

# A Study for IoT Data Analysis on Multi Use Old Building Inspection

Chang-Won Yoon<sup>1,2</sup>, Jun-Ho Huh<sup>1,2,3\*</sup>

## Abstract

It is important to maintain the safety of buildings to ensure that they are in good condition in order to prevent the risks posed by aging buildings. Recently, drones have been used to diagnose the safety of buildings using various sensors, but changes in the safety of buildings require not only external changes in condition, but also evaluation of internal factors such as vibration, tilt, and subsidence, which are difficult for workers to check with the naked eye. In this paper, we summarized the isolation forest algorithm process and formulas, and conducted experiments on IoT data analysis by applying the isolation forest, extended isolation forest, PCA and Isolation algorithms in order. We analyzed the condition change of the structure based on IoT data of vibration, tilt, and altitude change to check the condition change of the building. In particular, considering previous study's method of directly inputting individual sensor data's multi-dimensional vectors into extended isolation forest, we used the risk score reflecting the relative contributions of the tree variables, which are gyro, tilt, and altitude data. This allowed us to visualize risk levels as a single time-series graph, facilitating the identification of change intervals and patterns. This visualization method can be used to build a system for real-time diagnosis of old buildings based on the condition change and defeat factors of the structure based on the accumulated data in the section where the score rose sharply. This paper demonstrates the applicability of real-time anomaly detection systems and it is expected to enable more precise safety diagnostics by comprehensively utilizing IoT data with drone-based inspections of buildings.

**Key Words:** Old Building Inspection, IoT Sensor, IoT Data Preprocessing, Isolation Forest, Risk Score.

## I. INTRODUCTION

In Republic of Korea, the proportion of old buildings over 30 years old nationwide is rapidly increasing every year, with 36.5% in 2017, 37.1% in 2018, and 37.8% in 2019, and the number of disasters and accidents caused by weather disasters is increasing. In particular, Busan has the highest proportion of old buildings in the country with 54.3% of all buildings as of the end of December 2019, and the

proportion of buildings over 20 years old is 75%, which is much higher than the national average [1-3] (Fig. 1).

In order to prevent risks caused by old facilities in advance, management to maintain the condition of facilities at an appropriate level is necessary, and until now, safety inspections of aging buildings have been primarily conducted visually by workers, which has led to reduced reliability of diagnostic results and increased potential risks. The use of drones for safety inspections of various aging buildings has improved on-site work efficiency and worker safety.

Recently, drones have been equipped with individual sensors such as multispectral sensors, RGB, LiDAR, etc. to acquire data on cracks, disappearance, etc. on the outside of buildings, and safety diagnosis is carried out based on this information. However, while drone-collected data evaluates visual changes in aging buildings, such as cracks and



Fig. 1. Cases of damage caused by structures.

Manuscript received April 07, 2025; Revised May 05, 2025; Accepted May 07, 2025. (ID No. JMIS-25M-04-006)

Corresponding Author (\*): Jun-Ho Huh, +82-51-401-4371, 72networks@kmou.ac.kr

<sup>1</sup>Department of Data Informatics, National Korea Maritime and Ocean University, Busan 49112, Republic of Korea, niceyun@empas.com, 72networks@kmou.ac.kr

<sup>2</sup>Department of Interdisciplinary Major of Ocean Renewable Energy Engineering, National Korea Maritime and Ocean University, Busan 49112, Republic of Korea, niceyun@empas.com, 72networks@kmou.ac.kr

<sup>3</sup>Sabbatical Year at School of Computer Science and Engineering, Yeungnam University, Gyeongsan 38541, Republic of Korea, 72networks@kmou.ac.kr

losses, it remains insufficient for assessing and analyzing internal displacements such as subsidence and vibrations.

Changes in building safety are not only influenced by external conditions but also by internal factors such as vibration, tilting, and subsidence, which are difficult for workers to detect visually. However, these factors are often overlooked in building safety diagnostic evaluations due to the limitation that they are difficult to detect immediately on-site.

Additionally, to confirm changes in the condition of buildings, it is necessary to collect data periodically over a certain period to manage the factors causing the changes. However, periodic data collection using drones has limitations in terms of the efficiency of building management and operation.

This paper proposes a method for identifying changes in building conditions by utilizing IoT sensor data, considering the limitations of existing diagnostic methods based on video data obtained from drones.

Over a seven-day period at a university-industry collaboration building, IoT sensor data was collected for vibration, tilt, altitude change. To achieve precise and stable monitoring of building conditions, it is necessary to evaluate and analyze IoT data that provides objective internal factor information, identify risk zones based on the Isolation Forest algorithm and risk scores, and comprehensively analyze the accumulated data on safety changes to develop a real-time diagnostic system.

## II. THEORETICAL BACKGROUND

### 2.1. System Configuration with IoT Sensors

Until recently, various studies utilizing IoT for old buildings have been actively promoted. In order to utilize IoT as an automated management technology for state-managed facilities, the IoT network-based facility remote diagnosis and automation, advanced network maintenance, and unmanned facility repair and reinforcement technology were applied to establish a comprehensive IoT-based maintenance plan in order to utilize the passive facility maintenance system as an active maintenance system [4].

In addition, an application technique that can efficiently combine spatial information such as semi-permanent major facilities and design and construction drawings by tracking the deformation of micro-displacement using GNSS/USN /IoT has been proposed [5], and the facility inspection method can be reduced in cost and improved in objectivity if the facility inspection video is used as a reference video and the inspection method is based on drone video centered on automatic processing [6].

On the other hand, a conceptual model of an S-LCC plat-

form for efficient operation of facility systems that automatically analyzes and collects electricity consumption data using IoT was derived [7], and a method for analyzing and processing big data to perform real-time, always-on safety checks on bridges, which are road facilities, through an IoT platform and support rapid decision-making by facility managers was presented [8].

