Kaur and Virk *Discover Artificial Intelligence* (2025) 5:346
https://doi.org/10.1007/s44163-025-00610-4

Discover Artificial Intelligence

RESEARCH Open Access

An IoT-Driven hybrid AI model for health monitoring of cows

Devinder Kaur[1*] and Amandeep Kaur Virk[1]

*Correspondence:
Devinder Kaur
devinder.mgc@gmail.com
[1]Department of Computer Science, Sri Guru Granth Sahib World University, Fatehgarh Sahib, Punjab, India

Abstract
Livestock diseases continue to pose significant challenges for smallholder dairy farmers, particularly in regions with limited access to veterinary services and real-time health monitoring infrastructure. To address this, the present study proposes a cost-effective IoT-enabled framework for early disease detection in dairy cattle, targeting smallholder farmers in resource-limited regions. A custom smart collar was developed to monitor body temperature, pulse rate, and activity levels in 150 cows across seven districts of Punjab. To analyse the collected data, a novel hybrid model, SM-GBoost-LSTM, was implemented, combining Gradient Boosting (GBoost) for structured features, Long Short-Term Memory (LSTM) networks for temporal patterns, and SMOTE for handling class imbalance. The model achieved high performance, attaining 93.56% accuracy, 91.42% precision, 77.77% recall, and 84.02% F1-score. The system also integrates real-time cloud analytics and a bilingual mobile application (English/Punjabi) to deliver timely health alerts. Overall, the proposed framework provides a scalable, field-validated, and affordable AI-driven solution for improving livestock healthcare in smallholder dairy farming.

Article highlights

- Proposes a cost-effective IoT-enabled smart collar design for continuous monitoring of key physiological parameters in dairy cattle.
- Introduces SM-BoostLSTM, a hybrid model integrating SMOTE (for class imbalance), Gradient Boosting (for structured feature learning), and LSTM (for temporal pattern recognition) in livestock health data.
- Demonstrates the efficacy of the framework through field-validated analysis and a bilingual (English–Punjabi) mobile app envisioned for delivering timely health alerts to farmers.

Keywords Internet of things (IoT), Machine learning classifiers, Cow health monitoring, Precision livestock farming (PLF)

1 Introduction

The dairy sector in India plays a vital role in the national economy, employing over 80 million people, primarily small and marginal farmers. It contributes about 4.11% to the national GDP and 25.6% to the agricultural GDP [1]. However, livestock health monitoring continues to face significant challenges. Traditional methods are largely dependent on manual observation, which lacks the ability to provide real-time disease detection. This often leads to delayed diagnosis, higher animal mortality, and considerable economic losses. Passive data logging, absence of intelligent analytics, and lack of automated alert systems further exacerbate the problem, limiting timely veterinary interventions and efficient disease management.

In this context, the adoption of advanced technologies is essential to ensure productivity, profitability, and animal welfare. The emergence of Precision Livestock Farming (PLF) has introduced a paradigm shift in livestock health management by enabling continuous and automated monitoring systems. Integration of the Internet of Things (IoT) with machine learning has been instrumental in transforming cattle surveillance through real-time assessment and data-driven decision-making.

Several studies have highlighted the importance of IoT-enabled wearable systems in health monitoring. For instance, motion sensors and body-worn devices can detect early symptoms of lameness, mastitis, or infections by analysing physical activity, body temperature, and heart rate [2–4]. These tools have proven effective in lowering treatment costs and reducing the dependence on labour-intensive practices. Furthermore, IoT-based technologies have shown promising results in controlling widespread livestock diseases such as Foot and Mouth Disease (FMD) and mastitis [5]. Recent advancements in Artificial Intelligence (AI)-integrated IoT systems for medical and industrial applications underscore the potential of such technologies to reshape health and safety frameworks across domains, including livestock. Studies such as BDLT-IoMT and ORAN-B5G demonstrate the effective integration of blockchain and machine learning to ensure robust, scalable, and secure IoT-based monitoring environments [6, 7].

Despite these advancements, many existing livestock monitoring solutions still fall short in areas like predictive analytics, continuous health evaluation, and integration with intelligent alert systems. Technologies such as RFID, though effective for identification, are not typically connected to machine learning infrastructures, reducing their capacity to support health inference. Moreover, the absence of integrated alert mechanisms for farmers significantly elevate the risk of delayed intervention, thereby accelerating the decline in animal health conditions.

To address these challenges, this study presents a novel, low-cost, field-validated IoT smart collar system integrated with cloud computing, a hybrid SM-GBoost-LSTM model, and a bilingual (English/Punjabi) mobile application for smallholder dairy farmers in Punjab. The system captures essential physiological parameters, transmits data to the cloud for analysis, and delivers timely health alerts through the mobile application. Multiple machine learning approaches were evaluated on the dataset, and the best-performing hybrid model was selected to ensure accurate, real-time disease prediction. This is among the first studies to implement such an end-to-end system, combining affordability, field validation, and intelligent decision support for smallholder farmers.

2 Related works

Numerous studies have investigated technology-driven solutions for cattle health monitoring, focusing on automated sensing, early disease detection, and data-informed livestock management. This section reviews key approaches, including sensor technologies, machine learning methods, and evaluation strategies aimed at improving predictive accuracy in modern farming systems.

Study [8] explored the potential of Mid-Infrared (MIR) spectrometry for detecting health issues in grazing dairy cows using ML algorithms. The dataset comprised 1,909 milk samples from Holstein–Friesian, Jersey, and crossbreed cows, analyzed using classifiers such as Random Forest (RF), Support Vector Machine (SVM), Neural Networks (NN), Convolutional Neural Networks (CNN), and ensemble models. Among these, NN exhibited the highest performance, showcasing its potential for early disease detection.

Another study [9] investigated the role of ML in detecting behavioural anomalies in dairy cows affected by Sub Acute Ruminal Acidosis (SARA). Data were gathered from 14 cows diagnosed with SARA and 14 healthy controls using a Real-Time Locating System (RTLS) and ruminal pH sensors. Various ML models—including K-Nearest Neighbors Regression (KNNR), Decision Tree Regression (DTR), Multi-Layer Perceptron (MLP), and Long Short-Term Memory (LSTM)—were evaluated, with KNNR emerging as the best performer. However, the study emphasized the need for further refinement to enhance its practical applicability.

In the domain of calving prediction, ML was applied to behavioural data from activity, lying, and rumination monitors [10]. Data from 53 Holstein cows were collected using HR Tags and IceQube sensors, and multiple ML models—including Random Forest (RF), Linear Discriminant Analysis (LDA), and Neural Networks (NN)—were assessed. The NN model, which combined data from both sensor types, achieved superior accuracy, reinforcing the effectiveness of ML in predicting calving events.

The study conducted by [11] focused on predicting the metabolic status of dairy cows during early lactation. The metabolic conditions of 334 cows were classified into three categories—good, average, and poor—based on plasma metabolite and hormone levels. Random Forest (RF) and Support Vector Machines (SVM) were the top-performing classifiers, with RF excelling in sensitivity and specificity. The key predictive parameters included milk yield, fat yield, protein percentage, and lactose yield, demonstrating the capability of machine learning for non-invasive metabolic monitoring.

For Clinical Mastitis (CM) detection, Decision-Tree induction was applied to sensor data from Automatic Milking Systems (AMS) [12]. Over 2.5 years, data were collected from nine Dutch dairy herds, capturing parameters like electrical conductivity, milk colour, and yield at the Quarter Milking (QM) level. The decision-tree model significantly enhanced CM detection while reducing false-positive alerts, making it a promising tool for automated health monitoring.

A separate study [13] explored the potential of machine learning in predicting subclinical mastitis, utilizing 364,249 milking instances recorded by an automated in-line monitoring system. Several ML models—including Deep Learning (DL), Naïve Bayes (NB), Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), and Gradient-Boosted Trees (GBT)—were tested, with GBT demonstrating the highest predictive accuracy.

A comparative study [14] analyzed the effectiveness of K-Nearest Neighbors (KNN) and SVM for detecting clinical mastitis using Internet of Animal Health Things (IoAHT) sensor data. Parameters such as milk conductivity, temperature, and cow activity were examined, with SVM outperforming KNN. The findings reinforce the potential of IoAHT to enhance mastitis detection and herd health management.

In the area of lameness prediction, ML techniques were applied to milk production and conformation traits from 2,535 cows across nine Australian dairy herds. Various models, including Naïve Bayes Random Forest, Multilayer Perceptron (MLP), and Logistic Regression, were compared, with Naïve Bayes (NB) emerging as the top performer [15]. However, the study acknowledged that additional environmental factors and a larger dataset would be necessary to improve accuracy.

Research work [16] utilized ML to rank environmental heat stressors in dairy cows based on their impact on physiological responses such as Respiration Rate (RR), Skin Temperature (ST), and Vaginal Temperature (VT). Data from 19 Holstein–Friesian cows were examined, incorporating variables like Air Temperature (AT), Relative Humidity (RH), Solar Radiation (SR), and wind speed. RF and NN provided the most accurate predictions, identifying air temperature as the most significant contributor to heat stress.

ML was also evaluated for dystocia (difficult calving) detection in dairy cows and heifers using 1.2 million calving records from Polish Holstein–Friesian herds. Factors such as calving age, gestation length, and prior calving difficulty were analyzed, with RF and Boosted Trees (BT) being tested [17]. Although BT exhibited higher sensitivity, both models faced challenges due to high false alarm rates, limiting their real-world applicability.

For Lumpy Skin Disease (LSD) prediction, ML models were applied to meteorological and geospatial data comprising 24,803 instances with 19 climate-related features. Ten classifiers, including RF and Light Gradient Boosting Machine (LGBM), were tested, with RF and LGBM outperforming others [18]. The study demonstrated the viability of machine learning for early LSD detection, which could be instrumental in preventive vaccination strategies.

