International Journal of Applied Mathematics

Volume 38 No. 4 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

IOT DRIVEN XGBOOST POWERED REAL-TIME ACCIDENT PREDICTION SYSTEM IN CONSTRUCTION SITES WITH ENRICHED FEATURES

S. Eben Paul Richard[1*], G. Beni[2]

1Research Scholar, Department of CSE, Noorul Islam Centre for Higher Education, Kumaracoil, India-629180, ebenpaulrichard@gmail.com

2Associate Professor, Department of CSE, Noorul Islam Centre for Higher Education, Kumaracoil, India-629180. Gsbeni2005@gmail.com

Abstract

Worker safety is the predominant requirement in the construction industry, which has made the real-time safety monitoring crucial. This work aims to ensure workers safety in construction industry by presenting an Internet of Things (IoT) driven framework that leverages data from physiological, environmental and locational sensors being integrated with Raspberry Pi 4. The dataset undergoes rigorous pre-processing activity to remove outliers, normalize and balance the dataset. This work utilizes enriched features that deal with statistical, temporal and frequency domains. XGBoost classifier is trained as a binary classification model with 'safe' and 'alert' classes. The performance of the proposed work is evaluated with standard performance measures such as accuracy, precision, recall and F-measure rates. The proposed work shows promising results and is suitable for real-time environment.

Keywords – IoT, construction site, worker safety, machine learning, sensors.

INTRODUCTION

Construction industry is one of the global hazardous sectors, where the workers are confronted by numerous challenges such as harmful materials, machinery, structural unpredictability, unfavorable weather conditions and so on [1]. The International Labor Organization (ILO) estimates that the construction sector involves about 7% of the global workforce that contributes about 30-40% occupational fatalities [2,3]. As far as Indian construction industry is concerned, over 400 fatalities happened with more than 850 serious injuries [4]. However, most of the safety measures in today's construction industry rely on post-accident protocols with first-aid teams and emergency alarms. Though standard safety rules are followed, they are not personalized for every worker or working site, as the health condition or the environmental condition varies from person and site respectively [5,6].

Worker safety must be the paramount focus in every industry, particularly in the construction sector, because to the inherent risks associated with its foundational activities. Human injuries or life loss is one facet of accidents and it may result in economic loss as well. A well-planned construction site can enjoy the benefits of site control, efficiency, better reputation and so on. The construction site can be made smart by embracing Internet of Things (IoT) [7].

Received: August 08 ,2025

Volume 38 No. 4 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

IoT is a network of numerous physical devices that are comprised of sensors, software and connectivity to make them collect, process and exchange data without human interference. IoT technology makes the real-time data processing effective, in almost all the domains such as healthcare, security, surveillance, manufacturing, agriculture and so on [8]. Additionally, it shows better performance in carrying out prediction-based methods in a minimal cost incurrence. Machine Learning (ML) and Deep Learning (DL) schemes can be integrated with IoT framework effectively [9]. Hence, IoT provides a key to reach digital world from the physical world by performing safety, security, automation, cost effectiveness and so on.

Construction industry is still not updated with recent technologies, when compared to other domains [10]. Hence, it is an absolute necessity to present a proactive smart system that can detect and predict accidents in real-time construction sites. Industry 4.0 and smart infrastructure tends to present smarter and safer solutions through the incorporation of low-cost sensors and ML models [11]. Manual supervision is not feasible, as monitoring all the workers in a site is impossible and is tiresome.

Understanding the pressing requirement, this article aims to present a ML powered IoT based safer solution to predict accidents in construction sites by employing several sensors. The proposed work extracts sensor features to track worker's and environmental conditions for accident prediction by ML model, while alerting the managerial team. This work involves multimodal sensing activity that combines physiological, locational and environmental data to present a proactive accident detection scheme by ML model. Explainable Artificial Intelligence (XAI) is integrated to support better interpretability of reasons behind accident prediction. The major contributions of the work are listed as follows.

- This work considers over 50,000 records, which are obtained from real-time sensors for ensuring construction safety.
- Raspberry Pi and IoT enabled sensors are utilized to collect physiological, locational and environmental data.
- Accident prediction model is proposed with better accuracy and generalization rates.
- The outcomes of the proposed model are enabled with XAI for better interpretability and understandability.
- Several performance measures are employed to evaluate the performance of the proposed model.

The remaining content of this article is organized in the following format. Section 2 studies the recent literature concerning workers safety in construction industry. The proposed accident prediction model is elaborated in section 3 and the performance of the proposed work is evaluated in section 4. Finally, the article is concluded in section 5.

REVIEW OF LITERATURE

This section reviews the existing literature based on workers safety and accident prediction frameworks.

