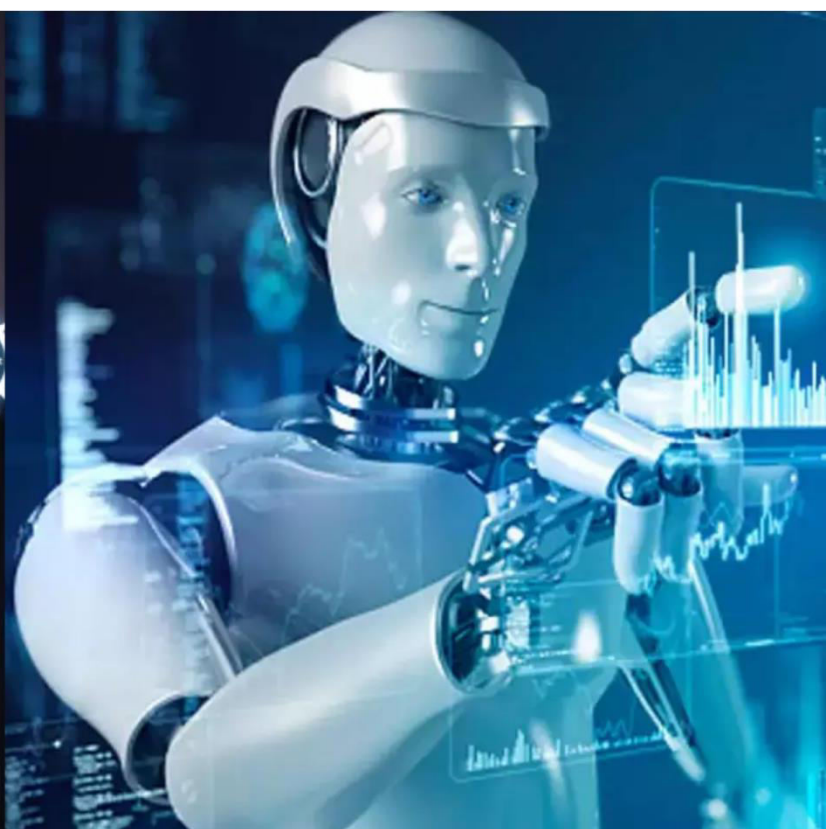




International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 9.935

Volume 15, Issue 10, October 2026

A Centralised AI-Driven Monitoring and Decision Support System for Primary Health Centres using CatBoost-Based Absenteeism Prediction

Dr. A. Askarunisa¹, Maha Sree P², Mathu Mathi T S³

Head of the Department, Department of Computer Science and Engineering, K.L.N. College of Engineering,
Tamil Nadu, India¹

Department of Computer Science and Engineering, K.L.N. College of Engineering, Tamil Nadu, India²

Department of Computer Science and Engineering, K.L.N. College of Engineering, Tamil Nadu, India³

ABSTRACT: Primary Health Centres (PHCs) and Sub-Centres are the first point of contact in public healthcare, yet district authorities still monitor them through manual registers that reach the district office late and inconsistently. This paper presents a Centralised PHC Monitoring System that gives the Deputy Director of Health Services (DDHS) a single role-based web and mobile platform for attendance, patient-service and equipment data. Incoming records are validated and normalised with lightweight NLP-based cleaning and stored in a central Microsoft SQL Server database. An analytics engine then (i) predicts staff absenteeism with a CatBoost classifier, (ii) detects anomalies in outpatient load using a z-score rule with short explanations, (iii) ranks facilities with a composite 0–100 performance score, and (iv) produces decision-support recommendations, such as staff transfers, that remain subject to human approval. Alerts are routed by severity and role to doctors or the District Health Officer. The system is implemented with React.js, FastAPI and SQL Server. A six-facility prototype demonstrates end-to-end operation, yielding a facility ranking that re-ranks automatically as outpatient data are logged, active anomaly alerts and a reasoned transfer suggestion, while unit, functional, integration and performance testing confirm correct behaviour across all layers. The results indicate that centralised data combined with lightweight predictive analytics can move district-level monitoring from retrospective reporting to proactive intervention.

KEYWORDS: Absenteeism prediction, anomaly detection, CatBoost, clinical decision support, healthcare analytics, primary health centres, public health informatics.