A comprehensive ML-based approach was used to detect common disorders in dairy cows, such as clinical mastitis, subclinical ketosis, lameness, and metritis [19]. The dataset included 131 sick cows and 149 healthy cows, incorporating features like milk yield, activity, rumination time, and milk electrical conductivity. Among eight tested ML models, Recursive Partitioning and Regression Trees (RPART) delivered the best classification performance.

An IoT-driven cattle health monitoring system was developed, integrating Wireless Sensor Networks (WSNs) and Artificial Neural Networks (ANN) for real-time disease prediction [20]. Data were collected from non-invasive body-area sensors, measuring temperature, heart rate, and jaw movement. The system employed Raspberry Pi for data processing and Message Queuing Telemetry Transport (MQTT) protocol for secure transmission. With its web-based dashboard and security enhancements, this system demonstrated high accuracy in cattle disease prediction, setting a new benchmark for IoT-based livestock monitoring.

Recent advances in deep learning, such as convolutional neural networks (CNN), have shown promising results in disease detection from medical images [21]. highlighted the potential of advanced machine learning techniques to improve diagnostic accuracy, which aligns with the hybrid modelling approach used in this study.

These studies collectively underscore the transformative impact of ML and IoT in dairy farming, enabling early disease detection, precision monitoring, and proactive herd management. Table 1 provides a description of existing studies on cattle health monitoring using Machine Learning (ML) and the Internet of Things (IoT). This comparison highlights the strengths and limitations of various approaches.

Table 1 Comparison of related works on ML and IoT-based cattle health monitoring

References	Technique used	Hardware device	Disease detected	ML classifier used	Results
[8]	Mid-Infrared Spectrometry	CombiFoss™ 7 (Foss Electric)	General health issues (lameness, mastitis, reproductive disorders)	RF, SVM, NN, CNN, Ensemble Models	NN: Sensitivity 61.74%, Specificity 97%, PPV 59.99%, AUC 79.37%
[9]	Real-Time Locating System (RTLS) & ruminal pH sensors	CowView (GEA Farm Technologies), eCow ruminal pH bolus	Subacute Ruminal Acidosis (SARA)	KNNR, DTR, MLP, LSTM	KNNR: 83% sensitivity, 66% false positives
[10]	Behavioural monitoring (activity, rumination, lying)	HR Tag (SCR Engineers Ltd.), IceQube (IceRobotics Ltd.)	Calving prediction	RF, LDA, NN	NN: 100% sensitivity, 86.8% specificity (daily); 82.8% sensitivity, 80.4% specificity (8-h window)
[12]	Sensor-based detection (conductivity, colour, yield)	Lely Astronaut AMS (A2/A3)	Clinical Mastitis	Decision Tree	99% specificity: 40% sensitivity; 97.9% specificity: 57.1% sensitivity; Severe CM detection: 64.3%
[13]	Automated milking data analysis	In-line monitoring system (CellSense®)	Sub-clinical Mastitis	GBT, DL, RF, NB, LR, DT	GBT: 84.9% accuracy, DL: high AUC (0.826), RF: lowest accuracy (82.3%)
[16]	Environmental heat stress ranking	Weather station (WS-16, Novalynx Corp.), temperature sensors	Heat stress in dairy cows	RF, NN, GBM, Penalized Regression	RF: RMSE 9.695 (RR), NN: RMSE 0.434 °C (VT), RF: RMSE 0.334 °C (ST)
[19]	Automatic monitoring & milking data	HR-Tag (SCR Engineers Ltd.), DataFlow (SCR Engineers Ltd.)	Clinical Mastitis, Subclinical Ketosis, Lameness, Metritis	RPART, SVM, RF, Extreme Gradient Boosting, Adaboost, Naïve Bayes, KNN, Logistic Regression	RPART: 81.58% accuracy, 92.86% precision, AUC 0.908

3 Methodology

This section presents the systematic approach adopted to develop and validate the cattle health prediction system, emphasizing the integration of data acquisition, analytical modelling, and user interaction components.

3.1 Data collection and preprocessing

A comprehensive dataset was collected from 150 cows across different districts of Punjab, India, including Fatehgarh Sahib, Ludhiana, Nawan Shahr, Patiala, Sangrur, Ropar, and Hoshiarpur over a period of two months per cow using an innovative custom designed "Smart Neck Collar" [22] depicted in Fig. 1. The dataset includes vital health parameters such as temperature, pulse rate, and activity level, recorded at 15-min intervals over 10 h per day, ensuring a rich and detailed dataset.

Figure 1 illustrates the custom-designed wearable smart neck collar used in the proposed livestock health monitoring system. The device features a durable adjustable strap for comfortable attachment around the neck of the cow and an enclosed electronics module housing key sensing and processing components. The embedded hardware includes temperature, pulse, and motion sensors, connected to an ESP32 microcontroller for on-device preprocessing of physiological and behavioural data. The enclosure

Fig. 1 Custom designed "Smart Neck Collar" for cow health status prediction.
[Reprinted from Kaur and Virk (2025), *Discover Internet of Things*, 5(1):12]

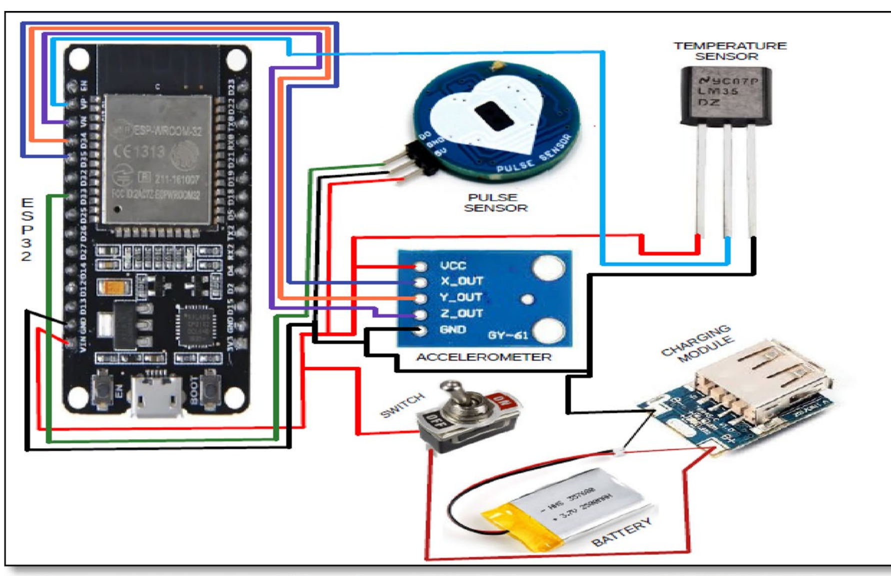

Fig. 2 Circuit diagram of the "Smart Neck Collar"

is lightweight, weather-resistant, and designed for sustained operation in outdoor farm environments. A detailed hardware design description is available in [22].

Figure 2 depicts the circuit diagram of the device.

For this study, a single daily timestamp was recorded for each cow to capture essential health status metrics. This approach balances data resolution with practical constraints such as storage, processing load, and interpretability for smallholder farmers. By focusing on daily summaries of physiological and behavioural parameters, the system effectively identifies meaningful health trends and potential disease occurrences without generating excessive or redundant data.

3.1.1 Hardware components used in "Smart Neck Collar"

The following hardware components have been used in the development of advanced Smart Neck Collar designed for real-time monitoring of cow health status. These components collectively form the technological framework that enables continuous data collection and processing.

The IoT-enabled Smart Neck Collar has been designed to provide continuous, real-time monitoring of key physiological parameters in dairy cattle while remaining cost-effective for smallholder farmers. The device integrates affordable, high-performance components (Table 2), including a pulse transducer for heart rate monitoring, a triaxial accelerometer for movement and behavior tracking, a waterproof temperature sensor for body temperature measurement, and a compact lithium-ion battery with an efficient power booster for stable operation. The sensors have been strategically positioned around the cow's neck to ensure accurate data acquisition while minimizing discomfort.

The lithium-ion battery provides approximately 12 h of continuous operation under normal conditions, and battery status is monitored via the power booster module. All sensors were calibrated prior to deployment: the pulse sensor was validated against standard veterinary heart-rate measurements, the accelerometer was checked with controlled motion tests, and the temperature sensor was calibrated using a reference thermometer. Each component was selected based on a balance of accuracy, durability, and affordability, ensuring that the collar delivers reliable data without incurring high costs. This careful selection and calibration of hardware makes the system practical for deployment in resource-limited settings.

3.1.2 Data preprocessing

To prepare the dataset for analysis, multiple pre-processing steps were applied:

1. *Handling Missing Values*: Missing data points were replaced with the mean of the respective column to maintain data integrity.

Table 2 Key hardware components of the IoT-enabled Smart Neck Collar

Component	Model/Type	Key features	Role in system
Pulse transducer	Pulse Sensor Amped	Open-source optical heart-rate sensor; based on photoplethysmography (PPG); integrates IR LED & photodetector on compact PCB; converts variations in reflected light intensity into analog voltage signals	Detects pulse rate by monitoring blood volume changes during cardiac cycles
Triaxial accelerometer	GY-61	Measures acceleration along X, Y, Z axes; three sensing elements aligned at right angles; enables accurate motion/vibration detection	Captures cattle movement and behaviour patterns for health monitoring
Temperature sensor	PT100-S Waterproof Stainless-Steel Probe (FTARP05 series)	4 mm diameter, 30 mm length stainless-steel probe; silver-plated copper cables; waterproof and highly durable; precise temperature readings	Monitors body temperature for disease detection in cattle
Lithium-ion battery	Orange 360mAh 1S 30C/60C LiPo Battery Pack	High performance, low internal resistance; robust discharge leads; cost-efficient and reliable	Provides sustained power supply for device operation
Power booster	5 V Step-up power module (134N3P)	Supports lithium battery charging; efficient voltage boost; USB connectivity; LED display for real-time voltage monitoring	Ensures stable power supply and charging protection for IoT device

2. *Anomaly Detection and Removal*: The Isolation Forest algorithm was used to detect and remove anomalies in pulse rate, temperature, and activity levels. Standard Scaler was applied to normalize features, ensuring a uniform range.
3. *Data Cleaning*: Outliers flagged as anomalies (– 1) were removed, leaving a high-quality dataset suitable for machine learning analysis.

