Rancidity Analysis Management System
Based on Machine Learning Using IoT Rancidity Sensors

Sung-Sam Hong,1 Kisoo Chang,1 Junhyung Lee,2 and ByungKon Kim2*

1Department of Computer Engineering, Gachon University,
IT College, 1342, Seonamdae-ro, Sugeong-gu, Seongnam-si, Gyeonggi-do, 13120, Korea
2Korea Institute of Civil Engineering and Building Technology,
283, Goyangdae-ro, Ilsanseo-gu, Goyang-si, Gyeonggi-do, 10223, Korea

(Received August 31, 2019; accepted November 18, 2019)

Keywords: rancidity, sensor, IoT, machine learning, data mining, road pavement quality management

Rancidity data can be used in various fields such as the quality analysis of food and raw materials used for construction. The rancidity of raw materials used in road pavement asphalt is currently only at the level determined by the temperature or visual factors. Although construction workers are managed individually and subjectively, such as by visual methods, they cannot be managed in practice. In this paper, we propose a system combining a rancidity sensor with an Internet of Things (IoT) communication module that collects and predicts rancidity measurements in real time at a site. The values measured by the sensor are periodically transferred to the Cloud through the IoT communication module, the validity of the data set is established, and the systematic management of the data is performed using machine-learning-based data analysis techniques. The results of an experiment showed a high classification prediction accuracy of 91.3% and a short-term pattern prediction accuracy of 96.6% (weighted scaling), confirming its excellent potential for raw material quality management. The results of this paper will be applied as a road pavement quality management system.

1. Introduction

Rancidity data can be used in diverse areas such as the quality analysis of food and raw materials for construction, and rancidity measurement is becoming an important method of determining the state of raw materials. In the case of oil used for frying food, the rancidity of the oil is measured by using a paper or from experience (such as how many times it has been used to fry chicken) in general restaurants, making accurate measurements difficult. The rancidity of raw materials used for pavement asphalt is determined merely by the temperature or visual estimation. Although rancidity can be managed by subjective measurement methods, such as a visual method and individual control, its systematic management at each position is not possible. Studies on systems allowing the transmission of measured data for monitoring and management and methods of analysis are scarce. In existing studies, although rancidity sensors,
particularly those for edible oils were investigated,\textsuperscript{(1–3)} the development of sensors combined with communication modules for the collection or analysis of data is insufficient.\textsuperscript{(4,5)}

The system proposed in this paper allows the real-time collection of rancidity measurements in the field and predictive analysis by coupling rancidity sensors and an Internet of Things (IoT) communication module. It periodically transmits the values of rancidity measured by sensors to the Cloud via the IoT communication module, constructs a data set of rancidity, and performs the systematic management of rancidity measurements using data analysis techniques based on machine learning. By integrating the data transmitted to the Cloud to construct the data set and processing the data to construct a training set, we have proposed a system that allows classification and short-term predictions. The proposed system is provided through a mobile application or the web and enables the extraction of a quality management pattern for rancidity by applying an IoT rancidity sensor and an intelligent data analysis technique. In this system, real-time monitoring and quality management are possible as features are extracted from the patterns, and quality management levels are classified through time series analysis to classify and predict the quality states of rancidity and to predict changes in quality so as to enable real-time monitoring and quality management. By the rational use of materials and the construction of a quality management system, the reliability of the quality of raw materials can be enhanced. Additionally, as reported in the paper, experiments on quality management, classification, and prediction as well as the prediction of changes in patterns have been carried out with the proposed system (Fig. 1) by using the developed IoT sensors to collect actually measured data from restaurants using oil for a given period. According to the results of the implementation, high accuracies for a classification prediction accuracy of 91.3% and a short-term pattern prediction accuracy of 96.66% were observed, confirming the excellent potential of the proposed system for the quality management of raw materials.

2. Related Works

2.1 Machine learning methods for IoT data analysis

IoT is set to revolutionize all aspects of our lives. The number of objects connected to IoT is expected to reach 50 billion by 2020, giving rise to an enormous amount of valuable
The data collected from IoT devices will be used to understand and control complex environments around us, enabling better decision making, greater automation, and higher efficiency, productivity, accuracy, and wealth generation. Data mining and other artificial intelligence methods will play a critical role in creating smarter IoTs, albeit with many challenges. There are eight well-known data mining algorithms for IoT data. These include deep learning artificial neural networks (DLANNs), which build a feedforward multilayer artificial neural network (ANN) for modeling high-level data abstractions. In Ref. 7, the results of three real IoT datasets show that C4.5 and C5.0 have higher accuracies, are memory-efficient, and have relatively higher processing speeds.