I. INTRODUCTION

Primary Health Centres (PHCs), upgraded PHCs and Sub-Centres form the backbone of public healthcare in many countries, delivering outpatient care, basic laboratory services, immunisation and maternal and child health services close to the community. In the district-level administrative structure considered in this work, these facilities report to a Deputy Director of Health Services (DDHS), who must ensure that every centre is adequately staffed, equipped and functioning. A single DDHS office may oversee many facilities, each with its own doctors, nursing staff, equipment and stock position.

In practice, this oversight is largely manual. Attendance and patient counts are written in registers and consolidated weekly or monthly, equipment shortages become visible only when a facility reports them, and workload differences between neighbouring facilities are rarely compared. Staff absence therefore tends to be noticed only after patients are affected, and one facility may be overloaded while another is under-used.

Machine learning (ML) is widely discussed for diagnosis and imaging [7], but its administrative and operational uses, such as predicting staff availability and patient load, receive less attention even though they decide whether a facility can deliver care at all. A monitoring platform that is both centralised and predictive would let district officers act before a service disruption occurs rather than after.

This paper presents such a platform. The main contributions are: 1) a role-based, web and mobile centralised data-collection platform for PHC attendance, patient services and equipment inventory with validation and NLP-based

cleaning at the point of entry; 2) an analytics engine that combines CatBoost-based absenteeism prediction, anomaly detection on outpatient load, a transparent composite performance score and a rule-based staff-transfer recommender; 3) a severity- and role-aware alert-routing layer that keeps the final decision with the authorised officer; and 4) a working implementation and a six-facility prototype evaluation covering functional, integration and performance behaviour.

The remainder of this paper is organised as follows. Section II reviews the literature, Section III presents the proposed methodology (architecture, analytical methods and implementation) with discussion, Section IV reports the results, and Section V concludes.

II. LITERATURE SURVEY

Public-health data platforms. Khalique et al. [1] proposed the PHF framework, which acquires, fuses and anonymises electronic health records from multiple primary facilities through an HL7 ADT gateway, supporting population-level analytics such as surveillance and immunisation tracking. Its standards-based design is attractive, but it presumes that basic digitisation already exists at every site, which may not hold for low-resource PHCs, and it targets patient-record analytics rather than facility administration.

Absenteeism prediction. Rista et al. [2] compared Random Forest, Decision Tree, SVM and k-NN classifiers on about 125,000 employment records and used the best model to rank the factors that influence absence. The study confirms that tabular attributes such as day, month and service length are informative, but it relies on a single organisation's data and ignores healthcare-specific seasonal triggers.

AI-based health monitoring. Kushwaha et al. [3] combined IoT sensors with CNN classifiers to monitor vital signs and predict disease. Such approaches reduce hospital burden but require costly real-time deep-learning inference and operate at the patient level, which makes them hard to deploy in resource-constrained primary care and unsuitable for facility-level administration.

Learning and anomaly-detection methods. CatBoost [4] uses ordered boosting and native handling of categorical variables to reduce prediction shift, which suits attributes such as facility, staff role and weekday. Chandola et al. [5] survey anomaly-detection techniques and show that the choice of method depends on the nature of the data; they do not address healthcare operational series such as daily outpatient counts. Scikit-learn [6] provides the preprocessing and evaluation tooling used in many such pipelines.

None of the above integrates attendance, service and inventory data from a network of primary facilities with prediction, anomaly detection, ranking and actionable recommendations for a district administrator. The present work addresses this gap.

III. PROPOSED METHODOLOGY AND DISCUSSION

A. System Architecture

The system follows a layered, modular architecture (Fig. 1): a frontend layer, a backend layer, a core data-processing layer, an AI and ML analytics engine, an alerting and notification layer, and a centralised database. Three user roles are supported. Doctors mark attendance, log patient data and view their facility dashboard and alerts. Staff mark attendance, enter patient counts and maintain the equipment inventory. The DDHS Admin monitors all facilities, reviews rankings and alerts, approves or dismisses suggested transfers and generates KPI reports.