Received: August 08 ,2025

International Journal of Applied Mathematics

Volume 38 No. 4 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

The work proposed in [12] conducts a review on accidents happened in construction sites, where workers are trapped between vehicles, machinery or structures. This work cites that the major causes of accidents include equipment failure, structural failures, trench collapse and so on. This work offers a global perspective by focusing on Malaysian case studies. In [13], a graph theory and network analysis are carried out to detect the interlinks in the safety factors of construction industry. Initially, a construction co-occurrence network is built with risk factors and the major hazards are detected using centrality metrics. Louvain algorithm is applied to environmental, incidental and demographic risks for clustering. The air quality and temporal factors are finally considered to be the key nodes. However, this work is based on historical data and dynamic data or IoT framework is not employed. The scalability of the work is not considered as well.

The authors of [14] present a worksite safety surveillance system for about 26 projects in India. This work employs a digital platform being handled by safety observers, who are trained to categorize hazards. Additionally, a predictive tool can estimate the likelihood of adverse outcomes such as fatalities. Though safety violations can easily be figured out this model, it relies on human observers. Besides, this work is not scalable, as it is labor intensive. Additionally, cost analysis has not been done.

The work in [15] presented a smart Personal Protective Equipment (PPE) by utilizing sensors into gloves and vests to track the worker's location, altitude, temperature and hand pressure. The collected data is uploaded to the real-cloud and a mobile app is developed. The goal of this work is to prevent fall, entry into high-risk zones by alerting supervisors and other kinds of injuries. This work ensures scalability, real-time application with cloud integration, but it suffers from hardware constraints and absence of ML framework for predictive assessment.

In [16], a computer vision-based ML system is presented to detect fall detection on construction sites with the footage of surveillance camera. The video streams are processed and detects unstable posture, sudden movement trajectory and balance shifts. Bounding boxes and skeletal modelling are used to track motions and the temporal features are extracted. ML models are trained and the supervisors are alerted. This work presents an alarming proactive model without any sensors or manual intervention, however it may suffer from environmental conditions that could affect video quality, annotation dependency and this model is not personalized.

A quantitative study with 165 construction professionals is surveyed in Nigeria [17], to examine wearable safety devices with respect to health and safety. This work conclude that wearable sensors help in mitigating hazards, sensing environment, tracking workers and productivity. On the other hand, it involves high deployment charges, technical and infrastructure issues and environmental constraints.

In [18], the recent research trends of safety provisioning insmart construction industry are studied. This work provides comprehensive study about safety technologies under six major topics for theoretical analysis. A job demand and resource model with resource conservation theory is presented in [19]. This

Received: August 08 ,2025

Volume 38 No. 4 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

work considers about 1,502 construction firms in Korea by employing hierarchical regression. However, this work lacks utilization of current technologies.

The financial costs incurred by Chinese construction projects are reviewed in [20] by employing an ensemble regression model. This work compares the performances of different classifiers such as Random Forest (RF), XGBoost and NGBoost with SHAP analysis upon 1606 accident records. However, this work is meant for cost prediction and not for safety. In [21], the authors presented an integrated real-time vision system to localize the workers, performsemantic segmentation and detect static with dynamic hazards. However, this work involves infrastructure dependency with surveillance cameras and computational complexity.

A real-time computer visionbased helmet detection system is presented in [22] by employing YOLOv5x. The proposed model is trained on 5000 images and proves better detection accuracy. Nevertheless, thiswork focusses only on helmets and not other wearables. The employment of YOLOv5x involves computational complexity. In [23], frequency analysis and analytic hierarchy process are utilized to detect the important factors leading to fatal falls. However, this work does not support IoT integration and scalability.

The work in [24] presented a visual warning system that is based on digital twin, deep learning and mixed reality. This work is tested on three quasi site scenarios for visualizing the hazard, yet this work involves cost and technical complexities. In [25], a comprehensive literature review is carried out on safety barriers concerning architecture, engineering and construction industries. Content analysis is performed under four classes such as management, culture, behavior and awareness.

In [26], key pattern recognition is carried out on construction industry of China. The key patterns of this work include type of accident, location, timing and trend prediction. However, this work is not predictive and performs data aggregate analysis. The work proposed in [27] reviews the occupational safety and health of Indian construction workers. In [28], a bibliometric review of research trends in construction safety is presented on scopus publications. Initially, seven different clusters in construction industry are formed to perform bibliometric analysis upon scopus database.

Motivated by these existing works, this work attempts to acquire multi-modal real-time data with IoT sensors with cloud integration powered by ML accident predictions. Dashboard based alerts are also a part of this work, in order to have better usability.

PROPOSED ML POWERED IOT INTEGRATED ACCIDENT PREDICTION SYSTEM FOR CONSTRUCTION SITES

The main goal of this work this is to present a IoT-cloud integrated ML technique trained accident prediction system with multimodal sensor data fusion. The key phases of this work involve data acquisition with sensors, data pre-processing, data storage, accident prediction, communication and visualization. All these phases explained in the following sections.