3.2 Feature extraction

Feature extraction was performed to transform raw sensor data into meaningful inputs for machine learning models.

1. The dataset was aggregated by date and cow ID, ensuring each cow had a single representative record per day.
2. Key features (pulse rate, temperature, and activity level) were summarized using their daily mean values to provide a concise representation of the cow's health trends.
3. Health status labels were determined based on daily physiological and behavioural measurements. Thresholds for body temperature, pulse rate, and activity levels were defined according to veterinary guidelines and relevant literature. These thresholds and labelling rules were validated in consultation with the staff of the Government Veterinary Hospital, Amloh, Punjab, who provided practical insights and confirmed normal and abnormal ranges.

A cow was labelled "unhealthy" if any recorded measurement exceeded its respective threshold at any point during the day, while a cow was labelled "healthy" only if all parameters remained within the normal ranges. This approach ensures that the dataset accurately reflects the daily health status of each cow, minimizes bias toward "healthy," and supports reliable model training and evaluation.

3.3 Model training and evaluation

After preprocessing the dataset, traditional machine learning classifiers were employed to analyse health patterns and predict the health status of cows. Models such as Random Forest, Decision Trees, Support Vector Machines (SVM), Logistic Regression, K-Nearest Neighbor (KNN), Neural Networks, and Naïve Bayes were selected due to their proven reliability on structured datasets, ease of implementation, and interpretability. These methods provided useful insights but were limited in their ability to detect all positive cases, particularly in scenarios where sensitivity to sick animals is essential.

To address these limitations, ensemble learning approaches were explored in the next phase of experimentation. Gradient Boosting and XGBoost were chosen for their ability to handle complex feature interactions, reduce bias–variance trade-offs, and deliver robust performance on structured datasets. These models demonstrated stronger predictive potential compared to traditional classifiers, making them more suitable for integration into the system.

Recognizing that cattle health data is inherently sequential, with physiological and behavioural indicators varying over time, the study then investigated a standalone Long Short-Term Memory (LSTM) model. LSTM networks are well-suited for capturing temporal dependencies in time-series data, making them a natural fit for livestock health monitoring. However, relying solely on LSTM was insufficient, which motivated the development of hybrid models.

In the final stage, hybrid frameworks were designed by combining LSTM with ensemble learners. This integration allowed the models to leverage both temporal sensitivity and the strong generalization capabilities of ensemble techniques. To further address the issue of class imbalance, data augmentation methods such as SMOTE were incorporated. These hybrid approaches proved to be the most effective, offering a balanced solution capable of detecting subtle health abnormalities while reducing misclassification.

3.4 Performance evaluation of tested models

The trained models were assessed based on the following performance metrics:

1. *Accuracy*—The proportion of correctly classified instances.
2. *Precision*—The ability to correctly predict diseased cows.
3. *Recall* —The ability of a model to detect all sick cows.
4. *F1 Score*—The harmonic means of precision and recall, particularly useful for imbalanced datasets.

By analysing these metrics, the best-performing model has been selected for real-time disease prediction, ensuring high accuracy and reliability.

Table 3 presents the evaluation metrics (Accuracy, Precision, Recall, F1 Score) for traditional ML models including Random Forest, Logistic Regression, Support Vector Machines, Decision Trees, Naïve Bayes, K-Nearest Neighbors and Neural Networks.

3.4.1 Key observations

- *Random Forest* provided good accuracy (81.78%) but relatively low recall (50.50%), suggesting that it missed more true positives.
- *Logistic Regression and SVM* demonstrated moderate performance, with Logistic Regression achieving the second-highest recall (76.51%), making it more effective for detecting positives but at the expense of precision.
- *Decision Tree* followed closely with 84.08% accuracy and a balanced precision (71.33%) and recall (66.22%), resulting in a strong F1 score of 68.68%.
- *Naive Bayes* achieved the highest recall (81.88%) but suffered from low precision and accuracy, indicating frequent false positives.
- *KNN* performed well in terms of accuracy (83.49%) and precision (69.05%), but its lower recall (51.79%) reduced the overall F1 score.
- *Neural Network* achieved the highest accuracy (86.20%) and best precision (73.80%), indicating strong overall performance with fewer false positives.

Figure 3 illustrates the performance of traditional machine learning models using graphical representation.

Table 3 Performance metrics of traditional machine learning models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
Random forest	81.78	64.22	50.50	56.54
Logistic regression	77.65	55.52	76.51	64.35
Support vector machine	79.36	58.52	74.50	65.55
Decision tree	84.08	71.33	66.22	68.68
Naive bayes	66.75	43.11	81.88	56.48
K-nearest neighbors	83.49	69.05	51.79	59.18
Neural networks	86.20	73.80	62.50	67.68

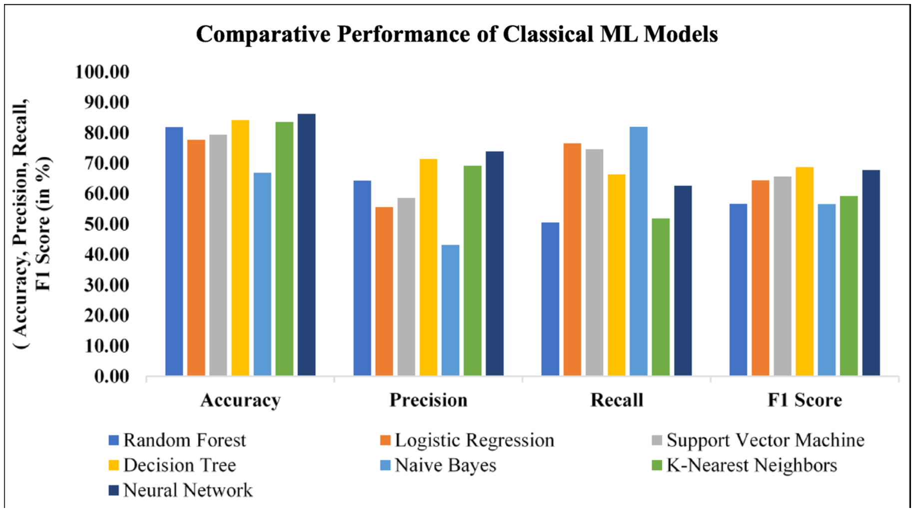

Fig. 3 Graphical representation of performance metrics for classical ML models

Table 4 Performance metrics of boosting based models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
Extreme gradient boosting	86.93	82.77	63.65	71.96
Gradient boosting	91.41	90.00	70.71	79.20

The evaluation of traditional classifiers demonstrated moderate performance across all metrics, with Neural Networks and Decision Trees achieving relatively better results compared to other models. However, a closer look at the precision–recall trade-offs revealed that none of the classifiers provided consistently high performance across all evaluation metrics.

Given these limitations, the results were not entirely satisfactory, prompting the exploration of ensemble learning approaches. Ensemble models are regarded as an approach to integrate the strength of multiple classifiers, address bias–variance trade-offs, and thereby enhance both the robustness and the predictive accuracy of the system.

After evaluating traditional classifiers, the next phase of experimentation focused on ensemble learning approaches. Specifically, Gradient Boosting and XGBoost were selected, as both are well-established techniques known for their ability to reduce bias–variance trade-offs and enhance predictive accuracy, particularly in structured datasets such as the one used in this study. Ensemble boosting methods Gradient Boosting and Extreme Gradient Boosting (XGBoost) showed noticeable improvements, particularly in Accuracy and F1-score, as shown in Table 4.

These results indicate that standard Gradient Boosting offers more consistent detection of health abnormalities in cows, making it the preferable choice for integration into the predictive module. Figure 4 presents the performance metrics of ensemble models, showing their accuracy, precision, recall, and F1 score for comparative evaluation.

Given that cow health patterns are influenced by time-based variations in physiological and behavioural data, the performance of a standalone Long Short-Term Memory (LSTM) model has been evaluated. LSTM networks are specifically designed to capture temporal dependencies in sequential data, making them well-suited for analysing time-series health indicators. The results of the LSTM model are presented in Table 5.

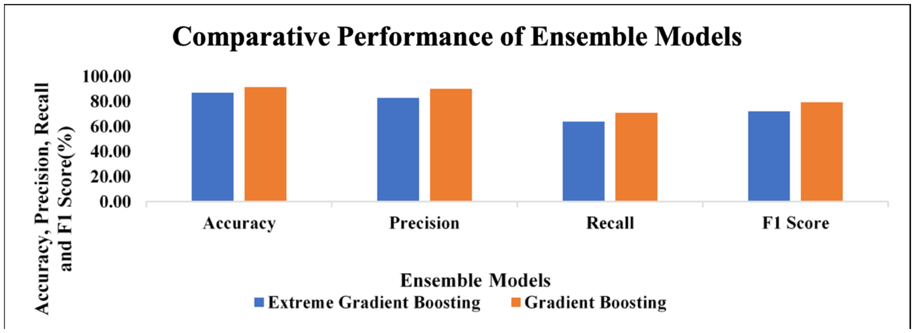

Fig. 4 Graphical representation of performance metrics for Ensemble models

Table 5 Performance of LSTM model for cow health prediction

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
LSTM	82.44	72.27	54.12	61.89

Table 6 Performance metrics of LSTM-enhanced hybrid models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
Random Forest with LSTM	86.59	93.99	52.46	67.34
Neural Networks with LSTM	89.00	78.04	72.91	75.39
Extreme Gradient Boost with LSTM	85.29	83.10	55.46	66.52
Gradient Boost with LSTM	91.69	89.34	72.70	80.17
SM-BoostLSTM (Gradient Boost with LSTM with SMOTE)	93.56	91.42	77.77	84.02

The standalone LSTM achieved an accuracy of 82.44%, with a precision of 72.27% and recall of 54.12%. The relatively lower recall indicates that while the model can make reliable predictions of unhealthy cows, it fails to capture all positive cases, potentially missing early signs of illness.