By fully automating the concrete crack detection framework through machine learning, it is possible to maintain facilities more effectively and efficiently than the existing expert-oriented facility maintenance [9], and to apply IoT devices using IoT crack meters to actual sites, it is necessary to establish an LTE communication network and develop a large-capacity battery in addition to its technical capabilities [10].

### 2.2. Monitoring Safety Changes in Building

Based on a paper on embedding PCB-based wireless inductive coupling corrosion potential sensors in reinforced concrete and soil to measure the degree of corrosion and long-term monitoring as a way to monitor safety changes in buildings, it was proposed to monitor buildings by interlocking IoT sensors and platforms [11], or to use a system and artificial intelligence algorithm that identifies and classifies six categories of data: building occupants, indoor environmental conditions, external environmental conditions, control systems and devices, equipment technology, and energy flow [12].

In addition, the possibility of using remote sensing systems as a tool to monitor asphalt road pavement conditions has been proposed [13], and studies have been conducted to detect safety changes in buildings with algorithms such as SVM and CNN models, parameter settings, and reported accuracy [14-18].

## III. IoT SENSOR BOARD FABRICATION AND PERFORMANCE EVALUATION

### 3.1. IoT Sensor Board Fabrication

The IoT sensor board used in this paper was designed and manufactured by the author using PCB to check the internal condition change factors of the building, and 20 pieces were manufactured, and considering that it is attached to the outside of the building, a battery and a solar panel were attached to the inside of the sensor board. In addition, a moisture-proof agent was sprayed on the case to prevent the penetration of rainwater.

To check the magnitude, tilt, and relative altitude of the building's vibration, the IoT sensor board is equipped with gyroscope, altitude sensor, and a geomagnetometer sensors, and the MCU uses ESP32-WROOM-32 microprocessor. It

Table 1. IoT sensor board specifications and images.

IoT sensor board specifications
- SENSOR: Gyro, Altimeter, Geomagnetometer
- MCU: ESP32-WROOM-32 (CPU: Xtensa® 32-bit LX6 microprocessors)
- SRAM: 520KB, FLASH: 8 MB
- POWER: DC24V → DC-DC (step down), solar cell → MPPT → rechargeable battery
- Sensor board size: 92.7×71 mm
- Protocol: MQTT

IoT sensor board prototype image



is configured to operate on top of the TCP/IP protocol using MQTT, an ISO standard publish-subscribe-based messaging protocol, to connect to remote locations with limited network bandwidth (Table 1).

### 3.2. IoT Sensor Board Performance test by KOLAS

In order to secure the objectivity of IoT data acquired from buildings, the prototype was tested for angular range, angular accuracy, acceleration accuracy, and data precision by requesting the Busan City \*\*\* public institution, and completed the KOLAS test results in January 2024.

Accuracy and precision are commonly used as performance metrics for sensors. Accuracy is a measure of whether a sensor consistently and accurately displays results that are true to the actual value, and precision is a measure of whether a sensor consistently displays results when measuring the same value multiple times.

Since the IoT sensor board produced in this paper aims to check the impact of internal condition changes in buildings, we

Table 2. IoT sensor board performance test indicators.

Performance	Unit	Test result
Angle range	Degree	±10 degrees
Angle accuracy	Degree	±0.01 degree
Acceleration accuracy	g	±0.01 g
Data precision	%F.S	±1%



Fig. 2. IoT sensor board performance test report (performance test [19]).

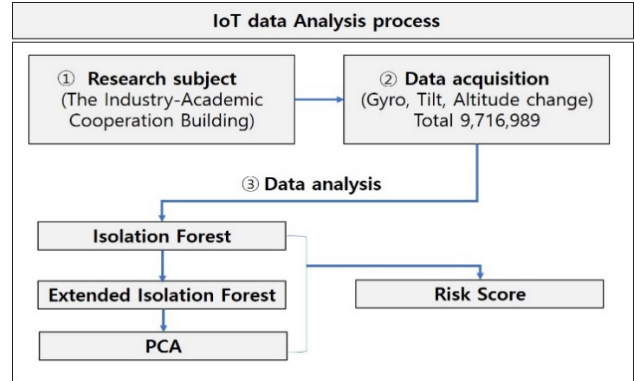


Fig. 3. IoT data analysis process.

adopted angle range, angle accuracy, acceleration accuracy, and data precision as performance indicators (Table 2, Fig. 2).

The entire data analysis process for anomaly detection and analysis from data collection, preprocessing, anomaly detection to risk score calculation is as shown in Fig. 3.

## IV. EXPERIMENTATION AND DATA COLLECTION

### 4.1. Research Subject

The Industry-Academic Cooperation Building of \*\*\* University in Busan was selected as the target building for the study. The building is used by many students and is suitable for the study of vibration, tilt, and settlement of the building itself, and it is easy to acquire data, so it was set as the target building of this study, and the location of the building is indicated by the blue square on Fig. 4.

### 4.2. Data Acquisition



Fig. 4. View of industry-academia cooperation center.

Table 3. IoT sensor board performance test indicators.

Data acquisition	Result
Total number of data	9,716,989
Start time	2024-04-02 18:32:21
End time	2024-04-08 23:59:59



Fig. 5. Installed on the rooftop of Research subject.

Table 4. IoT sensor board coordinates.