Received: August 08 ,2025

International Journal of Applied Mathematics

Volume 38 No. 4 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

3.1 Data Acquisition with Sensors with Raspberry Pi 4

This work tends to collect real-time data from sensors with Raspberry Pi 4 model B, which is known for its processing ability and its ability to work with IoT. Additionally, it supports a wider range of programming languages such as Python, Ruby, C, C++ and so on. It offers a powerful processor with better clock speed. This work attempts to extract physiological, environmental and locational data from sensors, which are then stored in cloud server and processed by ML algorithm for prediction of accidents in construction sites. All the utilized sensors are summarized below and shown in figure 2.

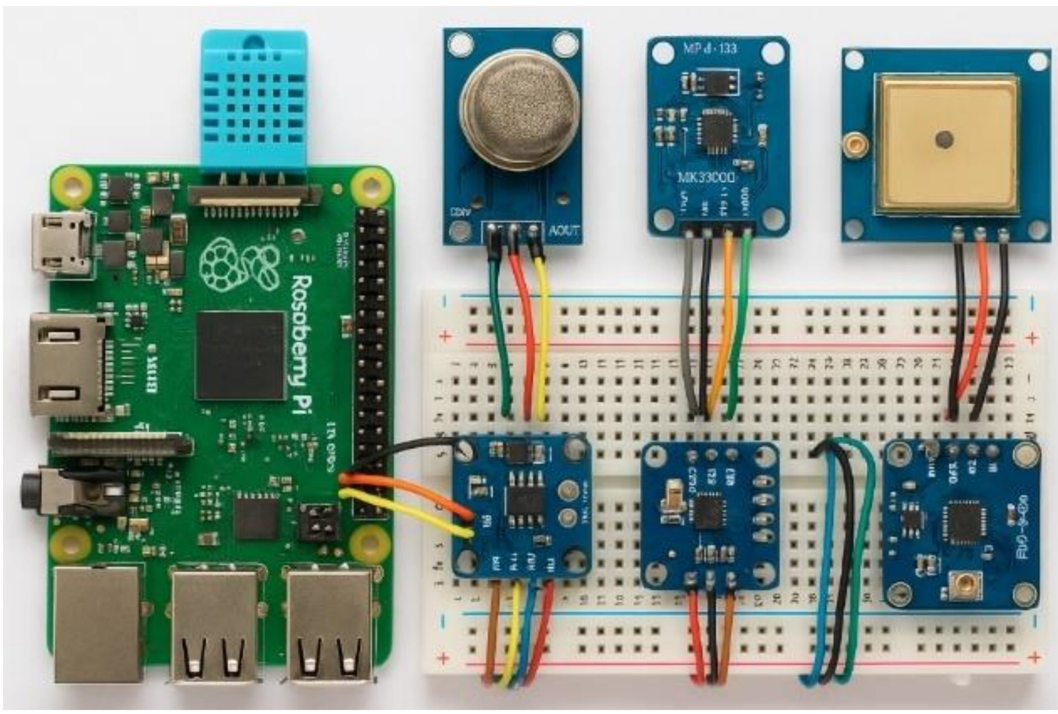

Fig.2. Raspberry Pi with sensors for accident prediction

3.1.1 Physiological data collection

The major physiological data required to predict worker safety are pulse rate, blood oxygen saturation and movement patterns. The pulse rate is measured in Beat Per Minute (BPM). With increased or decreased pulse rate, it is simple to detect worker's health condition instantly. Oxygen saturation in blood is measured by , where the normal range is in between 95 to 100%. In case of any gas exposure, fatigue or respiratory disorders, the falls down. The movement patterns are captured to know the motion and orientation that involves speed, direction, fall detection, rotation or sudden movement change. All the collected data can be combined to make a decision. For instance, high BPM+low may be the result of any gas exposure. High BPM+accelerometer fall can be an injury due to fall. Accelerometer fall + low + low BPM can indicate some critical issue. The reasons for the choice of these sensors are

Received: August 08 ,2025

International Journal of Applied Mathematics

Volume 38 No. 4 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

their invasiveness and continuous real-time data collection. This work utilizes MAX30102 (heart rate+and MPU6050 (Gyroscope/accelerometer) respectively.

3.1.2 Environmental data collection

Environmental data is necessary for predicting worker safety, as they help in detecting any harmful conditions, unsafe working spots and so on. The sensors employed to collect environmental data are temperature/humidity and quality of air to detect any gas leakage. When high temperature is sensed with humidity can indicate heat stress, abnormal gas levels can be detected. The sensors employed to collect environmental data are DHT22 (temperature/humidity) and MQ-135 (air quality) respectively.

3.1.3 Locational data collection

The locational data helps to collect the latitudinal and longitudinal coordinates, which can also be utilized to map zonal risks, in addition to worker tracking. With this data, the hazardous zones can be highlighted and the workers can be prevented to enter. Besides, in case of any emergency, the rescue operations can be carried out without any hassles. NEO-6M is utilized to collect the data.