To overcome the limitations of standalone LSTM, it was combined with machine learning models such as XGBoost, Gradient Boosting and Random Forest. While these ensemble learners are effective in handling structured data and modelling nonlinear relationships, they do not inherently capture sequential dependencies. By integrating LSTM with ensemble models, the hybrid frameworks combine temporal sensitivity with strong predictive generalization, leading to improved classification performance. The performance of the hybrid models is presented in Table 6.

3.4.2 Key observations

- *Random Forest with LSTM* achieved high precision (93.99%) but low recall (52.46%), indicating that while its positive predictions were accurate, it missed a large portion of true positives.
- *Neural Networks with LSTM* improved overall balance, reaching 89.00% accuracy and a much higher recall (72.91%) compared to Random Forest, which resulted in a stronger F1 score (75.39%).

- *Extreme Gradient Boost with LSTM* showed moderate performance (85.29% accuracy), with balanced precision (83.10%) and recall (55.46%), but its F1 score (66.52%) was lower than that of Neural Networks.
- *Gradient Boost with LSTM* significantly improved performance, achieving 91.69% accuracy and a strong F1 score (80.17%), demonstrating a better trade-off between precision (89.34%) and recall (72.70%).
- *SM-BoostLSTM (Gradient Boost + SMOTE)* delivered the best results overall, with the highest accuracy (93.56%), strong precision (91.42%), and improved recall (77.77%), leading to the best F1 score (84.02%). The application of SMOTE effectively enhanced minority class detection without sacrificing much precision.

Figure 5 presents the comparative performance of hybrid machine learning models. The comparison highlights how integrating boosting algorithms with LSTM improved overall performance as compared to individual models.

The evaluation of various machine learning models revealed that ensemble techniques, such as Gradient Boosting (GBoost), Extreme Gradient Boosting, and Random Forest, demonstrated superior performance when integrated with Long Short-Term Memory (LSTM) for predicting cow health status.

Among these, SM-BoostLSTM emerged as the most effective model, offering an optimal balance between precision and recall, which is crucial for early disease detection while minimizing false positives. The incorporation of SMOTE (Synthetic Minority Over-sampling Technique) significantly improved model performance by addressing class imbalance, leading to higher recall without a substantial drop in precision.

3.5 Development of mobile application

To enhance the accessibility and usability of the proposed cattle health monitoring system, a mobile application has been developed to provide real-time alerts to farmers.

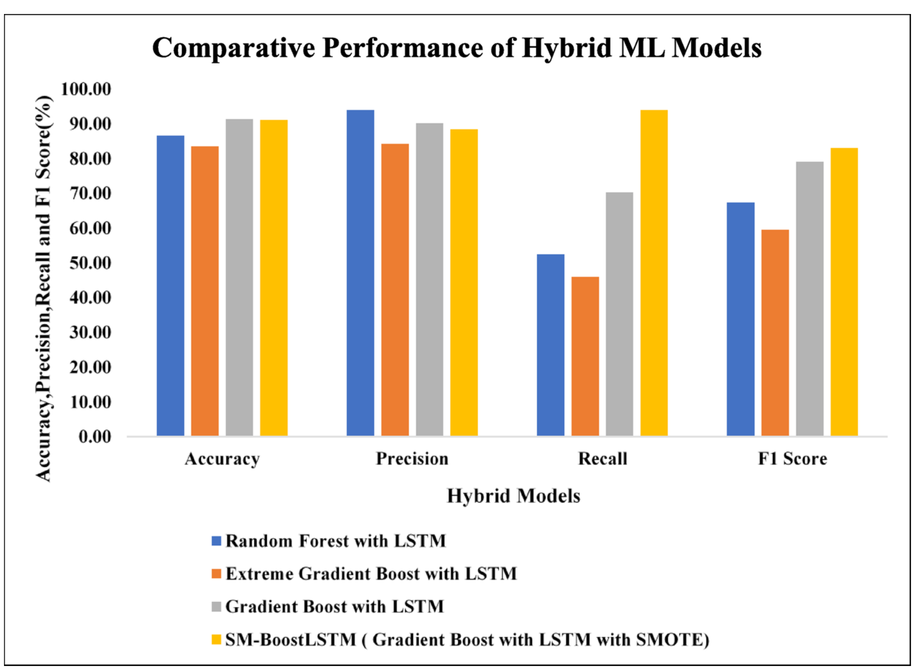

Fig. 5 Comparative performance of hybrid machine learning models

The application has been designed to notify farmers about any critical health conditions detected in cows based on real-time data analysis. The application is integrated with the cloud-based system that receives and processes health parameters from the custom-designed hardware device. Once the SM-BoostLSTM model detects any abnormalities or potential health risks, the application sends instant mobile notifications to the farmer.

The application offers a user-friendly interface designed to accommodate farmers with limited technical expertise. Furthermore, it supports two languages (English and Punjabi) to address the need of diverse users. It includes essential functionalities such as farm registration, cow registration, and real-time notifications. The Farm Registration screen allows farmers to input farm details, ensuring proper identification and data management. The Cow Registration screen enables the recording of individual cow profiles, including vital health parameters. The Notification section displays real-time alerts generated by the system, helping farmers take timely action to prevent health issues. The structured design ensures ease of use, making the system accessible even to users with limited technical expertise.

Figure 6a–c illustrate the user interfaces of the mobile application in English, specifically highlighting the Farm Registration, Cow Registration, and Notification screens, respectively. Similarly, Fig. 7a–c display these interfaces in Punjabi language. These screens provide an intuitive and comprehensive interface that facilitates effective and efficient management of both farm operations and cattle health.

3.6 Ethical and animal welfare considerations

This study was conducted in strict accordance with ethical guidelines to ensure the welfare and humane treatment of all animals involved. Non-invasive data collection methods were employed to minimize any potential discomfort or distress to the cattle. The IPR, Legal, and Ethical Matters Committee of Sri Guru Granth Sahib World University waived formal ethical approval, acknowledging that the observational and monitoring nature of the research posed no risk to the animals.

Informed consent was obtained from all participating dairy farm owners and cattle shelter representatives prior to data collection, promoting transparency and collaborative participation. Certification letters were collected from the participating farms, explicitly confirming awareness of the data collection procedures and its intended use

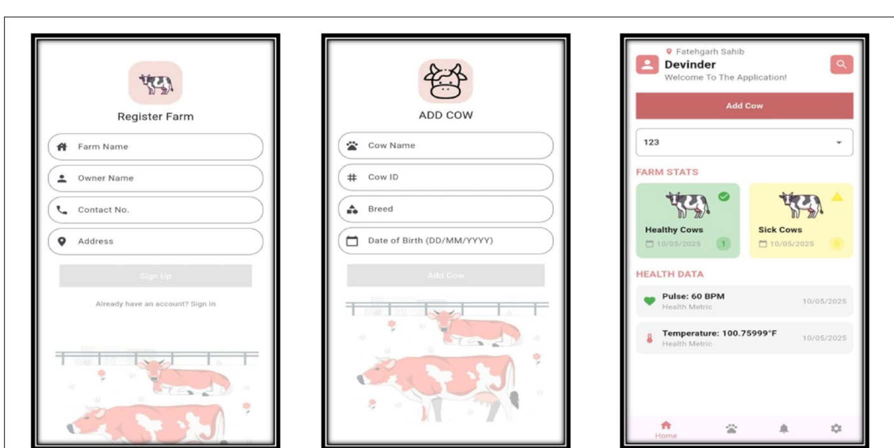

Fig. 6 **a** Farm registration, **b** Cow registration, **c** Notifications

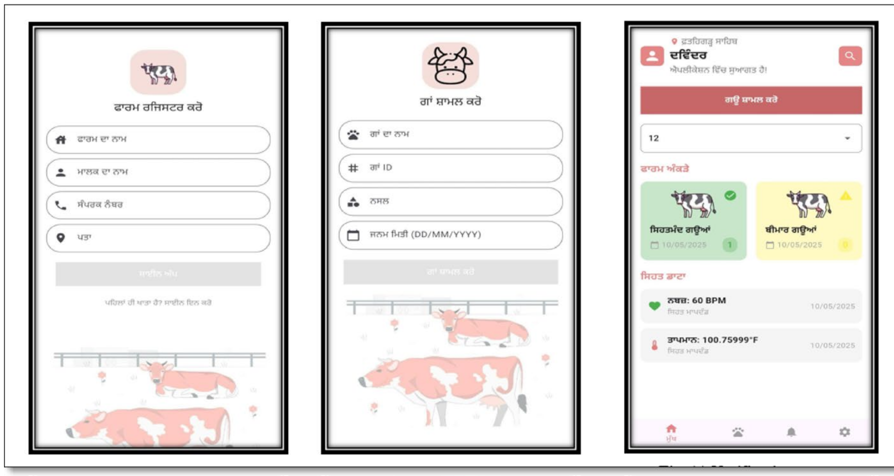

Fig. 7 **a** Farm registration (Punjabi version), **b** Cow registration (Punjabi version), **c** Notifications (Punjabi version)

for research. Data collection and monitoring were carried out under the supervision of qualified veterinary professionals, ensuring that animal welfare remained a primary concern throughout the study. Additionally, a certification issued by the supervising veterinary officer confirmed the correct and safe functioning of the device during trials.

Access to sensitive health data was restricted to the principal investigator and veterinary supervisors overseeing the fieldwork. This controlled access safeguarded the confidentiality and integrity of the data. Data management adhered to recognize the best practices aligned with standards for animal health information, including secure storage protocols to prevent unauthorized access or breaches. Since the system employed cloud-based storage via Firebase, potential risks such as data breaches or unauthorized third-party access were also considered. To mitigate these, stored data were anonymized to prevent identification of specific farms or animals, and access rights were restricted to authorized personnel only. Farmers were informed of these measures during the consent process.

