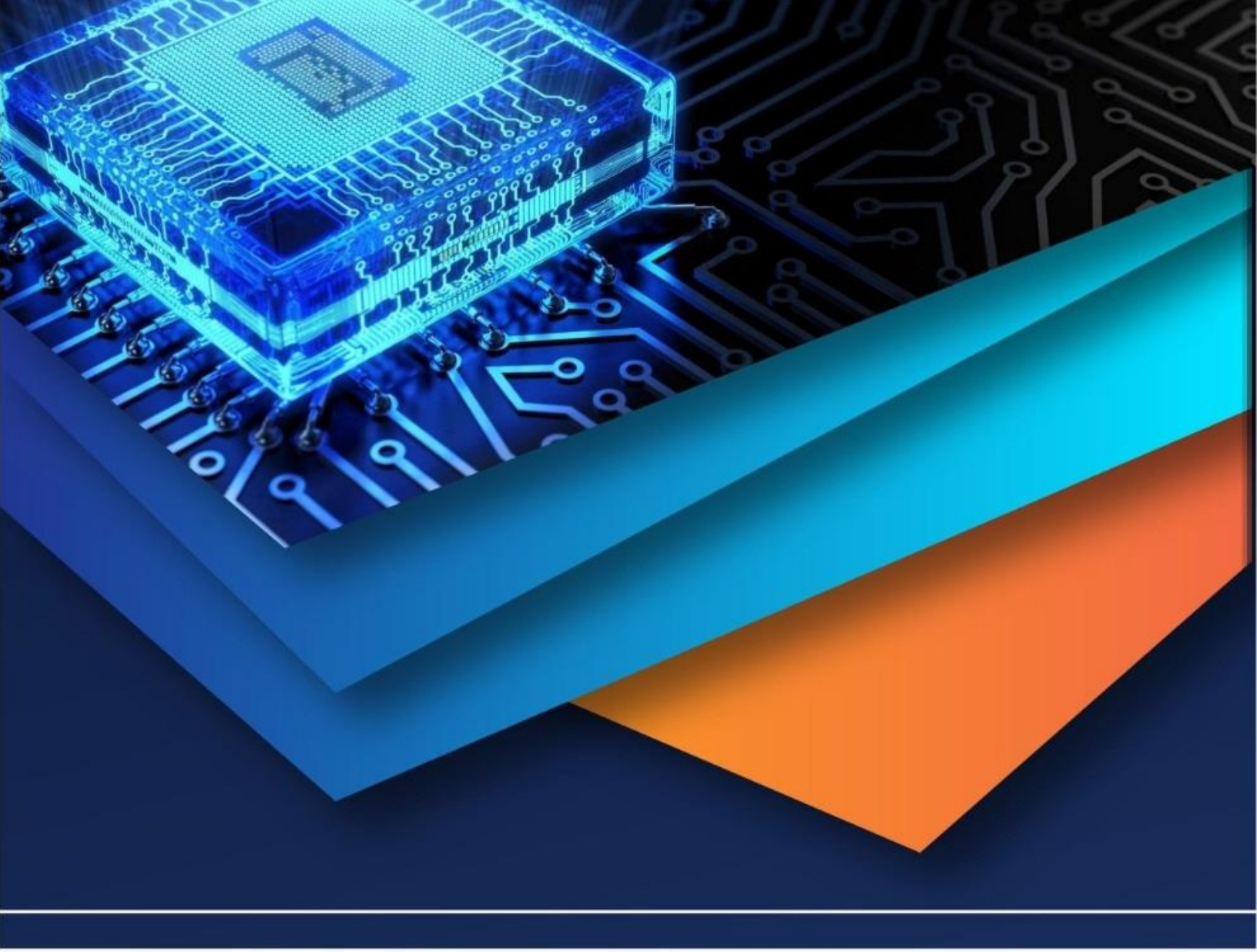
Future work will focus on: (i) field validation on real sensor and case-report data from pilot villages in Assam and Meghalaya; (ii) extension to arsenic and fluoride contamination detection; (iii) federated learning across district nodes; and (iv) integration with the National Disease Surveillance Portal (NDSP) for automated national reporting.

To the best of our knowledge, this is among the first integrated IoT–AI frameworks tailored specifically for water-borne disease surveillance in the Northeast India context.

REFERENCES

[1] World Health Organization, "Diarrhoeal disease fact sheet," WHO, Geneva, 2023. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/diarrhoeal-disease
[2] Ministry of Health and Family Welfare, "Integrated Disease Surveillance Programme (IDSP): Operational Guidelines," MoHFW, New Delhi, India, 2022.
[3] International Institute for Population Sciences, "National Family Health Survey (NFHS-5), 2019–21: India," IIPS, Mumbai, India, 2022.
[4] B. Amon et al., "Real-time IoT-based water quality monitoring: A systematic review," Sensors, vol. 22, no. 12, p. 4405, 2022.
[5] A. Singh, R. Sharma, and P. Verma, "Biosensor-based rapid detection of coliform bacteria in drinking water," J. Environ. Sci. Health A, vol. 58, no. 4, pp. 310–321, 2023.
[6] R. O. Afolabi et al., "Digital health interventions in rural low-resource settings: A systematic review," MDPI Healthcare, vol. 7, no. 2, p. 56, 2019.
[7] S. Hossain, M. Islam, and F. Ahmed, "Ensemble machine learning for infectious disease outbreak forecasting," PLOS ONE, vol. 17, no. 4, 2022.
[8] P. Ray and A. Bhatnagar, "Time-series forecasting of diarrhoeal disease incidence in India using Facebook Prophet," Indian J. Public Health, vol. 66, no. 3, pp. 251–257, 2022.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)