The real-time data is collected and are logged into a ‘dataset.csv’ file for manipulation. This dataset consists of a total of 50,000 records with 200 unique workers. The timeframe of the data collection is between June 25 to June 28. The logs are updated for every 5 seconds. The dataset is composed of the columns such as timestamp, worker_identifier, department, zone, temperature, humidity, air quality, heart rate, , acceleration in x,y,z, latitude and longitude.The dataset is visualized with 500 records in figure 1.

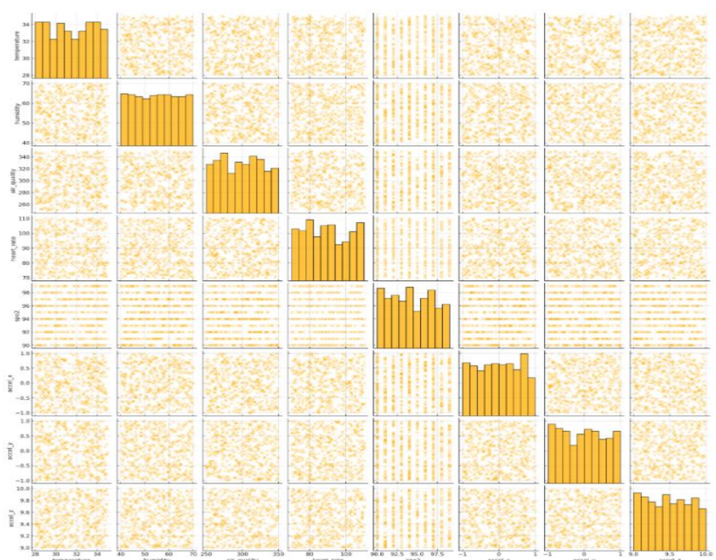

Fig.1. Dataset visualization

Received: August 08 ,2025

International Journal of Applied Mathematics

Volume 38 No. 4 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Hence, the real-time data is collected and stored on a local machine to enjoy the benefits low latency without any network delays, cost efficiency, ML compatibility, secure when compared to cloud storage. Understanding the benefits of local processing, this work is processed on a stand-alone system. However, in order to combat against memory or storage overhead, this work proposes to use local storage with periodical cloud sync. Cloud storage is necessary for backup and to avoid single point of failure. Thus, this work enjoys the merits of both local and cloud storage.

3.2 Data Pre-processing

As soon as the data collection and storage are done, it is necessary to pre-process the raw data, so as to make it suitable for ML based processing.In order to achieve reasonable accuracy and reliability, effective pre-processing is of utmost importance. The main goal of pre-processing is to transform the raw data to structured and normalized data, which is the basic requirement of ML framework. The key steps involved in this phase are timestamp parsing, missed value handling, datatype conversion, outlier detection, normalization, label creation and dataset balancing. The dataset before and after normalization is shown in figure 2.

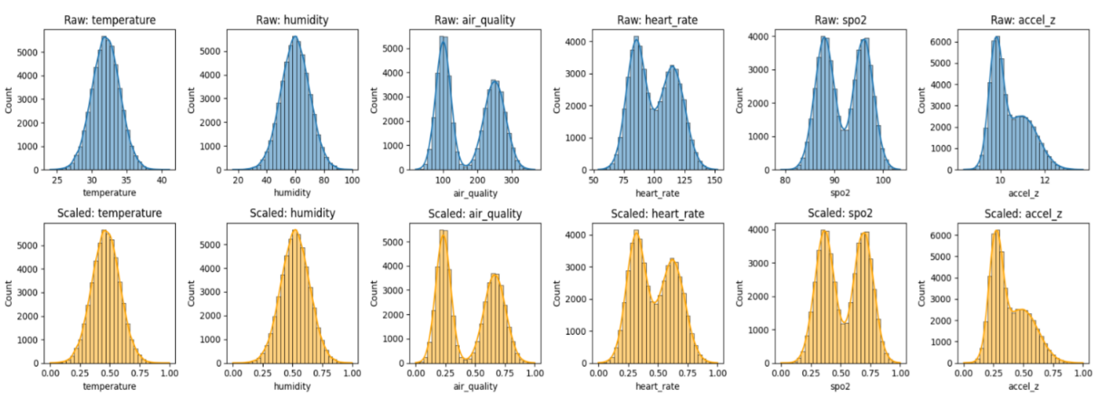

Fig.2. Visualizing normalized dataset

- Timestamp parsing: This step helps in aligning the time-series by converting the timestamps to datetime objects and the data is chronologically sorted for better analysis.
- Missed value handling: This step is to handle and manage the missed readings from the sensors. This issue is handled by performing data interpolation, in order to ensure data completeness. Data interpolation helps to maintain continuity in time-series data.
- Datatype conversion: The dataset involves both numerical and textual data. The fields temperature, SpO_2 and air quality are in float format. The categorical fields such as worker_id, zone, department are numerically converted. This step is crucial for training an ML model.
- Outlier detection: Outliers may present in the dataset and are tackled by z-score technique for improving the robustness of the prediction model.It is calculated by eqn.(1).

$$Z_s = \frac{x_i - \mu}{\sigma} \qquad (1)$$

Received: August 08 ,2025

International Journal of Applied Mathematics

Volume 38 No. 4 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

In the above equation, x_i is the value of feature, μ and σ indicate the mean and standard deviation respectively. A value is considered as outlier, when a data point is highly deviated from its neighborhood points.