This comprehensive ethical framework—covering animal welfare, informed consent, certification, verification, restricted data access, and secure cloud data management—supports responsible use, ownership, and storage of livestock health data. Adherence to these principles reinforces the scientific rigor, reliability, and social responsibility of the research while addressing important considerations regarding data privacy and ethics in agricultural and veterinary settings.

4 Proposed system

The proposed system combines a smart neck-mounted IoT device with a cloud-driven machine learning framework to facilitate continuous, real-time monitoring of cattle health. Physiological signals captured by the wearable sensors are wirelessly transmitted to the cloud, where the chosen ML model processes the data to detect early signs of health issues such as lameness and metabolic disorders. Tailored for smallholder dairy operations, this low-cost, non-invasive solution improves animal health management by enabling timely intervention and promoting both welfare and farm productivity.

4.1 System architecture

Figure 8 illustrates the layered architecture of the proposed IoT-based cow health monitoring system, visually mapping the data flow from on-animal sensing to end-user delivery. The model is divided into four primary layers — *Perception, Network, Cloud, and Application*. This arrangement not only organizes functional responsibilities but also supports modularity, scalability, and secure real-time operation.

- *Perception Layer (Data Acquisition and On-Device Preprocessing):* At the base of the system, the smart neck collar device, equipped with sensors, continuously collects physiological (e.g., body temperature, pulse rate) and behavioural (e.g., activity levels) data. The ESP32 microcontroller performs on-device preprocessing, including noise filtering, normalization, and feature extraction, reducing the transmission load while improving analytical accuracy. Importantly, the system supports local caching with batch uploads, ensuring that data are reliably stored when network connectivity is unavailable and transmitted to the cloud once the network is restored. This feature enables continuous monitoring and data integrity, particularly in rural or resource-limited settings with intermittent connectivity.
- *Network Layer (Secure Data Transmission)*: The processed data is securely transmitted to the cloud via Wi-Fi (IEEE 802.11), as depicted by the Wi-Fi and HTTPS icons in the Fig. 8. All communication is encrypted using TLS, ensuring data integrity. For areas with intermittent connectivity, the system supports local caching with batch uploads when the network becomes available.
- *Cloud Layer (Data Storage, Machine Learning, and Analytics):* In the cloud segment of the figure, Firebase Realtime Database and Cloud Storage handle secure data storage. Incoming data triggers the hybrid ML model pipeline to detect early signs of health anomalies in real time.

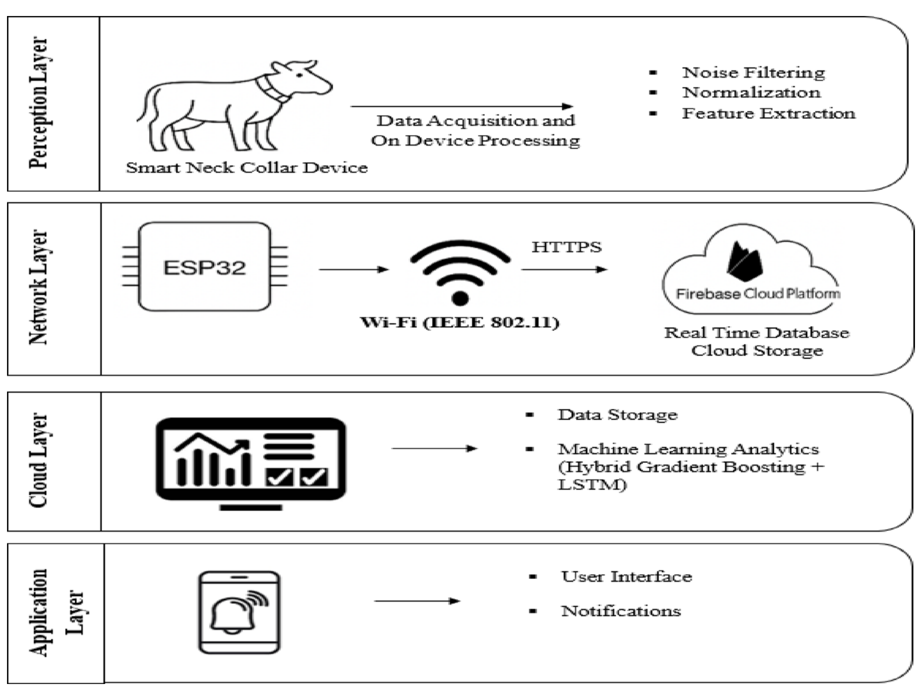

Fig. 8 Layered architecture of the proposed IoT-based livestock health monitoring system

- *Application Layer (User Interface and Notifications):* The top layer of the figure depicts the farmer-friendly mobile application, which retrieves predictions securely from the cloud and delivers intuitive notifications. Abnormal health conditions trigger immediate alerts, enabling timely intervention and improved livestock welfare.

By visually linking each stage of sensing, transmission, analysis, and user interaction, Fig. 8 emphasizes how the proposed system achieves reliable, secure, and actionable health monitoring in practical farm environments.

4.2 Communication protocols

The system utilizes HTTPS as the primary protocol for secure data transmission between IoT devices and the cloud backend. This approach aligns with best practices for cloud communication security and supports seamless integration with Firebase services, which mandate HTTPS for all data exchanges.

1. *Security and Privacy*: User and device authentication are enforced via Firebase Authentication, which provides robust identity management and access control [23]. Combined with HTTPS, this layered security framework protects sensitive livestock health information from unauthorized access, ensuring data privacy and compliance with modern security standards relevant to IoT healthcare systems [24].
2. *On-Device Preprocessing:* The smart collar device implements preprocessing algorithms locally, including noise filtering, normalization, and feature extraction of sensor data. This edge processing reduces data size, lowers network load, and enhances the accuracy of transmitted data [25]. Such on-device computation is critical for maintaining low latency and energy efficiency in resource-constrained settings [26].
3. *Cloud-Based Prediction with SM-BoostLSTM:* The cloud analytics system utilizes the hybrid model SM-BoostLSTM that incorporates SMOTE to address data imbalance, Gradient Boosting for modelling intricate nonlinear relationships in structured data, and LSTM networks to capture temporal dynamics in time-series sensor data [25, 26]. This integrated approach significantly improves the accuracy of cattle health classification, enabling timely identification of anomalies and diseases [27].
4. *Real-Time Analytics and Notifications*: Real-time processing of incoming data streams facilitates immediate health status assessment and alert generation, which are communicated to end users via a dedicated mobile application.

4.3 Workflow of the proposed model

The overall workflow of the proposed SM-BoostLSTM framework is illustrated in Fig. 9. The process begins with sensor-based data collection, followed by systematic preprocessing steps including handling missing values, anomaly detection and removal, and data cleaning. Relevant temporal and statistical features are then extracted for model training.

To address class imbalance in the dataset, the Synthetic Minority Oversampling Technique (SMOTE) was applied after feature extraction and within each training fold during cross-validation, rather than before the train–test split. This approach ensured that oversampling was performed only on the training data, thereby preventing potential data leakage that could occur if synthetic samples were generated using information from the

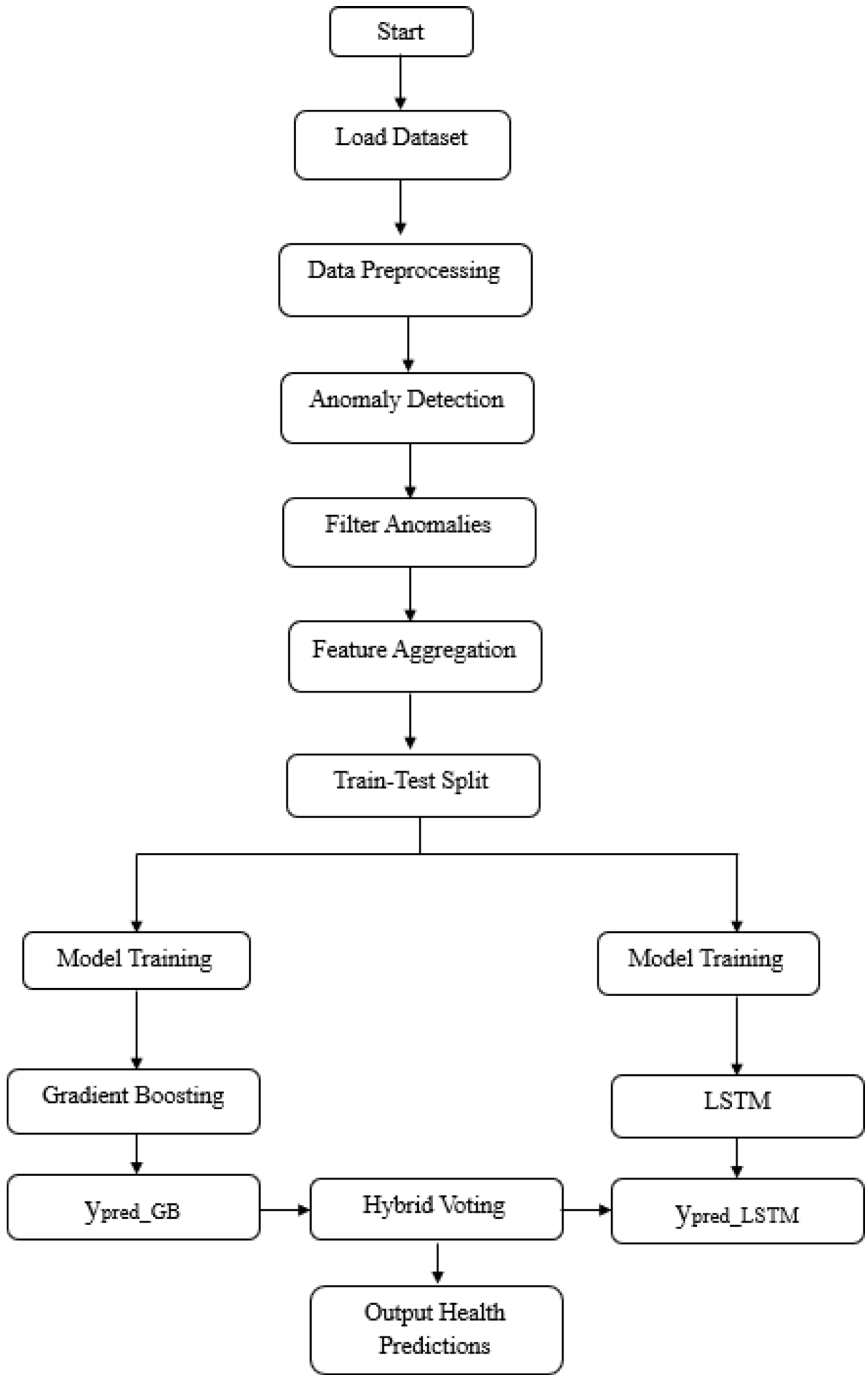

Fig. 9 Overall workflow of the proposed SM-BoostLSTM framework

test set. Applying SMOTE within folds enabled the model to learn from a balanced class distribution while preserving the independence and integrity of the validation data.