No.	Latitude	Longitude
1	35; 8; 43.4418000000004767	129; 0; 30.8332790000131354
2	35; 8; 43.2445199999929386	129; 0; 30.5017189999925904
3	35; 8; 43.2445199999929386	129; 0; 30.5017189999925904
4	35; 8; 43.2445199999929386	129; 0; 30.5017189999925904
5	35; 8; 43.2445199999929386	129; 0; 30.5017189999925904
6	35; 8; 43.2445199999929386	129; 0; 30.5017189999925904
7	35; 8; 43.2445199999929386	129; 0; 30.5017189999925904
8	35; 8; 44.3835600000020847	129; 0; 30.7976400000043178
9	35; 8; 44.3835600000020847	129; 0; 30.7976400000043178
10	35; 8; 43.9083600000013163	129; 0; 31.8290389999747276
11	35; 8; 43.76219900000043581	129; 0; 31.9424400000134483
12	35; 8; 43.76219900000043581	129; 0; 31.9424400000134483
13	35; 8; 43.6754389999987325	129; 0; 32.1548390000243671
14	35; 8; 43.5976800000061715	129; 0; 32.145120000010431
15	35; 8; 43.40615900000055375	129; 0; 32.1184789999970235
16	35; 8; 43.3892400000041079	129; 0; 32.0943589999806136
17	35; 8; 43.3989589999982428	129; 0; 32.1094799999846146
18	35; 8; 43.3989589999982428	129; 0; 32.1094799999846146
19	35; 8; 43.19771900000033673	129; 0; 30.9160789999878034
20	35; 8; 43.19771900000033673	129; 0; 30.9160789999878034

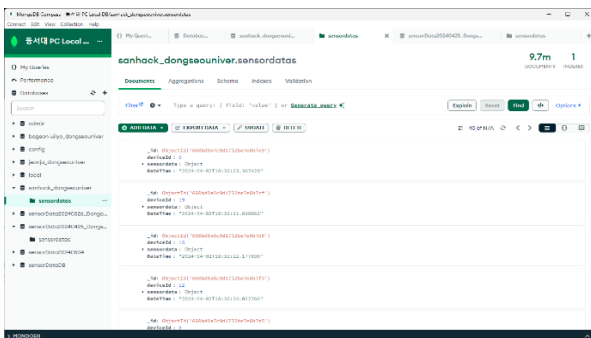


Fig. 6. IoT sensor data in MongoDB.

IoT sensors produced in this study were installed on the rooftop of the industry-academic cooperation center for a total of approx. 7 days from 2024.04.02. to 2024.04.08.

The total number of data acquired during this period was 9,716,989, which is much more real-world data than previous studies with thousands or tens of thousands of data points (Table 3). A total of 20 sensor boards were installed at a certain distance on the rooftop space of the Industry-Academia Cooperation Center, and the installation location of each sensor board is shown in Fig. 5 and Table 4 in terms of latitude and longitude coordinates. Each position value is automatically acquired from the sensor board.

The gyro, barometric pressure, and altimeter data generated by the IoT sensors were stored in real time by installing a local computer on the roof of the Industry-Academic Cooperation Center and utilizing Ethernet and Wi-Fi networks to minimize the loss of original data.

Fig. 6 shows that the data acquired from the IoT sensor board is stored in MongoDB on the desktop PC in the experimental environment.

## V. DATA ANALYSIS

### 5.1. Data Preprocessing and Feature Extraction

The data acquired from IoT sensors was exported from MongoDB to CSV format to reduce unnecessary conversion processes, and includes gyroscope data and barometric pressure sensor data. The data is recorded in milliseconds, which is characterized by the ability to detect instantaneous changes in the industry-academia cooperation center.

In the Fig. 7, the timestamp column is represented in milliseconds, and the sensordata.gyro[0], gyro[1], and gyro[2] columns represent the angular velocity in each x, y, and z axis direction, which is used to analyze small vibrations and slope changes of the structure.

The sensordata.atm[0] column represents the air pressure at the point where the sensor is located, and is used to analyze the elevation change and settlement of the structure.

The gyroscope data is used to analyze the vibration and tilt of the structure, checking the x-axis angular velocity, y-axis angular velocity, and z-axis angular velocity, and the air pressure and altitude data is used to check whether the structure is settling or not. The columns required for the study were selected from the entire CSV file and modified

```

# PS C:\Users\H\I\anaconda3\envs\heristudio\python.exe "C:\import pandas as pd.py"
id deviceid sensordata.gyro[0] sensordata.gyro[1] sensordata.gyro[2] sensordata.degree[1] sensordata.degree[2] sensordata.atm[0] timestamp
0 66bda4c4773e1eb7e3 1 -0.02 -0.40 ... -0.74 90.00 998 2024-04-02T10:32:23.631574
1 66bda4c4773e1eb7e6 17 -0.11 -0.07 ... 0.30 70.81 998 2024-04-02T10:32:23.602016
2 66bda4c4773e1eb7e6 6 -0.06 -0.30 ... -0.21 90.00 998 2024-04-02T10:32:23.245751
3 66bda4c4773e1eb7e9 13 0.36 0.40 ... 3.10 90.00 998 2024-04-02T10:32:23.021000
4 66bda4c4773e1eb7f1 5 -0.10 -0.08 ... -0.12 90.00 997 2024-04-02T10:32:23.655232
[5 rows x 16 columns]

```

Fig. 7. IoT sensor raw data.



```
PS C:\Users\HP> & C:\Users\HP\anaconda3\envs\nerfstudio\python.exe "hi/import pandas as pd1.py"
```

timestamp	gyro_x	gyro_y	gyro_z	pressure	gyro_magnitude	tilt_angle	relative_altitude	altitude_change
2024-04-02 18:32:23.031574	-0.02	-0.40	-0.01	998	0.406625	-92.867405	0.000000	0.000000
2024-04-02 18:32:23.082816	-0.11	-0.07	0.05	998	0.139642	-147.528888	0.000000	0.000000
2024-04-02 18:32:23.245751	-0.06	-0.30	-0.12	998	0.328634	-101.309932	0.000000	0.000000
2024-04-02 18:32:23.855232	-0.18	-0.08	0.17	997	0.260192	-156.037511	8.456112	8.456112
2024-04-02 18:32:22.793142	-0.09	-0.11	0.03	998	0.145258	-129.289487	0.000000	-8.456112

Fig. 8. IoT sensor feature extraction.

to gyro\_x, gyro\_y, gyro\_z, pressure, and DateTime to pre-process and extract features.