- Normalization: The numerical values available in the dataset should be normalized for uniform processing and is performed by min-max normalization. Normalization is necessary to understand features under varied scales and ML models need it to gain knowledge. It is computed by the following formula.

$$\overline{x} = \frac{x - x_{min}}{x_{max} - x_{min}};\ \overline{x} \in [0,1] \tag{2}$$

In the above equation, x, x_{min} and x_{max} are the original, minimum and maximum values of the features respectively. This normalization technique does not change the distributional shape and preserves the distance between the points.

- Label creation: Label creation is the tedious step, which could impart knowledge to the ML model about safe and unsafe situations for predicting accidents. This work involves two classes such as 'safe' and 'unsafe/alert'.
- Dataset balancing: The dataset is quite unbalanced with more of safe samples, which may introduce biased classification towards majority samples under a specific class. Hence, the dataset needs to be balanced and is attained by Synthetic Minority Over-Sampling Technique (SMOTE), which distributes the equal distribution of samples across two classes. Both the classes are populated with 25,000 records respectively.

A well-pre-processed dataset helps in attaining better accuracy and generalization rates of a prediction system and supports in training phase of a ML algorithm.

3.3 Feature Extraction

As soon as the data is pre-processed, the key features are extracted from the data. This phase helps in transforming the raw data to informative features, which a ML model can process and learn. This work extracts three kinds of features such as statistical, temporal and frequency domain features, as explained below.

3.3.1 Statistical features

Statistical features are measures that are captured from sensors, which are found to be basic yet powerful for ML model. The statistical features being considered are mean, min/max, Standard Deviation (SD), range, rolling mean, skewness and kurtosis. The features along with the formula are shown in table 1.

Table 1. Statistical features

Sno	Feature	Formula	Description
1	Mean	$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$	x_i- Data values of time window n- number of values in window

Received: August 08 ,2025

International Journal of Applied Mathematics

Volume 38 No. 4 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

2	Min/Max	$mn(x) = \min\{x_1, x_2, \dots, x_n\}$ $mx(x) = max\{x_1, x_2, \dots, x_n\}$	x_i- Data values
3	SD	$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2}$	μ-Mean σ- SD
4	Range	$R = \max(x) - \min(x)$	x- Data values
5	Rolling mean	$RM_t = \frac{1}{k}\sum_{i=t-k+1}^{t} x_i$	k- window size x_i- Value at time t
6	Skewness	$S = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^3}{\sigma^3}$	x_i- Data values of time window μ-Mean σ- SD
7	Kurtosis	$K = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^4}{\sigma^4}$	x_i- Data values of time window μ-Mean σ- SD

With mean, the central value for a time window can be obtained. SD measures the data variation from the mean, such that instability can be figured out easily. Min and max values find the smallest and greatest value. Rolling window features are meant for operating over a specific count of recent samples. The real-time trends can be understood with this. The range values help in detecting the patterns of the readings. The asymmetry of data distribution is computed by skewness and the extreme fluctuations of readings is computed by kurtosis.

3.3.2 Temporal features

These features are time related and can also be known as time-windowed values. These features can extract frequency of events in a specific time window. As time is considered, end-of-day fatigue can be detected with different features.Data with respect to time is captured to track the trends, delays and periodicity. Some of the features are time of day, rolling window based, lagging features, event duration and rate of change.The formula for computing different features and their descriptions are presented in table.

Table. Temporal features

No	Feature	Formula	Description
1	Hour based	$Hour(t) = ExtractHour(TS_t)$ $Shift(t) = \begin{cases} Early\ 6 \le Hour < 10 \\ Mid\ 10 \le Hour < 14 \\ Late\ 14 \le Hour \le 18 \end{cases}$	TS-Timestamp

Received: August 08 ,2025

International Journal of Applied Mathematics

Volume 38 No. 4 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

2	Rolling mean	$RM_t^{(k)} = \frac{1}{k} \sum_{i=t-k+1}^{t} x_i$	k-Window size t-present time index
3	Rolling SD	$RSD_t^{(k)} = \sqrt{\frac{1}{k} \sum_{i=t-k+1}^{t} (x_i - \mu_t)^2}$	k-Window size t-present time index μ_t-Rolling mean over last 'k' points
4	Lag features	$x_{t-1}, x_{t-2}, \dots, x_{t-k}$	
5	Work duration	$Worktime_t = TS_t - shifttime_{worker}$	TS – Timestamp
6	Rate of change	$\Delta x_t = x_t - x_{t-1}; \frac{dx}{dt} = \frac{x_t - x_{t-1}}{\Delta t}$	Δx_t- Rate of change in variable

3.3.3 Frequency based features

Frequency space transformation is carried out to extract these sorts of features and these features are meant for detecting repetitive vibration patterns. These features are beneficial for motion-based sensors such as accelerometer. The time series data are transformed from time to frequency domain with the help of Fast Fourier Transform (FFT).Table 3 shows the frequency based features.