The balanced training data were then used to develop the SM-BoostLSTM hybrid model, which integrates a Gradient Boosting classifier and a Long Short-Term Memory (LSTM) network to leverage both structured and temporal aspects of the cattle health dataset. After independent model training, the outputs from both components were combined using a hybrid voting mechanism to generate the final disease prediction. When both models agreed, their shared prediction was accepted; in cases of disagreement, the Gradient Boosting output acted as the tie-breaker due to its superior consistency with structured physiological data. This fusion strategy effectively combines the interpretability and stability of Gradient Boosting with the temporal learning capabilities of LSTM, resulting in a system that demonstrates enhanced robustness, precision, and generalization. The final hybrid model was deployed in a mobile application for real-time cow health monitoring, and its performance was evaluated using predictive accuracy to ensure both technical reliability and user accessibility.

Although daily mean values of physiological parameters (pulse rate, temperature, and activity level) have been used as individual feature representations, the LSTM component of the proposed SM-BoostLSTM framework processes multi-day sequences for each cow. Each input sequence thus consists of several consecutive daily records, enabling the LSTM to learn temporal dependencies and health trends across days rather than within a single day. This design captures gradual variations in physiological signals that precede health anomalies while maintaining robustness against missing intra-day data.

This sequence-level modelling ensures that the LSTM retains its temporal learning capability and does not degenerate into a purely feed-forward layer, as it leverages inter-day temporal dynamics rather than single-day observations.

The proposed SM-BoostLSTM framework is designed to monitor and predict the onset of common and economically significant cattle health conditions, including Bovine Respiratory Disease (BRD), Lameness, Metabolic Disorders, and Bovine Viral Diarrhoea (BVD). These conditions were selected because they cause detectable physiological and behavioural deviations that can be captured by sensor-based measurements of body temperature, pulse rate, and activity level.

Specifically, elevated body temperature and accelerated pulse rate are early indicators of BRD and BVD, while reduced mobility is characteristic of lameness and metabolic disorders. By continuously analysing these interrelated parameters, the proposed system can infer abnormal physiological trends and provide early warnings of disease onset. This disease–parameter mapping was established based on veterinary literature and validated through consultation with veterinary experts.

4.4 Implications of the proposed system

The proposed system has important implications for smallholder dairy farming by enabling continuous monitoring and early disease detection which are described as follows:

1. *Real-time health monitoring*

- Unlike traditional methods that rely on manual inspections or RFID-based tracking, this smart collar continuously measures temperature, pulse rate, and activity level.
- Farmers receive instant alerts via a mobile app, allowing for quick interventions.

2. *Machine learning for early disease detection*

- The proposed system employs a hybrid model, SM-BoostLSTM, which demonstrated superior performance in terms of accuracy,precision, recall and F1 score.
- This balanced performance—marked by high recall and strong precision—effectively minimizes both false positives and false negatives, enabling reliable early detection of diseases such as mastitis, lameness, and metabolic disorders.
- Compared to traditional threshold-based monitoring techniques, SM-BoostLSTM offers a substantial improvement in predictive accuracy, supporting timely and data-driven interventions in livestock health management.

3. *Cost-effectiveness & scalability*

- Uses low-power components (e.g., 360 mAh LiPo battery) for long operational life, reducing maintenance costs.
- Modular design makes it easily adaptable for farms of different sizes.

4. *Improved dairy productivity*

- By detecting diseases early, it reduces milk production losses.
- Enhances overall farm efficiency and profitability.

4.5 Comparative analysis of existing systems and the proposed system

To evaluate the effectiveness and novelty of the proposed IoT-based cattle health monitoring system, a comparative analysis has been conducted against existing systems reported in the literature. Table 7 presents a structured comparison based on key aspects

Table 7 Proposed system vs. Existing systems

Aspect	Existing systems	Proposed system
Target users	Primarily developed for large-scale commercial farms with access to advanced infrastructure (e.g., [8, 9, 13, 28, 29])	Specifically tailored for small-scale and resource-constrained farms, with emphasis on affordability and ease of use
Hardware complexity	Use of high-end or multiple hardware devices: CombiFoss™ 7 [8], CowView & pH bolus [9], IceQube [10], Lely Astronaut AMS [12], Calving monitoring systems [28]	Utilizes a custom-built, lightweight smart neck collar with multi-sensor integration, suitable for widespread deployment
ML techniques used	Variety of classifiers used: RF, SVM, CNN, LSTM, GBT, DT, etc. across different systems ([8–10, 13, 19, 28, 29])	Employs a hybrid GBoost + LSTM + SMOTE model optimized for both temporal dynamics and classification performance
Model Performance	Performance varies: high specificity but low sensitivity in some (e.g., [29]: 99% specificity, 40% sensitivity; [28]: reported metrics for calving detection)	Delivers balanced and high-performance metrics across accuracy, precision, recall, and F1-score, ensuring both sensitivity and specificity are maintained
Infrastructure dependency	Dependent on complex systems like robotic milking [12], real-time locating systems [9], and weather stations [16]	Requires only basic infrastructure and is operable in rural field conditions with Wi-Fi communication
User accessibility	Limited local language support; often not designed for direct farmer interaction ([9, 13, 28])	Includes a bilingual Mobile app (English and Punjabi), enhancing usability for local farmers and enabling real-time health alerts
Scalability and cost	High cost and technical complexity reduce feasibility for small-scale use ([9, 12, 28])	Low-cost design with modular hardware and cloud integration enables scalable deployment in rural and underserved areas
Environmental adaptability	Typically suited for controlled or high-tech environments ([8, 9, 16, 28])	Designed and tested in the field under real environmental and farm conditions in Punjab, India

such as hardware complexity, machine learning algorithms used, model performance and environmental adaptability.

The proposed system revolutionizes precision livestock farming by offering real-time, intelligent, and cost-effective monitoring using IoT and machine learning, surpassing traditional methods in managing cow health and dairy farms.

5 Result

To identify the most effective machine learning approach for early disease detection in cattle, multiple models were evaluated using the collected sensor data. The results of all the tested and evaluated models are presented in Table 8. Traditional classifiers achieved moderate performance, with the Neural Network (86.20% accuracy, F1-score 67.68%) and Decision Tree (84.08% accuracy, F1-score 68.68%) performing relatively better than other models.

Ensemble models showed a notable improvement, with Gradient Boosting (91.41% accuracy, F1-score 79.2%) outperforming XGBoost across most metrics. This confirmed the advantage of ensemble learning in reducing bias–variance trade-offs compared to traditional classifiers.

Hybrid approaches further improved the results by integrating LSTM with ensemble models. In particular, the SM-BoostLSTM (Gradient Boost with LSTM and SMOTE) achieved the best overall performance with 93.56% accuracy, 91.42% precision, 77.77% recall, and 84.02% F1-score. This demonstrates that incorporating temporal sequence modelling (via LSTM) along with oversampling for class imbalance (via SMOTE) significantly enhances predictive accuracy and robustness.

Overall, the results indicate a clear progression in performance from traditional models → ensemble methods → hybrid approaches, highlighting the importance of combining advanced techniques for reliable disease detection in cows.

Overall, the proposed SM-BoostLSTM approach offers the best trade-off between accuracy, precision, and recall, making it more reliable for real-world cow health prediction scenarios compared to the other tested model combinations.

Table 8 Performance comparison of traditional, ensemble and hybrid machine learning models

Traditional models				
Technique	**Accuracy(%)**	**Precision(%)**	**Recall(%)**	**F1 Score(%)**
Random forest	81.78	64.22	50.50	56.54
Logistic regression	77.65	55.52	76.51	64.35
Support vector machine	79.36	58.52	74.50	65.55
Decision tree	84.08	71.33	66.22	68.68
Naive bayes	66.75	43.11	81.88	56.48
K-nearest neighbors	83.49	69.05	51.79	59.18
Neural network	86.20	73.80	62.50	67.68
LSTM	82.44	72.27	54.12	61.89
Ensemble models				
Gradient boosting	91.41	90	70.71	79.2
Extreme gradient boosting	86.93	82.77	63.65	71.96
Hybrid models				
Random forest with LSTM	86.59	93.99	52.46	67.34
Neural networks with LSTM	89.00	78.04	72.91	75.39
Extreme gradient boost with LSTM	85.29	83.1	55.46	66.52
Gradient boost with LSTM	91.69	89.34	72.7	80.17
SM-BoostLSTM (Gradient Boost with LSTM and SMOTE)	93.56	91.42	77.77	84.02

Table 9 Confusion Matrix for SM-BoostLSTM on 150 Cows

	Predicted healthy (0)	Predicted sick (1)
Actual healthy (0)	113	2
Actual sick (1)	8	27