In the Fig. 8, gyro\_magnitude indicates the magnitude of the vibration by combining the gyroscope data of the x, y, and z axes, and the larger the value, the stronger the overall vibration of the industry-academia collaboration center.

tilt\_angle is a tilt angle that uses gyroscope x, y axis data and expresses the degree of abnormal tilt of the industry-academia collaboration center in degrees (°).

relative\_altitude is the relative altitude (m) of the sensor location using barometric pressure data, which measures the subsidence phenomenon as a change in the relative altitude of the industry-academia collaboration center.

altitude\_change is the change in relative altitude from the previous measurement, and a positive value indicates an increase in altitude, while a negative value indicates a decrease. In the case of subsidence, negative values are consistently displayed.

## 5.2. Data Analysis

In this study, we use IoT sensor data to test changes in buildings based on time with establish anomaly intervals, and analyze cumulative displacement outliers.

In general, LSTM algorithm is mainly used for time series prediction, which predicts the future based on past data, and has the advantage of estimating future changes, but this study aims to examine the change status of the structure from the currently collected data, and utilizes the Isolation Forest algorithm, an unsupervised learning-based anomaly detection model used to identify anomalies with a small data volume.

The analysis environment is Intel(R) Core (TM) i9-14900 K 3.20 GH, 128 GB RAM, NVIDIA RTX 4090, Windows 11 pro, Visual Studio Code, and the Python version used is 3.10.14.

### 5.2.1. Isolation Forest Formulas

From the entire datasets, a tree(iTree) randomly selects a sampling  $\psi$ (psi), and randomly selects  $q$  attribute to be used as branching criteria from all attributes in the datasets.

e.g., this study has tree features gyro\_magnitude, tilt\_angle, altitude\_change.

Attributes={gyro\_magnitude,tilt\_angle,altitude\_change}

$q \sim \text{Uniform (Attributes)}$

A branch point  $p$  is randomly selected between the minimum and maximum values of the selected attribute  $q$ . If  $x_p$  is less than the  $p$ , the data is split into a subtree  $T_L$  (Tree Left); if  $x_p$  is equal or larger than the  $p$ , then data is split into a subtree  $T_R$  (Tree Right).

$p \sim \text{Uniform}(\min(x_q), \max(x_q))$

$x_p < p \rightarrow T_L, \quad x_p \geq p \rightarrow T_R$

split condition:  $x \in T_L$ , if  $x_q < p$

$x \in T_R$ , if  $x_q \geq p$

Through split condition, the data  $x$  is moved from the root of iTree along the tree nodes until it becomes isolated, such as until only one sample remains at the leaf node, or until the pre-defined height limit is reached. And this is called the path length and is denoted by  $h(x)$ . The average value of  $h(x)$  across all trees is calculated.

height limit =  $\log_2(\psi)$

$h(x)$  = depth of  $x$  in the Isolation Tree

$E[h(x)] = 1/t \sum hi(x)$

where,  $i: 1 \sim t, \quad t$ : number of iTrees,

$hi(x)$ : path length of data  $x$  in the  $i$ th tree

Since the path length varies depending on the number of samples, it cannot be compared based on length value. Therefore, it is needed to mathematically approximate the expected average path length of normal data when all samples are selected with the same probability, and divide the path length by the expected average path length obtained through this process.

The expected average path length obtained here is called the normalization constant  $c(\psi)$ .

$c(\psi) = 2H(\psi - 1) - (2(\psi - 1)/\psi)$

where,  $H$ : harmonic number,  $H(i): \ln(i) + \gamma$  ( $\gamma \approx 0.5772$ )

If data  $x$  is isolated within a short path in the iTree, it is likely to be an outlier. Therefore, for a specific data  $x$  in the range of 0 to 1, the expected path length based on the sub sample size  $\psi$  is normalized to obtain the outlier score  $s(x, \psi)$ . The data points are then sorted in descending order of outlier scores, and the top contamination % with the lowest scores are labeled as outliers. The contamination rate serves as the threshold value for determining the outlier score.

$S(x, \psi) = 2^{-(E[h(x)]/c(\psi))}$

### 5.2.2. Parameter Settings

The Isolation Forest algorithm is like to a binary tree from regression tree. It involves randomly selecting a sample, called  $\psi(\text{psi})$ , to create a tree structure(iTree). It is divided random values at each node. This process calculates the distance of an object from a leaf node, which is isolated as the outlier score.

A random sample of  $\psi(\text{psi})$  sample is selected from the entire data set to create a single iTree. Next, a variable  $q$  of data  $x$  is randomly selected from the iTree, which is used as the splitting axis by randomly selecting one of all attributes at each node division. Then uniform distribution is used to sample a split value  $p$  between the minimum and maximum values of the selected variable.

This process is recursively repeated until the sub trees satisfy splitting criteria, either until a single sample remains or until the maximum depth limit (height limit) is reached.

The number of trees( $t$ ) was set to 100 in this experiment. Although the variance decreases as the number of trees( $t$ ) increases and it converges to ROC AUC value at  $t=100$ , but based on previous literature indicating that additional performance improvements are limited despite increased computational cost, this parameter was set to 100.

The sub-sampling size( $\psi$ ) was set to 256 based on prior literature. This consideration accounts for the fact that outliers can easily become isolated even in small samples.