No	Feature	Formula	Description
1	Dominant Frequency	$F_{dmnt}(k) = \arg \max_k (\lvert X_k \rvert)$	X_k- Frequency component by FFT $\lvert X_k \rvert$ - Magnitude of complex coefficient
2	Energy	$E = \sum_{k=0}^{N-1} \lvert X_k \rvert^2$	$\lvert X_k \rvert^2$- Power of each frequency component
3	Entropy	$H = -\sum_{k=0}^{N-1} p_k . \log_2(p_k)$ $p_k = \frac{\lvert X_k \rvert^2}{\sum_{j=0}^{N-1} \lvert X_j \rvert^2}$	p_k - Normalized power at k^{th}frequency
4	Ratio of high-to-low band power	$HL_{ratio} = \frac{\sum_{f_k \in F_{high}} \lvert X_k \rvert^2}{\sum_{f_k \in F_{low}} \lvert X_k \rvert^2}$	F_{high}-High frequency F_{low}- Low frequency

The periodical patterns or vibrations can be easily detected, which is not possible with raw data. This work extracts the FFT of each axis (x,y,z) and the dominant frequency can be utilized to detect jump or fall. The highest amplitude is measured by the dominant frequency. Certain other features such as energy, entropy and ratio of high-to-low band power are extracted for better analysis. The spectral energy computes the total signal power in the frequency domain, while the randomness of a frequency

Received: August 08 ,2025

Volume 38 No. 4 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

distribution is quantified by entropy. The energy levels in higher and lower bands are compared to detect rapid movements.

3.4 Accident Prediction by XGBoost

With the extracted features, the eXtreme Gradient Boosting (XGBoost) is trained. The XGBoost is an efficient and scalable model that is based on gradient boosting decision trees that can prove effective performance over tabular data. The concept of XGBoost is to build a strong prediction model, while combining several weak decision trees during the training process with an objective to minimize the loss function by employing II order Taylor approximation. The objective function of the XGBoost is given by equation 3.

$$OBJF^{(t)} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \omega(f_t) \quad (3)$$

$$\omega(f) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2 \quad (4)$$

Here, $f_t(x) \in \mathbb{T}$, which is the decision tree at step t, T and w_j indicate the count of leaves and weight of leaf j respectively. γ and λ are regularization parameters. The $OBJF$ is optimized by the following equation.

$$OBJF^{(t)} \approx \sum_{i=1}^{n} \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t(x_i)^2\right] + \omega(f_t) \quad (5)$$

The II order gradient is given by

$$h_i = \frac{\partial^2 l(y_i, \hat{y}_i^{(t-1)})}{\partial \hat{y}_i^{(t-1)^2}} \quad (6)$$

Each tree divides the data into several leaves, as follows. Let leaf j involves several instances I_j. $G_j = \sum_{i \in I_j} g_i$; $H_j = \sum_{i \in I_j} h_i$.

The optimal leaf weight (lw_j) and total gain are computed by

$$lw_j^* = -\frac{G_j}{H_j + \lambda} \quad (7)$$

$$Gain_j = -\frac{1}{2} \sum_{j=1}^{T} \frac{G_j^2}{H_j + \lambda} + \gamma T \quad (8)$$

The initial prediction is started with $\hat{y}^{(0)}$. For every iteration, gradients and h_i are computed and a decision tree is built to reduce the regularization loss. The prediction is updated by the following equation.

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + f_t(x_i) \quad (9)$$

The major reasons for employing XGBoost classifier to predict accidents are its learning speed, overfitting issue handling, better interpretation ability and effective regularization.

RESULTS AND DISCUSSION

The performance of the proposed accident prediction system is evaluated on a stand-alone system with 16 GB RAM on MATLAB2021A version. The train/test ratio of this work is set as 80/20. The efficiency of this work is proven by analyzing it with standard performance measures such as accuracy, precision, recall, F-measure rates and confusion matrices. The analysis is carried out in different aspects such as to highlight the performance of combination of different feature sets such as statistical, temporal and frequency, in contrast to individual feature set respectively. Different classifiers such as Support Vector Machine (SVM) and Random Forest (RF) are employed and their performances are showcased.

Received: August 08 ,2025

International Journal of Applied Mathematics

Volume 38 No. 4 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

$$\partial = \frac{T_{Pos}+T_{Neg}}{T_{Pos}+T_{Neg}+F_{Pos}+F_{Neg}} \quad (10)$$

$$\wp = \frac{T_{Pos}}{T_{Pos}+F_{Pos}} \quad (11)$$

$$\Re = \frac{T_{Pos}}{T_{Pos}+F_{Neg}} \quad (12)$$

$$\mathcal{F} = 2 \times \frac{(\wp \times \Re)}{(\wp + \Re)} \quad (13)$$

The proportion of positive samples that were accurately recognized in comparison to the total count of samples that were indeed positive is represented by $\Re$, which is frequently referred to as the sensitivity or true positive rate. The harmonic mean of $\wp$ and $\Re$ is used to ascertain the value of $\mathcal{F}$.