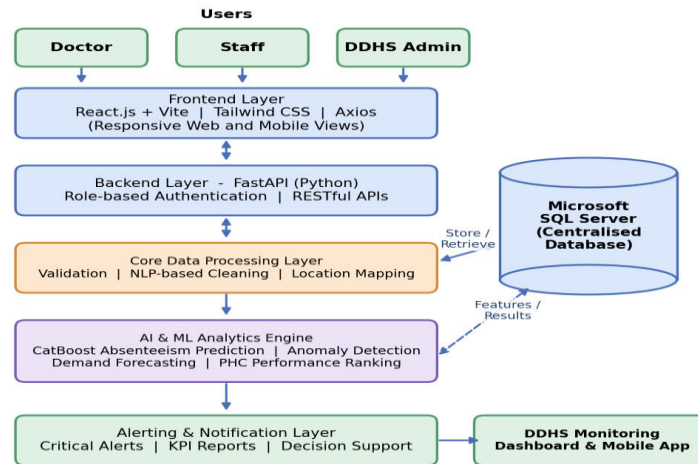


Fig.1. Layered architecture of the Centralised PHC Monitoring System

The React.js frontend collects inputs and calls RESTful FastAPI endpoints. The backend authenticates each user, checks the role and forwards the payload to the core processing layer, where validation, NLP-based cleaning and location mapping are applied. Cleaned data is stored in Microsoft SQL Server. The analytics engine reads historical and current data from the database, writes its outputs back, and passes them to the alerting layer, which fires alerts, produces KPI reports and feeds the DDHS dashboard and mobile application.

The level-1 data-flow diagram in Fig. 2 contains four processes, two external entities and three data stores. Process 1.0 verifies credentials and roles against the user-accounts store (D1). Process 2.0 validates and writes attendance, outpatient and equipment entries to the operational-data store (D2). Process 3.0 reads features from D2, applies the ML models and saves predictions, anomalies and ranking scores in the analytics-results store (D3). Process 4.0 generates alerts, KPI reports and decision-support items for the DDHS Admin, who reviews and approves or dismisses them. Stored data supports trend analysis and model retraining.

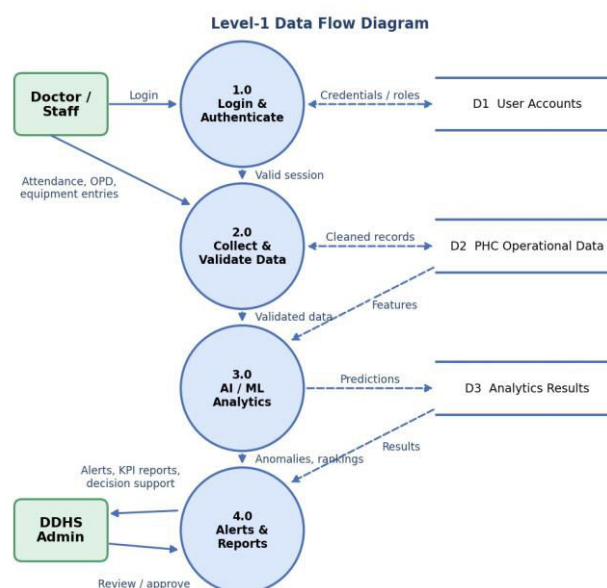


Fig.2. Level-1 data-flow diagram showing processes 1.0–4.0 and data stores D1–D3.

B. Data Collection, Validation and Cleaning

Every submission passes through a validation stage before storage. Required fields must be present; admitted-patient and outpatient (OP) counts must be numeric and non-negative; and attendance status must be one of Present, Absent or On Leave. Free-text fields such as equipment and facility names are normalised by trimming, collapsing repeated spaces and applying consistent case, so that one item or facility is not stored under several spellings. This light NLP-based cleaning is important for the reliability of reports and rankings. Each record is linked to a facility and region, and facilities are classified as PHC, upgraded PHC or Sub-Centre so that comparisons and recommendations are location-aware.

C. Staff Absenteeism Prediction

Absenteeism is cast as binary classification of whether staff member i will be absent on a given day. Features are the facility, role, weekday and month, the number of absences in the last 7 and 30 days, leave days in the last 30 days, and the staff member's distance to the facility. Facility, role and weekday are categorical and are passed to CatBoost [4] directly, avoiding manual one-hot encoding. The model returns the absence probability

$$p_i = P(y_i = 1 | \mathbf{x}_i) (1)$$

and a staff member is flagged when $p_i \geq 0.7$. The predicted absenteeism rate of facility f with n_f staff is

$$A_f = (1 / n_f) \sum_{i \in f} p_i (2)$$

which is displayed to the DDHS office so that replacements can be arranged in advance.

D. Anomaly Detection on Outpatient Load

For the daily OP series x_t of a facility, the standardised deviation is

$$z_t = (x_t - \mu) / \sigma (3)$$

where μ and σ are the mean and standard deviation of the series. A day is flagged as anomalous when $|z_t| > 2.5$. Statistical rules of this type are among the standard techniques catalogued in [5] and are appropriate for short, low-dimensional operational series. Each flagged anomaly is accompanied by a short explanation; when no external language-model key is configured, a rule-based explanation is produced so that the system continues to operate.