Contamination was set to 0.3% (0.003). This was set how to detect outliers caused by internal factors of the building and to minimize external factors such as this building's construction work, vehicle movement, and sensor problem issues. Out of the total 9,716,989 data points, the period from 6:00 to 18:00 on April 3rd, 2024 was considered as the ground truth for pseudo labels, achieving precision $\approx 92.0\%$ , recall $\approx 92.0\%$ , and ROC AUC $\approx 93.0\%$ , which indicates that the parameters were appropriately set.

The height limit was set to  $\log_2(\psi)$  based on previous literature. The maximum height of a binary tree is  $\log_2(\psi)$ . If the height limit is small branching ends at the root node, limiting a distinction between normal and abnormal values. If it is large, normal values have longer paths and abnormal values have shorter paths, improving outlier detection.

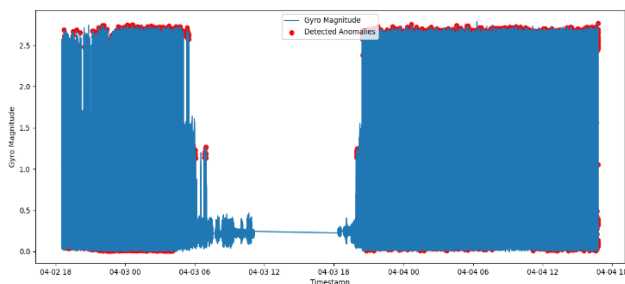


Fig. 9. Magnitude outlier detection result.

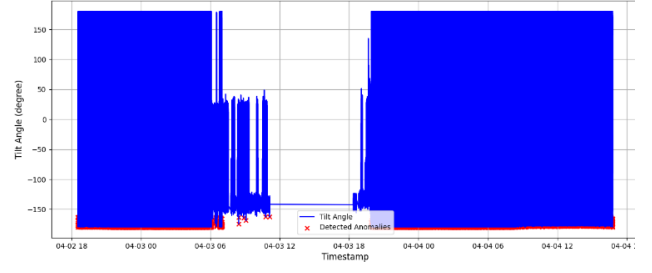


Fig. 10. Tilt angle outlier detection result.

### 5.2.3. Magnitude of Vibration Outlier Analysis

A total of 28,405 anomalies were identified by analyzing the gyro\_magnitude data, which represents the magnitude of the vibration, and shown in Fig. 9, where the x-axis is time and the y-axis is magnitude of the vibration, the blue line represents the magnitude of the vibration measured over time, and the red area represents the anomalies. From 18:00 on April 2 to 6:00 on April 3, high vibration was detected in almost the entire area, and from 18:00 on April 3, the vibration increased again.

### 5.2.4. Tilt Angle of Outlier Analysis

A total of 26,942 anomalies were identified by analyzing the tilt\_angle data. The tilt angle measured by the sensor in blue is the tilt angle (unit: degree), and the points in red indicate the anomalies in tilt. In Fig. 10, most of the anomalies are around  $-150$  degrees ( $^{\circ}$ ), indicating that the industry-academia collaboration center moved abnormally in one direction relative to the sensor.

### 5.2.5. Risk Score

Next, the data from the sensors were integrated to determine the risk level of the building in a time series. By integrating the data on the magnitude of vibration (gyro\_magnitude), tilt (tilt\_angle), and relative altitude change (altitude\_change), the risk score was calculated on an hourly basis.

The risk score is a single, comparable metric of multiple signals (sensor characteristics), and a moving average was used to smooth out short-term noise and identify trends in the continuous signals of the sensor data.

The weights used in the risk score are based on three data points: the magnitude of the structure's vibration (gyro\_magnitude), the change in tilt (tilt\_angle), and the change in relative altitude (altitude\_change).

The criteria for selecting the weights is based on features shown in Fig. 8 of IoT data. In isolation forest algorithm, iTree randomly selects a sample and randomly selects  $q$  attribute to be used as branching criteria from all attributes.

As shown in Table 5, the magnitude of the vibration and the tilt angle are each weighted 0.4 because they directly reflect

Table 5. Risk score.

Data	Weight	Contents
gyro_magnitude	0.4	Indicates the vibration state of the building and is weighted highly because the greater the vibration, the higher the risk.
tilt_angle	0.4	Tilt changes are highly weighted as they are associated with building imbalances, swaying, etc.
Altitude_change	0.2	Elevation changes are given less weight than relative vibration and tilt changes because this is considered more indirect.

the displacement of the building, and the relative elevation change is weighted 0.2 because it can be affected by environmental factors such as air pressure and temperature.

Upon examining the observed data, gyro\_magnitude and tilt\_angle change values showed high sensitivity, while altitude\_change values exhibited a relatively stable characteristic. To determine the weights for three data points, we derived five combinations that sum to 1.0 and conducted a preliminary review considering the sensitivity characteristic of them mentioned above. As results, the weight was set to gyro\_magnitude 0.4, tilt\_angle 0.4, altitude\_change 0.2 for reflection in all data change values.

For the purposes of this study, a risk score is defined as follows.

$$\text{risk\_score} = (\text{gyro\_magnitude average} \times \text{weight } 0.4) + (\text{tilt\_angle average} \times \text{weight } 0.4) + (\text{altitude\_change average} \times \text{weight } 0.2)$$

Fig. 11 visualizes the weighted score (magnitude of oscillations, change in slope, change in relative elevation) and the score per component as a time series evolution, with red dots (markers) indicating points where the score rises sharply from the previous point in time.

The risk score is almost constant for most of the time intervals, indicating that the building's condition change is affected by both vibration and tilt changes, and that height changes such as settlement are relatively small and constant.