Table 4. Performance Results

Model	∂ (%)	$\wp$ (%)	$\Re$(%)	$\mathcal{F}$(%)
Statistical Features (SF) + RF	77.5	76.19	80	77.95
Temporal features (TF) + RF	72.5	71.43	75	73.14
Frequency domain features (FF) + RF	75.3	72.5	79.4	75.73
Statistical Features (SF) + XGBoost	89.38	87.50	90.91	89.13
Temporal features (TF) + XGBoost	83.13	81.18	86.25	83.65
Frequency domain features (FF) + XGBoost	90	86.36	95	90.49
SF+TF+FF+RF	97.5	96.34	98.75	97.54
SF+TF+FF+XGBoost	**99.5**	**99.34**	**99.75**	**99.54**

The results are shown in graphical format as follows.

Received: August 08 ,2025

International Journal of Applied Mathematics

Volume 38 No. 4 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

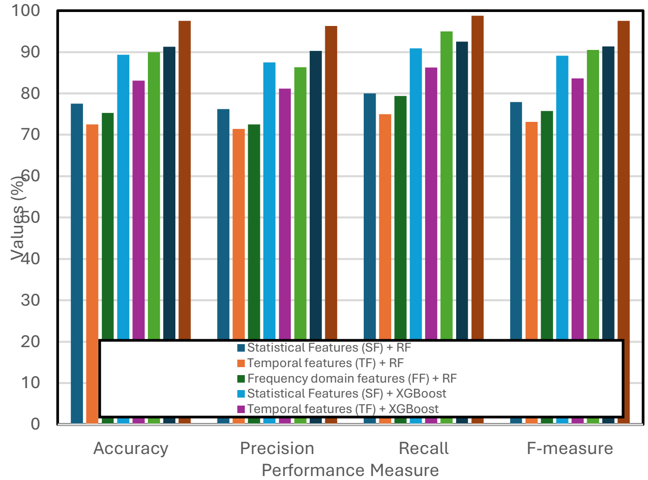

Fig.3. Performance analysis

The performance of the proposed work is proven by the experimental results. The combination of statistical, temporal and frequency domain based features prove better performance with 99.5% accuracy rates. The XGBoost classifier performs well than RF classifier, owing to its effective regularization and predicting ability with regular updations.

CONCLUSION

This article presents a XGBoost powered accident prediction system to ensure construction worker's safety using IoT sensors. Physiological, environmental and locational data are all collected by sensors with Raspberry Pi 4 and a dataset is built with 15 columns including label. The dataset contains 50,000 records for 200 workers collected in three days of time. Periodical cloud data storage is employed by this work, in order to avoid single-point-of-failure and massive data storage. Yet, all the data processing is done on a stand-alone computer. The data pre-processing is initially done to normalize, detect outliers and to balance the dataset. The pre-processed dataset is then processed to extract statistical, temporal and frequency based features. XGBoost classifier is employed to learn the feature base and to carry out accident prediction under two classes, safe and alert. The performance of the proposed work is evaluated and the results show the efficacy of the proposed approach. In future, this work can be carried out in real-time with wearable sensors on a live construction site.

References

[1] Karthick, S., Kermanshachi, S., Pamidimukkala, A., &Namian, M. (2023). A review of construction workforce health challenges and strategies in extreme weather conditions. International journal of occupational safety and ergonomics, 29(2), 773-784.

[2] Takala, J. (2020). Burden of injury due to occupational exposures. Handbook of disability, work and health, 105-126.

Received: August 08 ,2025

Volume 38 No. 4 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

[3] Almaskati, D., Kermanshachi, S., Pamidimukkala, A., Loganathan, K., & Yin, Z. (2024). A Review on Construction Safety: Hazards, Mitigation Strategies, and Impacted Sectors. Buildings, 14(2), 526.

[4] https://www.industriall-union.org/over-400-workers-killed-in-india-in-due-to-workplace-safety-lapses

[5] Shamsuddin, K. A., Ani, M. N. C., Ismail, A. K., & Ibrahim, M. R. (2015). Investigation the Safety, Health and Environment (SHE) protection in construction area. International Research Journal of Engineering and Technology, 2(6), 624-636.

[6] Awolusi, I., Marks, E., & Hallowell, M. (2018). Wearable technology for personalized construction safety monitoring and trending: Review of applicable devices. Automation in construction, 85, 96-106.

[7] Jia, M., Komeily, A., Wang, Y., & Srinivasan, R. S. (2019). Adopting Internet of Things for the development of smart buildings: A review of enabling technologies and applications. Automation in construction, 101, 111-126.