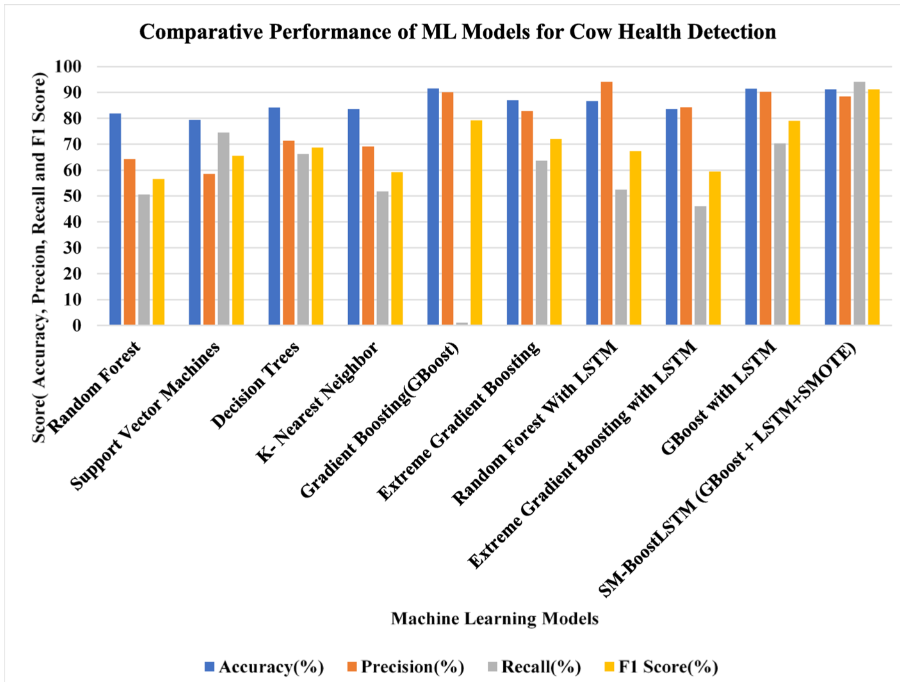

Fig. 10 Comparison of precision, accuracy, recall, and F1 scores for various machine learning techniques applied to cow health detection

Table 9 depicts the confusion matrix for SM-BoostLSTM using a sample size of 150 cows.

The SM-BoostLSTM model achieves the best performance, correctly predicting 113 healthy cows and 27 sick cows. Only 2 healthy and 8 sick cows are misclassified. This highlights the effectiveness of combining Gradient Boosting with LSTM and applying SMOTE, resulting in improved recall, balanced predictions, and higher overall accuracy.

Figure 10 presents a comparative performance analysis of multiple machine learning models applied to cow health status prediction. This comparison highlights the trade-offs between models and underscores the superior performance of the proposed SM-BoostLSTM approach, which achieves high recall while maintaining strong precision, thereby ensuring timely and reliable health alerts.

The proposed SM-BoostLSTM framework integrates a Gradient Boosting (GB) classifier with a Long Short-Term Memory (LSTM) neural network, preceded by class balancing via the Synthetic Minority Oversampling Technique (SMOTE). Gradient Boosting hyperparameters were optimized using GridSearchCV with fivefold cross-validation, yielding an optimal configuration with 150 estimators, a learning rate of 0.05, maximum tree depth of 4, a minimum samples split of 2, and a subsample ratio of 0.8 (random_state = 42). The LSTM component was implemented in Keras with a TensorFlow backend and configured with a single LSTM layer containing 50 units and ReLU activation, followed by a dropout layer with a rate of 0.2 to reduce overfitting, and a dense output layer with sigmoid activation. The model was trained using the Adam optimizer (learning rate = 0.001) and binary cross-entropy loss for 10 epochs with a batch size of 32.

The input to the LSTM comprised sequences of shape (1, 3), corresponding to one timestep with three features: pulse rate, temperature, and activity level. Predictions from the GB and LSTM models were combined using a hybrid voting mechanism to generate the final classification output. This explicit specification of architecture and tuned parameters ensures reproducibility enablingother researchers to replicate the proposed approach in similar problem domains.

Crucially, to ensure rigorous statistical evaluation, fivefold cross-validation was employed not only during hyperparameter tuning but also for final model assessment. Cross-validation accuracy scores were calculated across folds, yielding an average accuracy with reduced variance due to multiple train-test splits. This approach mitigates biases arising from a single random split and provides a more robust estimate of generalization performance.

Overall, this multi-faceted evaluation framework adequately addresses concerns regarding the sufficiency of the train-test split and demonstrates that the proposed model reliably predicts cattle health status across varied conditions.

6 Discussions

The evaluation of various machine learning models revealed that ensemble techniques, such as Gradient Boosting (GBoost) and Extreme Gradient Boosting, demonstrated superior performance when integrated with Long Short-Term Memory (LSTM) for predicting cow health status. Among these, SM-BoostLSTM emerged as the most effective model. This combination outperformed all other approaches, offering an optimal balance between precision and recall, which is crucial for early disease detection while minimizing false positives.

While the integration of SMOTE, Gradient Boosting, and LSTM enhances class balance, temporal feature learning, and predictive accuracy, this hybrid approach may incur higher computational costs compared to single-model classifiers. Although exact training and inference times were not recorded during this study, the sequential nature of LSTM processing and the iterative boosting mechanism of Gradient Boosting are expected to increase processing requirements. These demands are typically manageable in desktop or cloud environments but could present challenges for on-device or low-resource deployments in rural farm settings. Future work will benchmark the complete SM-BOOSTLSTM pipeline in terms of training time, inference latency, and energy consumption, and will explore model compression and optimization techniques to improve efficiency for field deployment.

The designed IoT-based health monitoring system is inherently scalable to larger herds due to its modular architecture. The smart collar device operates independently with low power consumption, enabling simultaneous monitoring of multiple animals without overwhelming local resources. Data from numerous devices can be aggregated and processed efficiently on the cloud platform, which supports real-time analysis and alerting regardless of herd size. This scalability makes the system adaptable to both smallholder farms and large commercial dairy operations, promoting widespread adoption in diverse farming environments.

While the focus of this study was to develop and validate an accurate and easy-to-use cattle health monitoring system, practical factors such as energy efficiency, cost-effectiveness, and potential for long-term deployment were also carefully considered during the design process. The hardware uses low-power components, including the ESP32 microcontroller known for its low energy consumption and built-in connectivity. Energy-efficient sensors were chosen to reduce power usage without sacrificing data quality, allowing the device to operate for extended periods on battery power—an important feature for farm environments where frequent charging is difficult.

Cost is an important factor in selecting components, with the aim of keeping the overall system affordable for small- to medium-scale farmers who often have limited resources. By using widely available commercial sensors and open-source software platforms, the production cost remains low while maintaining reliable performance.

6.1 System limitations

Despite the promising performance of the proposed IoT-enabled cow health monitoring system, several practical limitations must be acknowledged. One key concern is the potential degradation of sensor performance over time due to factors such as hardware wear, environmental interference, or calibration inconsistencies. These issues can undermine data integrity, potentially leading to misclassification of health conditions and delayed interventions. Addressing these issues, require routine maintenance schedules and sensor recalibration protocols to maintain operational accuracy.

Another significant limitation involves the reliance of the system on continuous internet connectivity for real-time data synchronization and cloud-based analysis. In many rural dairy farming regions, inconsistent or limited network access can disrupt data transmission and delay health alerts, thereby reducing the timeliness and reliability of the system response. This limitation may directly affect decision-making during critical health events.

Enhancements to the system architecture are envisioned to mitigate these issues. These include the incorporation of onboard data storage with offline processing to ensure data buffering during connectivity lapses, and the development of more robust, fault-tolerant sensor networks equipped with self-diagnostic capabilities. Moreover, integrating alternative communication protocols such as LoRaWAN or mesh networking could significantly improve system performance in low-connectivity environments.

6.2 Challenges in system implementation

While the layered architecture of the system addresses core functionalities such as data collection, processing, and user communication, several implementation-level challenges remain that may affect scalability and efficiency:

1. *Latency:* Achieving real-time responsiveness is vital for timely health anomaly detection. Although on-device preprocessing reduces the volume of data transmission, delays during cloud communication and inference can still impact the overall responsiveness of the system. To address this, edge computing strategies and asynchronous data handling mechanisms have been incorporated to improve latency performance.
2. *Bandwidth Limitations:* Rural deployments often face bandwidth scarcity due to unreliable or slow internet connections. Relying on Wi-Fi and secure HTTPS protocols

can intensify these limitations. To optimize performance under such constraints, the system employs local preprocessing to reduce data volume and uses batch uploads to synchronize data when network access is restored.

3. *Interoperability:* While HTTPS enables secure communication with cloud services, it can restrict compatibility with other IoT ecosystems that operate using lighter protocols like MQTT. To ensure long-term scalability and integration with heterogeneous devices, future system iterations should adopt modular interfaces and standardized communication protocols.
4. *Resource Constraints:* The wearable smart collar must operate efficiently within strict power and processing limits. To ensure prolonged battery life and minimal computational load, lightweight preprocessing algorithms are used. These algorithms strike a balance between conserving energy and preserving data quality for accurate analysis in the cloud.

To overcome these technical limitations and deployment challenges, future development efforts will focus on enhancing system autonomy, improving robustness under field conditions, and expanding connectivity options. Incorporating self-healing sensor networks, supporting multi-protocol communication stack, and enabling adaptive data management will be key to ensure broader applicability and long-term reliability of the system in diverse agricultural environments.

7 Conclusions

This study presents a novel IoT-based system for real-time cow health monitoring, integrating custom-designed wearable sensor technology with advanced hybrid machine learning models to enhance disease prediction accuracy. A comprehensive dataset enabled rigorous evaluation of both traditional and hybrid approaches, with the Gradient Boosting (GBoost) combined with Long Short-Term Memory (LSTM) networks delivering the best performance in terms of accuracy, precision, recall, and F1-score. SMOTE technique was employed to address class imbalance and optimize model effectiveness.