However, the width of the score decreases significantly



Fig. 11. Risk score result.

in the time interval from 06:00 to 18:00 on April 3, 2024, indicating that some outliers are missing and that the Isolation Forest algorithm is not able to reflect the trend of outliers over time.

The result of the test with the isolation forest algorithm shown in Fig. 11 is that the risk\_score graph is a result of judging whether the data obtained from the tree-based structure is an outlier, which has limitations in analyzing atypical outliers (correlation between variables). Therefore, we applied extended isolation forest to check for complex outlier changes using risk\_score, which is a weighted sum of gyro\_magnitude, tilt\_angle, and altitude\_change, as a single variable in continuous data, and checked the results.

Extended isolation forests represented the points in time when one or more of the following fluctuated rapidly: vibration, tilt, or elevation change, and showed how risk scores and outliers increased together with sustained vibration or tilt change.

The x-axis represents time for approx. 7-days from April 2 to April 9, 2024, and the y-axis represents the change in state based on a weighted sum of the magnitude of the risk\_score's oscillation, slope change, and relative elevation change. The blue line shows the change in risk\_score over time, and the red dots indicate outliers detected by the extended isolation forest (Fig. 12).

After applying the extended isolation forest, redundancy and noise among variables were not eliminated, so we used principal component analysis (PCA) to identify the main direction of variance of the data acquired from the sensors, and then applied the underlying isolation forest. This is a research method proposed by Liu et al. [18], and we implemented it for the test of this study to confirm the results.

The extended Isolation Forest was applied to improve the range and segmentation resolution of outlier detection, but the redundancy and noise problems among variables were not eliminated. The risk\_score variable is composed of the weighted sum of gyro\_magnitude, title\_angle, and altitude\_changedml, which tend to change simultaneously in time, so we used principal component analysis (PCA), which is used to maximize the total variance while converting high-dimensional data into low-dimensional space, to identify the main direction of variance of the data acquired from the sensor, and then applied isolation forest based on the results. This is a research method proposed by Liu et al.



Fig. 12. Extended isolation forest result.

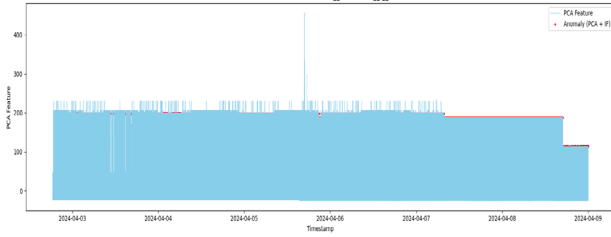


Fig. 13. PCA & isolation forest result.

[18] and was applied to the research environment of this study to confirm the results.

The x-axis shows the time flow, and the y-axis shows the dimensionality reduction by PCA based on three variables: gyro\_magnitude, tilt\_angle, and altitude\_change. The blue part is the full signal, while the red dots represent outliers detected by PCA and isolation forest. Starting on April 2, 2024, the sensor values show variability, with possible temporary building vibrations or tilt changes between April 5 and April 6. From April 8, it has been found to be stable.

Fig. 13 shows the correlation of sensors rather than a single variable in tracking building condition changes and detecting anomalies based on IoT sensors. This shows that by applying PCA and isolation forest in combination, the changes that were limited to be detected by the existing isolation forest could be identified.

The performance comparison and improvement of the algorithms used in this paper, PCA and isolation forest showed improvement overall, but showed little changed in altitude\_change. isolation forest basically has good performance in anomaly detection, but it was difficult to distinguish anomalies at specific points or there were some points, and Extended Isolation was able to see complex anomalies better while complementing the limitations of it.

PCA and isolation forest were shown to remove redundancy and noise of three data (gyro\_magnitude, tilt\_angle, altitude\_change), and showed good results based on Risk score.

## VI. CONCLUSION

In this paper, we analyzed the state change of the structure based on the vibration, tilt, and relative elevation change data collected using IoT sensors for about 7 days from April 2 to April 9, 2024, for the Industry-Academic Cooperation Center of Busan \*\*\* University. The total number of data collected was about 9.71 million, and based on this, we analyzed the factors that affect the state change inside the building.

The magnitudes of vibration and tilt changes showed a constant trend in most time intervals, but large changes were observed in certain time periods, while elevation changes were relatively small and stable. A weighted sum of the magnitude of vibration, tilt, and relative elevation changes

was used to calculate a risk score, and the score was visualized as a time series to check the association between changes in the risk score and physical events.

In addition to the isolation forest algorithm, the extended isolation forest algorithm was used to analyze the IoT sensor data, and it was shown to effectively separate the data distribution and precisely detect outliers in buildings.

In addition, the combination of PCA and isolation forest was shown to detect some undetected time intervals based on the correlation between variables.

This study confirmed that the accumulated data of IoT can be utilized to build a system that diagnoses structural condition changes and defect factors in real time.

Changes in data on internal factors of these structures are difficult to identify in the static state of the structure, and it is expected that it can be utilized for integrated safety management by supplementing diagnostic methods based on video data of the structure.

The results of this study suggest the possibility of building a data-based risk zone identification and real-time detection system, and given that certain abnormalities were concentrated in some time periods among thousands of data, it is expected that IoT big data for public facilities or structures used by a large number of people can be built in the future, and potential risks that occur under certain conditions or accumulate changes due to internal factors can be detected through long-term observation, contributing to the safety management of society and citizens.

And also, this study has the limitations as follows and needs to be improved by a continuous study in the future.

### 6.1. Potential Sensor Noise

This paper was conducted based on data collected from IoT sensors over seven days at specific building. As making IoT sensor boards production, we were focused to ensure stability and sustainability. However, during data collection, the changes unrelated to actual internal physical changes within the building may have occurred due to insufficient battery power, or unstable network connection, or and sensor's malfunctions, and detected as normal noise.