[8] Gupta, B. B., &Quamara, M. (2020). An overview of Internet of Things (IoT): Architectural aspects, challenges, and protocols. Concurrency and Computation: Practice and Experience, 32(21), e4946.

[9] Al-Garadi, M. A., Mohamed, A., Al-Ali, A. K., Du, X., Ali, I., &Guizani, M. (2020). A survey of machine and deep learning methods for internet of things (IoT) security. IEEE communications surveys & tutorials, 22(3), 1646-1685.

[10] Chen, X., Chang-Richards, A. Y., Pelosi, A., Jia, Y., Shen, X., Siddiqui, M. K., & Yang, N. (2022). Implementation of technologies in the construction industry: a systematic review. Engineering, Construction and Architectural Management, 29(8), 3181-3209.

[11] Talebkhah, M., Sali, A., Gordan, M., Hashim, S. J., &Rokhani, F. Z. (2023). Comprehensive review on development of smart cities using industry 4.0 technologies. IEEE Access, 11, 91981-92030.

[12] Darda'u Rafindadi, A., Kado, B., Gora, A. U. M., Dalha, I. B., Haruna, S. I., Ibrahim, Y. E., & Ahmed Shabbir, O. (2025). Caught-In/Between Accidents in the Construction Industry: A Systematic Review. Safety, 11(1), 12.

[13] Kumi, L., Jeong, J., Jeong, J., Son, J., & Mun, H. (2025). Network-based safety risk analysis and interactive dashboard for root cause identification in construction accident management. Reliability Engineering & System Safety, 256, 110814.

[14] Muralidhar, G., Vinoth, A., Uniyal, U., & Sen, K. N. (2025). Enhancing Construction Safety Through Workplace Surveillance and Systematic Follow-Ups. American Journal of Industrial Medicine.

[15] Rashidi, A., Woon, G. L., Dasandara, M., Bazghaleh, M., &Pasbakhsh, P. (2025). Smart personal protective equipment for intelligent construction safety monitoring. Smart and Sustainable Built Environment, 14(3), 835-858.

[16] Liu, X., Xu, F., Zhang, Z., & Sun, K. (2025). Fall-portent detection for construction sites based on computer vision and machine learning. Engineering, Construction and Architectural Management, 32(3), 1499-1521.

[17] Ibrahim, K., Simpeh, F., & Adebowale, O. J. (2025). Benefits and challenges of wearable safety devices in the construction sector. Smart and Sustainable Built Environment, 14(1), 50-71.

[18] Seo, H. J., & Yoon, Y. G. (2025). Analyzing Research Trends in Smart Construction Safety: A Topic Modeling Approach. Buildings, 15(4), 520.

[19] Seo, H. J., Rhee, S. Y., & Yoon, Y. G. (2025). Enhancing Worker Safety Behaviors Through the Job Demands–Resources Approach: Insights from the Korean Construction Sector. Buildings, 15(3), 486.

Received: August 08 ,2025

Volume 38 No. 4 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

[20] Xia, X., Xiang, P., Khanmohammadi, S., Gao, T., &Arashpour, M. (2024). Predicting safety accident costs in construction projects using ensemble data-driven models. Journal of Construction Engineering and Management, 150(7), 04024054.

[21] Jeelani, I., Asadi, K., Ramshankar, H., Han, K., & Albert, A. (2021). Real-time vision-based worker localization & hazard detection for construction. Automation in Construction, 121, 103448.

[22] Hayat, A., & Morgado-Dias, F. (2022). Deep learning-based automatic safety helmet detection system for construction safety. Applied Sciences, 12(16), 8268.

[23] Rafindadi, A. D. U., Napiah, M., Othman, I., Mikić, M., Haruna, A., Alarifi, H., & Al-Ashmori, Y. Y. (2022). Analysis of the causes and preventive measures of fatal fall-related accidents in the construction industry. Ain Shams Engineering Journal, 13(4), 101712.

[24] Wu, S., Hou, L., Zhang, G. K., & Chen, H. (2022). Real-time mixed reality-based visual warning for construction workforce safety. Automation in Construction, 139, 104252.

[25] Maliha, M. N., Abu Aisheh, Y. I., Tayeh, B. A., & Almalki, A. (2021). Safety barriers identification, classification, and ways to improve safety performance in the architecture, engineering, and construction (AEC) industry: Review study. Sustainability, 13(6), 3316.

[26] Xu, Q., & Xu, K. (2021). Analysis of the characteristics of fatal accidents in the construction industry in China based on statistical data. International journal of environmental research and public health, 18(4), 2162.

[27] Samanta, S., &Gochhayat, J. (2023). Critique on occupational safety and health in construction sector: An Indian perspective. Materials Today: Proceedings, 80, 3016-3021.

[28] Akinlolu, M., Haupt, T. C., Edwards, D. J., &Simpeh, F. (2022). A bibliometric review of the status and emerging research trends in construction safety management technologies. International Journal of Construction Management, 22(14), 2699-2711.

Received: August 08 ,2025