E. PHC Performance Score and Ranking

Each facility receives a composite score on a 0–100 scale from three operational indicators: normalised OP load, staff attendance and resource availability,

$$S_f = 100 [w_1 \min(q_f / q_{\max}, 1) + w_2 (1 - a_f) + w_3 r_f] (4)$$

where q_f is the average daily OP count, $q_{\max}$ a normalisation ceiling, a_f the observed absentee rate, r_f the fraction of equipment types that meet their required count, and the weights satisfy $w_1 + w_2 + w_3 = 1$. Facilities are sorted by S_f and the ranking is refreshed as new data arrives.

F. Decision Support and Alert Routing

To suggest resource redistribution, the load per staff member of facility f is computed as

$$L_f = q_f / \max(n_f, 1) (5)$$

and a transfer from the lightest-loaded facility l to the heaviest-loaded facility h is proposed when $L_h > 1.5 L_l$. The suggestion is shown with its reason and Approve and Dismiss options, so the system recommends but never executes a transfer. A routing engine selects recipients by severity and role: a facility-level issue goes to the concerned doctor, whereas critical or district-wide conditions go to the District Health Officer.

G. Implementation

The platform is implemented as four modules: data preprocessing and location information, attendance monitoring, healthcare-service monitoring, and AI alerting and notification. Table I lists the technology stack. The backend exposes REST endpoints for attendance, patient logging, equipment management, absenteeism prediction, ranking and transfer suggestions, and returns JSON to the frontend; Pydantic models validate request bodies automatically.

TABLE I TECHNOLOGY STACK

Layer	Technology / role
Frontend	React.js with Vite, Tailwind CSS, Axios; responsive web and mobile views
Backend	Python, FastAPI [8]; role-based authentication, RESTful APIs
Database	Microsoft SQL Server; users, attendance, patients, equipment, analytics results
ML / analytics	CatBoost classifier, Scikit-learn [6], Pandas, NumPy
Tools	VS Code, Git, Postman; Windows Server or Ubuntu Linux hosting

Attendance module. Doctors and staff record present, absent or on-leave status; the screen confirms the logged status with role and facility, shows daily summary counts, lists the facility attendance log and keeps a 30-day history that feeds the prediction features. Healthcare-service module. Staff log admitted and OP counts, with cumulative totals and a recent-volume chart. The equipment screen stores units present and required for each item, computes the gap and flags the item OK or NEEDED. Alert module. Anomalies, absentee flags, ranking and transfer suggestions are produced by the analytics engine and delivered through the routing engine and the DDHS dashboard.

The hardware requirement is modest: a server with an Intel i5/Ryzen 5-class processor, 8 GB RAM (16 GB preferred) and 256 GB SSD, and client devices with any modern browser, which keeps the solution within reach of district-level IT budgets.

IV. RESULTS

A. Prototype Evaluation

The prototype was populated with data for six facilities (two upgraded PHCs, two PHCs and two Sub-Centres) and exercised end to end from login through data entry, analytics and alerting, using doctor, staff and administrator accounts. Fig. 3 shows the DDHS dashboard, and Table II compares two dashboard snapshots captured as data were logged: S1 from a doctor account (average 42 OPs per facility over seven days, one active anomaly, Ambattur Upgraded PHC ranked first with 67.2/100) and S2 from the administrator account (average 50 OPs, two active anomalies, Perambur PHC ranked first with 72.4/100). No absentees were flagged over the last five days in either snapshot.

The two snapshots show that the ranking is recomputed from live data. Between S1 and S2 the OP count of Perambur PHC rose from 57 to 102, which raised its score from 66.9 to 72.4 and moved it from second to first place, while the other facilities' OP counts did not change. The staff view of the same facility is consistent with S1: its patient screen shows 57 cumulative OPs and 10 cumulative admissions.