The system stands out for its end-to-end design—from data acquisition using a smart collar, to cloud-based analysis, and finally farmer notification via a bilingual (English and Punjabi) mobile application. This pipeline ensures real-time alerts, enabling early detection of diseases and supporting timely interventions. Its affordability, adaptability to local infrastructure, and focus on smallholder farmers contribute to its scalability and suitability in resource-limited agricultural settings.

Future work will focus on expanding the dataset by including more animals and diverse regions, incorporating additional physiological and behavioural indicators, and investigating advanced learning architectures such as attention-based models to enhance the generalizability of the system across different regions and cattle breeds. Although quantitative benchmarking using publicly available datasets is currently limited due to differences in testing conditions, continuous collaboration with veterinarians has validated the accuracy and reliability of the health alerts produced by the system. Subsequent research will aim to strengthen benchmarking by incorporating baseline heuristic models and conducting direct comparisons with conventional veterinary diagnostics, thereby further demonstrating robustness and practical applicability of the system.

Acknowledgements
The authors extend their sincere gratitude to the staff of the Government Veterinary Hospital, Amloh, District Fatehgarh Sahib, as well as the farm owners and support staff of private dairy farms across various districts of Punjab for their invaluable support and cooperation during the data collection process.

Author contributions
Devinder Kaur—Conceptualization, methodology, data collection, formal analysis, investigation, data curation, writing—original draft preparation, writing—review and editing. Amandeep Kaur Virk—Supervision, methodology guidance, writing—review and editing.

Funding
This study was carried out without financial support from external sources.

Data availability
The data supporting the findings of this study are available from the corresponding author upon reasonable request. Due to the sensitive nature of the data, which includes sensor information from dairy cows, it will be shared on a case-by-case basis to ensure its appropriate and ethical use.

Declarations

Ethics approval and consent to participate
This research involved non-invasive experimentation on cows to collect vital health parameters while ensuring their well-being. All cows, irrespective of age or breed, were treated with care and respect. Data collection was conducted under the supervision and with the consent of a professional veterinary team, prioritizing the animals' welfare. No harmful procedures were used, and veterinarians confirmed that the sensor device caused no lasting harm. Ethical approval was waived by the IPR, Legal, and Ethical Matters Committee of Sri Guru Granth Sahib World University, as the study involved only non-invasive observation and monitoring, ensuring that no harm or distress was caused to the animals.Informed consent was obtained from the dairy farm owners and gaushala representatives prior to data collection.

Consent to publish
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 7 September 2025 / Accepted: 24 October 2025
Published online: 21 November 2025

References

1. Press Information Bureau, Government of India. Dairy sector employs over 80 million people, mostly small and marginal farmers: Shri Parshottam Rupala [Internet]. 2023 May 11 [cited 2025 May 19]. Available from: https://www.pib.gov.in/PressReleasePage.aspx?PRID=1929183
2. Nagl L, Schmitz R, Warren S, Hildreth TS, Erickson H, Andresen D. Wearable sensor system for wireless state-of-health determination in cattle. In: Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 2003 Sep 17–21; Cancun, Mexico. p. 3012–5.
3. Suresh A, Sarath TV. An IoT solution for cattle health monitoring. IOP Conf Ser Mater Sci Eng. 2019;561(1):012106. https://doi.org/10.1088/1757-899X/561/1/012106.
4. Tippannavar SS, Nayak P, Punith G. Smart collar for cattle tracking and health monitoring using IoT on ESP32. Int J Innov Res Adv Eng. 2023;10(3):81–5.
5. Vyas S, Shukla V, Doshi N. FMD and mastitis disease detection in cows using internet of things (IoT). Procedia Comput Sci. 2019;160:728–33. https://doi.org/10.1016/j.procs.2019.11.019.
6. Khan AA, Laghari AA, Baqasah AM, Alroobaea R, Gadekallu TR, Sampedro GA, et al. ORAN-B5G: a next-generation open radio access network architecture with machine learning for beyond 5G in industrial 5.0. IEEE Trans Green Commun Netw. 2024;8(3):1026–36. https://doi.org/10.1109/TGCN.2024.3396454.
7. Khan AA, Laghari AA, Baqasah AM, Bacarra R, Alroobaea R, Alsafyani M, et al. BDLT-IoMT—a novel architecture: SVM machine learning for robust and secure data processing in internet of medical things with blockchain cybersecurity. J Supercomput. 2025;81(1):1–22.
8. Contla Hernández B, Lopez-Villalobos N, Vignes M. Identifying health status in grazing dairy cows from milk mid-infrared spectroscopy by using machine learning methods. Animals (Basel). 2021;11(8):2154. https://doi.org/10.3390/ani11082154.
9. Wagner N, Antoine V, Miahhalon MM, Lardy R, Silberberg M, Koko J, et al. Machine learning to detect behavioural anomalies in dairy cows under subacute ruminal acidosis. Comput Electron Agric. 2020;170:105233.
10. Borchers MR, Chang YM, Proudfoot KL, Wadsworth BA, Stone AE, Bewley JM. Machine-learning-based calving prediction from activity, lying, and ruminating behaviors in dairy cattle. J Dairy Sci. 2017;100(7):5664–74. https://doi.org/10.3168/jds.2016-11526.
11. Xu W, van Knegsel AT, Vervoort JJ, Bruckmaier RM, van Hoeij RJ, Kemp B, et al. Prediction of metabolic status of dairy cows in early lactation with on-farm cow data and machine learning algorithms. J Dairy Sci. 2019;102(11):10186–201.
12. Kamphuis C, Mollenhorst H, Heesterbeek JA, Hogeveen H. Detection of clinical mastitis with sensor data from automatic milking systems is improved by using decision-tree induction. J Dairy Sci. 2010;93(8):3616–27. https://doi.org/10.3168/jds.2010-3228.

13. Ebrahimi M, Mohammadi-Dehcheshmeh M, Ebrahimie E, Petrovski KR. Comprehensive analysis of machine learning models for prediction of sub-clinical mastitis: deep learning and gradient-boosted trees outperform other models. Comput Biol Med. 2019;114:103456. https://doi.org/10.1016/j.compbiomed.2019.103456.
14. Ankitha K, Manjaiah DH. Comparison of KNN and SVM algorithms to detect clinical mastitis in cows using internet of animal health things. In: Satapathy SC, Zhang YD, Bhateja V, Majhi R, editors. Intelligent data engineering and analytics. Advances in Intelligent Systems and Computing. . Singapore: Springer; 2021.
15. Shahinfar S, Khansefid M, Haile-Mariam M, Pryce JE. Machine learning approaches for the prediction of lameness in dairy cows. Animal. 2021;15(11):100391. https://doi.org/10.1016/j.animal.2021.100391.
16. Gorczyca MT, Gebremedhin KG. Ranking of environmental heat stressors for dairy cows using machine learning algorithms. Comput Electron Agric. 2020;168:105124. https://doi.org/10.1016/j.compag.2019.105124.
17. Zaborski D, Proskura WS, Grzesiak W, Różańska-Zawieja J, Sobek Z. The comparison between random forest and boosted trees for dystocia detection in dairy cows. Comput Electron Agric. 2019;163:104856. https://doi.org/10.1016/j.compag.2019.104856.
18. Dofadar DF, Abdullah HM, Khan RH, Rahman R, Ahmed MS. A comparative analysis of lumpy skin disease prediction through machine learning approaches. In: 2022 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET); 2022 Sep 20–22; Dhaka, Bangladesh. p. 1–4. Available from: https://doi.org/10.1109/IICAIET55139.2022.9936742
19. Zhou X, Xu C, Wang H, Xu W, Zhao Z, Chen M, et al. The early prediction of common disorders in dairy cows monitored by automatic systems with machine learning algorithms. Animals (Basel). 2022;12(10):1251. https://doi.org/10.3390/ani12101251.
20. Arshad J, Siddiqui TA, Sheikh MI, Waseem MS, Nawaz MA, Eldin ET, et al. Deployment of an intelligent and secure cattle health monitoring system. Egypt Inform J. 2023;24(2):265–75. https://doi.org/10.1016/j.eij.2023.04.001.
21. Avcı İ, Alzabaq A. A new respiratory diseases detection model in chest X-ray images using CNN. Traitement du Signal. 2023;40(1):145. https://doi.org/10.18280/ts.400113.
22. Kaur D, Virk AK. Smart neck collar: IoT-based disease detection and health monitoring for dairy cows. Discover Internet Things. 2025;5(1):12. https://doi.org/10.1007/s43926-025-00109-5.
23. Rescorla E. HTTP over TLS. RFC 2818. Internet Eng Task Force; 2000 May. Available from: https://tools.ietf.org/html/rfc2818
24. Google. Privacy and security in Firebase. Firebase Documentation. Available from: https://firebase.google.com/support/privacy. Accessed on: 20 August 2025.
25. Google. Firebase Authentication overview [Internet]. Available from: https://firebase.google.com/docs/auth. Accessed on: 20 August 2025.
26. Conti M, Dehghantanha A, Franke K, Watson S. Internet of things security and forensics: challenges and opportunities. Future Gener Comput Syst. 2018;78:544–6. https://doi.org/10.1016/j.future.2017.07.060.
27. Chen M, Ma Y, Song J, Lai CH, Hu B. Smart clothing: connecting human with clouds and big data for sustainable health monitoring. Mob Netw Appl. 2016;21(5):825–45.
28. Sumi K, Zin TT, Kobayashi I, Horii Y. A study on cow monitoring system for calving process. In: Proceedings of the IEEE 6th Global Conference on Consumer Electronics (GCCE); 2017 Oct 24–27; Nagoya, Japan. Piscataway (NJ): IEEE; 2017. p. 1–2. Available from: https://doi.org/10.1109/GCCE.2017.8229442.
29. Moe AST, Tin P, Aikawa M, Kobayashi I, Zin TT. AI-powered visual E-monitoring system for cattle health and wealth. Smart Agric Technol. 2025;12:101300. https://doi.org/10.1016/j.atech.2025.101300.

Publisher's note