To delete such unrelated sensor noise from datasets and to ensure data's reliability, it is necessary to additionally conduct pre-filtering algorithms.

### 6.2. Environmental Interferences

External events such as construction works, vehicle movement, and weather changes (e.g., strong winds occurring) near the target building can cause errors in the accuracy of IoT data. It is necessary to collect data on such environmental interferences events as secondary data, and to post-process outliers that occurred during that time period.



### 6.3. Generalizability to Other Types of Buildings

In this study, the tested method was applied to a particular building in terms of location, structure, and environment. However, although the framework of this study can be applied to other types of buildings, researchers need to consider the operational differences of IoT sensors when applying them to various building types to improve IoT sensors and optimize data distribution and characteristics. This will enable the method to be applied to other types of buildings.

## REFERENCES

- [1] Internet resource, Falling Rocks at a Construction Site in Busan, Yonhap News, Feb. 2019 <https://www.yna.co.kr/>
- [2] Internet resource, Earthen slope collapse, NoCut News, Oct. 2019 (<https://www.nocutnews.co.kr/>)
- [3] Y.S. Kim, *Current Status of Old Buildings and Future Tasks*, National Assembly Research Service, Apr. 2020.
- [4] M. B. Seo, "Automatic management technology of infrastructure facility based on IoT," in *Proceedings of the Korea Academia-Industrial Cooperation Society*, Nov. 2014, pp. 294-296.
- [5] S. Yeon, C. H. Yeon, and J. Kim, "Application technique on the construction disaster protection based on spatial information and USN/IoT/UAV," in *Proceedings of the Korean Contents Association General Conference*, May 2018, pp. 123-124.
- [6] H. Y. Kim, K. A. Choi, and I. P. Lee, "Drone image-based facility inspection; Focusing on automatic process using reference images," in *Proceedings of the Journal of the Korean Society of Geospatial Information Science*, Jun. 2018, pp. 21-32.
- [7] T. K. Park, "A plan for establishing iot-based building maintenance platform (S-LCC): Focusing a concept model on the function configuration and practical use of measurement data," in *Proceedings of the Korea Academia-Industrial Cooperation Society*, Feb. 2020, pp. 611-618.
- [8] J. Lim, S. You, and Y. Kim, "Monitoring system based on bigdata platform for safety management of road facilities," in *Proceedings of the Korean Institute of Information Technology*, Nov. 2020, pp. 139-151.
- [9] B. Ji, "Machine learning-based concrete crack detection framework for facility maintenance" in *Proceedings of the Korean Geo-Environment Society*, Oct. 2021. pp. 5-12.
- [10] J. Y. Park, J. S. Shin, J. B. Won, J. W. Park, and M. Y. Park, "Development of low-power IoT sensor and cloud-based data fusion displacement estimation method for ambient bridge monitoring," in *Proceedings of the Computational Structural Engineering Institute of Korea*, Oct. 2021, pp. 301-308.
- [11] K. Perveen, G. E. Bridges, S. Bhadra, and D. J. Thomson, "Corrosion potential sensor for remote monitoring of civil structure based on printed circuit board sensor," *IEEE Transactions on Instrumentation and Measurement*, vol. 63, no. 10, Oct. 2014, pp. 2422-2431.
- [12] Mahdavi, Ardeshtir, and M. Taheri, "An ontology for building monitoring," *Journal of Building Performance Simulation*, vol. 10, no. 5-6, pp. 499-508, 2016.
- [13] Y. Pan, X. Zhang, G. Cervone, and L. Yang, "Detection of asphalt pavement potholes and cracks based on the unmanned aerial vehicle multispectral imagery," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 10, pp. 3701-3712, Oct. 2018.
- [14] L. Khelifi and M. Mignotte, "Deep learning for change detection in remote sensing images: Comprehensive review and meta-analysis," *IEEE Access*, vol. 8, pp. 126385-126400, 2022.
- [15] H. Ghanbari, M. Mahdianpari, S. Homayouni and F. Mohammadimanesh, "A meta-analysis of convolutional neural networks for remote sensing applications," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 3602-3613, 2021.
- [16] H. Jiang, M. Peng, Y. Zhong, H. Xie, Z. Hao, and J. Lin, et al. "A survey on deep learning-based change detection from high-resolution remote sensing images," *Remote Sensing*, vol. 14, no. 7, p. 1552, 2022.
- [17] Panja, S. N. Patowary, S. Saha, and A. Nag, "Anomaly detection in IoT using ex-tended isolation forest," in *International Symposium on Artificial Intelligence*, Cham, pp. 3-14, 2022.
- [18] D. Liu, H. Zhen, D. Kong, X. Chen, L. Zhang, and M. Yuan, et al., "Sensors anomaly detection of industrial internet of things based on isolated forest algorithm and data compression," *Scientific Programming*, p. 6699313, 2021.
- [19] KOLAS (Korea Laboratory Accreditation Scheme), *Performance Test, Research Institute of Medium & Small Shipbuilding*.

## AUTHORS



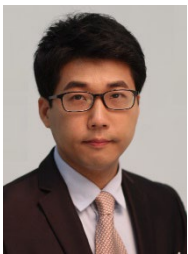
**Chan-Won Yoon** completed the two doctoral courses at the Department of Applied Economics, Hanyang University, Republic of Korea and at the Department of Technology Management, Korea University of Technology and Education, Republic of Korea, in February 2016. In 2001, he worked as a senior researcher at the National Information Industry Promotion Agency (NIPA), Ministry of Science and ICT, Republic of

Korea, and in 2005, he received the Bluetooth Industry Achievement Award from the Bluetooth Industry Council of the Korea Radio Promotion Association, Ministry of Science and ICT, Republic of Korea. He has participated in ICT researches on the Policy improvement for Small power wireless device transmission (ETRI, 2005), the Research on strategy for activation of safety and security-related Telematics services (NIA, 2006), the Research on strategy for overseas expansion of WiBro in Latin America and Southeast Asia (Korea Communications Commission, 2009), the Development of broadband wireless PON combining technology using network coding (KAIST, 2009), the Establishment of next-generation convergence infrastructure (Ministry of Knowledge Economy, 2009).