TABLE II RANKING SNAPSHOTS S1 AND S2

Facility	Score S1	OPs S1	Score S2	OPs S2
Perambur PHC	66.9	57	72.4	102
Ambattur Upgraded PHC	67.2	76	67.2	76
Kolathur Sub Centre	59.9	30	59.9	30
Tondiarpet PHC	58.8	30	58.8	30
Villivakkam Sub Centre	58.8	30	58.8	30
Anna Nagar Upgraded PHC	58.8	30	55.3	30

The OP-level chart makes the imbalance across the network immediately visible: in S2 two facilities handle 102 and 76 OPs while the remaining four handle 30. Anna Nagar Upgraded PHC's score fell from 58.8 to 55.3 although its OP count stayed at 30, which shows that the score responds to indicators other than OP load. The decision-support panel recommends moving Dr. V. Iyer from Anna Nagar Upgraded PHC to Kolathur Sub Centre, with the reason that the destination handles a much higher OP load per staff member. Because the reasoning is displayed, the officer can verify it before pressing Approve or Dismiss.

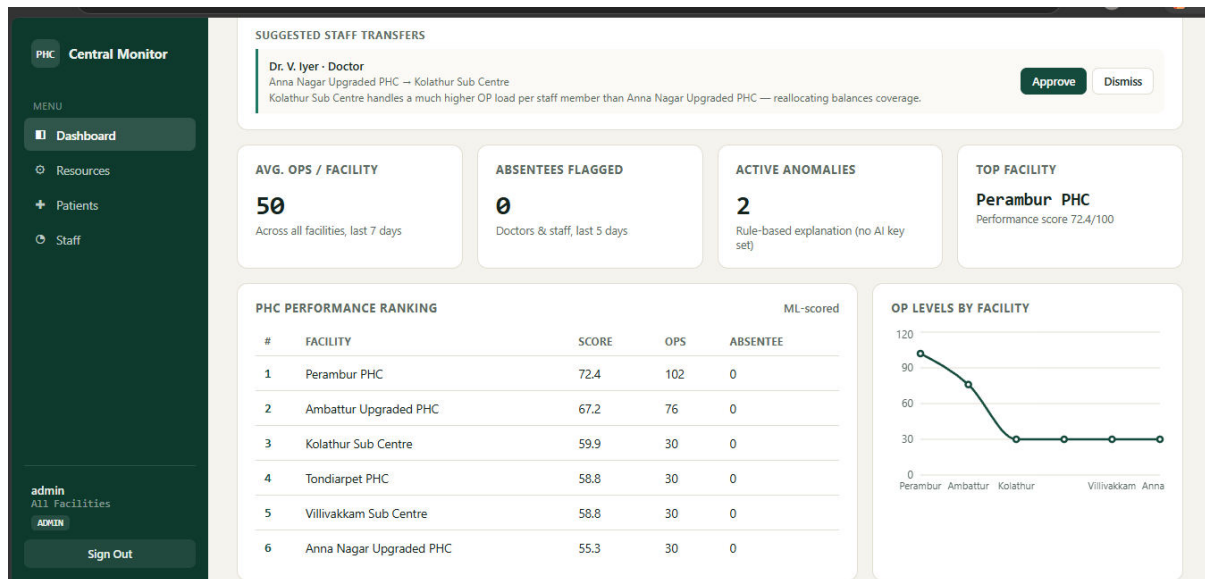


Fig.3. DDHS dashboard of the prototype: suggested staff transfer, summary indicators, ML-scored PHC performance ranking and OP levels by facility.

Attendance and equipment. On the attendance screen of Perambur PHC, the facility log lists three staff members (one staff member and two doctors), all marked present, giving 3 present, 0 absent and 0 on leave for the day. The equipment screen of the same facility tracks seven equipment types with 23 units in total; four types are below the required count and three are fully stocked, which is the condition that triggers a restocking alert. Table III lists the first four items, where the gap is computed as required minus present.

TABLE III EQUIPMENT INVENTORY OF PERAMBUR PHC (FIRST FOUR ITEMS)

Equipment	Present	Required	Gap	Status
Digital BP Monitor	6	8	2	NEEDED
Autoclave Sterilizer	2	2	0	OK
ECG Machine	1	2	1	NEEDED
Oxygen Concentrator	4	4	0	OK

B. System Testing

Testing covered four levels. Unit tests exercised authentication for each role, validation and cleaning functions with malformed input, derived equipment gap and status, probability bounds of the absenteeism predictor, and the scoring and anomaly functions on known inputs. Functional tests followed the workflow from login to alerting; representative cases are listed in Table IV. Integration tests verified the path from the React interface through FastAPI to SQL Server, from stored data to the analytics engine and dashboards, and from critical conditions to routed notifications, including graceful handling of error responses. Performance tests simulated several facilities submitting data concurrently and examined API response time, dashboard refresh time and database query speed for a large number of stored records.