He has participated in Policy researches on the Technical engineering education model of Korea Polytechnic University (KDI Graduate School, 2011) and the Honduras National ICT Master Plan (KDI, 2011), and also he has conducted the Roadmap development for the ICT convergence smart printing service industry (Jeju Techno Park, 2016), the Mid-to long-term development strategy for fostering the VR industry in Jeju (Jeju Techno Park, 2016). He has jointly conducted Economy research on the Study on the current status and development tasks of the domestic wireless communication device industry (2017) with the Gyeonggi Headquarters of the Bank of Korea and the Daegu Gyeongbuk Headquarters of the Bank of Korea, Republic of Korea.

Since April 2018, he has been an Associate Professor at the Department of Artificial Intelligence Applications, Dongseo University, Busan, Republic of Korea, and was awarded the Manpower Training Citation by the Ministry of Education, Republic of Korea, in November 2022.

In February 2025, he completed PhD course in Data Informatics, Korea Maritime and Ocean University, Republic of Korea.



**Jun-Ho Huh** received B.S. in Science Degree from Department of Major of Applied Marine Sciences, And, B.E. in Engineering Degree (Double Major) from Department of Computer Engineering from Jeju National University, Jeju, Republic of Korea in Aug. 2007. And he received M.A. degree in education from the Department of Computer Science Education, Pukyong National University, Republic of Korea, in Aug. 2012.

Also, he received the Ph.D. in Engineering Degree from Department of Computer Engineering, Pukyong National Republic of Korea in Feb. 2016. He received Best Paper Minister Award (Ministry of Trade, Industry and Energy, Korean Government) the

16th International Conference on Control, Automation and Systems (Oct. 2016), ICROS with IEEE Xplore. Also, he received Springer Nature Most Cited Paper Award, Human-centric Computing and Information Sciences most cited paper award 2019 (Research published in the journal between 2016–2018; SCIE). Also, he was awarded the Commendation for Meritorious Service in the Promotion of Busan's Data Industry in December 2023 (Commendation by the Mayor of Busan, Republic of Korea Government).

Since July 2023, he has been an Editor-in-Chief (EiC) of The Journal of Global Convergence Research, Global Convergence Research Academy, Pukyong National University, Republic of Korea. Also, since Apr 2016, he has been an Associate Editor (AE) at Journal of Multimedia Information System (JMIS), Korea Multimedia Society (EI/KCI indexed).

Since 2017, he has been an Technical Committee (TC) at IFAC (International Federation of Automatic Control), CC 1, TC 1.5. Also, since 2017, he has been an TC at IFAC, CC 3, TC 3. And, since 2017, he has been an TC at IFAC, CC 7 (Transportation and Vehicle Systems), TC 7.2. (Marine Systems). Also, since 2020, he has been an (TC) at IFAC, CC 2, TC 2.6.

He was a Research Professor at Dankook University at Jukjeon, Yongin, Republic of Korea, from July 2016 to September 2016. Also, he was an Assistant Professor with the Department of Software, Catholic University of Pusan, Republic of Korea, from December 2016 to August 2019. Also, he was an Assistant Professor with the Department of Data Informatics, National Korea Maritime and Ocean University, Republic of Korea, from Sep., 2019 to Sep., 2021. Also, he was an Associate Professor (Tenured) with the Department of Data Informatics, National Korea Maritime and Ocean University from Oct., 2021 to Mar., 2024.

Since April 2025, he has been a Full Professor (Tenured) with the Department of Data Informatics/Data Science, National Korea Maritime and Ocean University, Republic of Korea. Also, since September 2020, he has been an Center Chair (Director) of Big Data Center for Total Lifecycle of Shipbuilding and Shipping at National Korea Maritime and Ocean University. Also, since April 2025, he has been Join a Full Professor (Tenured) (Sep. 2022–Mar. 2025: Associate Professor) of Global R&E Program for Interdisciplinary Technologies of Ocean Renewable Energy (BK 21 Four Research Group) of Interdisciplinary Major of Ocean Renewable Energy Engineering, National Korea Maritime and Ocean University, Republic of Korea. Also, since April 2025, he has been a Full Professor (Tenured) (Mar. 2024–Mar. 2025: Associate Professor) of Research Group of Marine Safety and Disasters, Dept. of Convergence Study on the Ocean Science and Technology, Ocean Science and Technology School (OST School: KMOU & KIOST: Korea Institute of Ocean Science and Technology). He is the Book Author of "Smart Grid Test Bed Using OPNET and Power Line Communication" pp. 1-425, IGI Global, USA, 2017. Also, He is the Book Author of "Principles, Policies, and Applications of Kotlin Programming" pp. 1-457, IGI Global, USA, 2023. And, he has authored/edited 10 books and edited 10 special issues in reputed Clarivate Analytics Index journals. Elsevier/SCOPUS World Top 2% Scientist (2023, 2024). Also, he has published more than 117 articles in Clarivate Analytics Index (SCI/SCIE/SSCI indexed) with over 4,300 citations and has an h-index of 34. He is spending his sabbatical year at Yeungnam University (School of Computer Science and Engineering) from Mar. 2025 to Feb. 2026.

<https://scholar.google.co.kr/citations?user=cr5wjNYAAAJ&hl>