TABLE IV REPRESENTATIVE FUNCTIONAL TEST CASES

Test case	Expected behaviour
Invalid credentials	Access denied with an error message
Submit attendance	Status stored; present count updated
Negative patient count	Entry rejected with validation message
Add equipment entry	Gap, days left and status calculated and shown
View ranking	Facilities listed in descending score order
Unequal OP load per staff	Transfer suggestion with reason and Approve/Dismiss
Staff opens admin-only function	Access not permitted

C. Discussion

Compared with the manual process described in Section I, the platform removes the consolidation delay, enforces consistent data formats at entry, gives the DDHS a single comparable view of all facilities, and turns the data into prioritised, explained actions. Targeted escalation reduces the chance that a critical alert is lost among routine notifications.

The study has limitations that should be stated plainly. First, the evaluation in this paper is functional: it demonstrates that the pipeline operates correctly on sample data, but it does not report predictive accuracy metrics such as AUC or F1 for the absenteeism model, because these require a labelled attendance history from real facilities. Second, the ranking weights, the 0.7 flagging threshold, the 2.5 anomaly limit and the 1.5 load ratio are configurable heuristics that need calibration against field data. Third, poorly tuned thresholds can cause alert fatigue, and delayed or inaccurate input can cause false triggers. Fourth, handling of patient-level records in a deployment would need to follow applicable health-data privacy rules. A pilot across a district, with retrospective validation of predictions against recorded attendance, is the logical next step.

V. CONCLUSION

This paper presented a Centralised PHC Monitoring System that unifies attendance, patient-service and equipment data from PHCs and Sub-Centres, applies CatBoost-based absenteeism prediction, z-score anomaly detection, composite performance ranking and rule-based transfer recommendations, and delivers severity-aware alerts to the appropriate officer. A React, FastAPI and SQL Server implementation was validated on a six-facility prototype and through multi-level testing. The design keeps humans in control of every decision while giving them timely, explained evidence.

Future work includes pilot deployment with quantitative evaluation of the absenteeism model against alternatives such as Random Forest and gradient-boosting baselines, outbreak prediction from service data, GIS-based district monitoring, predictive resource allocation, telemedicine and IoT integration, state-level data consolidation, and blockchain-based record integrity.

ACKNOWLEDGMENT

The authors thank the Department of Computer Science and Engineering, K.L.N. College of Engineering, for the facilities and guidance provided during this work.

REFERENCES

- [1]F. Khalique, S. A. Khan, and I. Nosheen, “PHF: A framework for public health monitoring, analytics and research,” IEEE Access, vol. 7, pp. 101309–101326, 2019.
- [2]A. Rista, L. Ahmedi, and Z. Kastrati, “Predicting and analyzing absenteeism at workplace using machine learning algorithms,” in Proc. 43rd Int. Conv. Inf., Commun. Electron. Technol. (MIPRO), Sep. 2020.
- [3]P. Kushwaha et al., “Internet of Things based health monitoring system with AI powered disease prediction,” IEEE, 2024.
- [4]L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, “CatBoost: Unbiased boosting with categorical features,” in Adv. Neural Inf. Process. Syst. (NeurIPS), vol. 31, 2018.
- [5]V. Chandola, A. Banerjee, and V. Kumar, “Anomaly detection: A survey,” ACM Comput. Surv., vol. 41, no. 3, Art. 15, Jul. 2009.
- [6]F. Pedregosa et al., “Scikit-learn: Machine learning in Python,” J. Mach. Learn. Res., vol. 12, pp. 2825–2830, 2011.
- [7]A. Rajkomar, J. Dean, and I. Kohane, “Machine learning in medicine,” N. Engl. J. Med., vol. 380, no. 14, pp. 1347–1358, 2019.
- [8]S. Ramírez, “FastAPI framework documentation.” [Online]. Available: <https://fastapi.tiangolo.com>

BIOGRAPHY

Dr. A. Askarunisa, M.E., Ph.D., is the Head of the Department of Computer Science and Engineering, K.L.N. College of Engineering, Tamil Nadu, India.

Maha Sree P is an undergraduate student in the Department of Computer Science and Engineering, K.L.N. College of Engineering, Tamil Nadu, India.

Mathu Mathi T S is an undergraduate student in the Department of Computer Science and Engineering, K.L.N. College of Engineering, Tamil Nadu, India.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details