ANNs and DLANNs can provide highly accurate results but have high computation cost and complexity. Through the performance comparison of eight typical algorithms, the efficiency of IoT and machine learning algorithms was assessed, and each feature was presented in Ref. 6. As such, the technology is being used in various fields ranging from data collection to analysis and prediction in conjunction with IoT sensors and machine learning methods.

2.2 Support vector machine (SVM)

The classification model was constructed using the SVM as the machine learning model. The basic principle of the SVM begins with the classification of SVM input patterns into two classes {+1, −1} through supervised learning. When training groups are classified into two classes, the hyperplane that separates the training patterns can be determined. The hyperplane here refers to the cutting plane that separates individual groups. The boundaries of the patterns determining this plane are called support vectors.

There are three key stages in the SVM algorithm. The first is to find the boundaries, the second to find the hyperplane that maximizes the margin or distance of the hyperplane from each boundary, and the last to determine which side of the hyperplane the given test record belongs to.

- Find boundary of each class
  The outline obtained when all points corresponding to one class in the data set is the boundary of the class. Figure 2 shows a boundary, which is referred to as the convex hull (Fig. 3). Each class has its own convex hull, and the convex hulls of different classes do not intersect.
• Find the best hyperplane to satisfy conditions
When finding the hyperplane, a mathematical method is employed, where any hyperplane has two attributes of $x_1$, $x_2$ as follows, and an expression such as $H: b + \omega = 0$ is given. Examining this expression, we have $x = (x_1, x_2)$ and $b$ is an intercept called the deflection. This expression is similar to the standard form of a linear equation, and the equation describing the optimum hyperplane is given as $H_0: b_0 + \omega_0 = 0$. The margin is given by $Margin = \frac{2}{\omega_0 \omega_0}$. From this expression, the optimum hyperplane is obtained.

• Classification of test records
Once the boundaries and the hyperplane are determined, it is calculated on which side of the hyperplane a new test record to be classified lies. This can be done easily by substituting the test record $x$ into the mathematical expression for the hyperplane. If the calculated value is $+1$, it belongs to the positive class, while it belongs to the negative class in the case of $-1$.

• Features
The SVM is an algorithm used for classification and prediction together with such algorithms as decision-making trees and neural networks. It is a model allowing the linear transformation of the training set, which shows better performance, making it easier to transform the data set.

3. Intelligent Rancidity Analysis System Based on IoT

The proposed system can transmit the sensor data by using rancidity sensors capable of measuring temperature and rancidity and by coupling a Bluetooth communication module to the sensors. Figure 4 shows a hardware of rancidity sensors. Sensors and modules are designed and developed, and the data measured by the sensors is transmitted to the data collection server via gateway. By installing one gateway for each field, the gateway transmits rancidity and temperature data, and these gateways will later be developed into diversified IoT gateways in an expandable form.$^{(11,13)}$ For the data analysis module, an algorithm that supports a classification function (rancidity quality state) and a prediction function (quality change) based on the

Fig. 4. (Color online) Rancidity sensor.
Rancidity measurements has been proposed by our other studies. It employs a newly developed technology that analyzes to enable control and management from the viewpoint of monitoring and maintains the rancidity of raw materials. Figure 5 shows the process of our proposed system. The communication protocol between the sensor and the server is shown in Fig. 6. The data measured by sensors is transmitted to the gateway via the Bluetooth communication module, then the gateway processes the data according to the prescribed protocol for transmission to the server.

The data transmitted by using the protocol in Rest API form is collected, saved, and managed through a data collection server based on a web application server that enables real-time monitoring. Data analysis is carried out in the analysis module using the collected data. Discussed in detail below are the contents of the analysis module.

### 3.1 Model for quality classification and prediction

The classification and prediction functions respectively classify and predict the quality states through the proposed classification model based on machine learning, as well as identify and evaluate the current quality state following the identification of management levels. Through the model obtained by the accumulation of data, a standardized data model allowing the identification of oil management states is constructed. The classification function is characterized as follows.

---

**Fig. 5.** (Color online) Process of proposed system.

**Fig. 6.** (Color online) Communication protocol.
1. Signal preprocessing and feature extraction from the rancidity pattern as a function of the time flow of the location using each raw material.
2. Labeling of the existing data according to the prescribed levels with the quality management level determined by prior knowledge as a target variable (Good, Normal, Poor, or Very Poor).
3. Management label classification by generating an analysis model through data learning and by inputting data collected from individual businesses.
4. In the present module, each feature from the data collected for 500 days has learning sets constructed using feature sets with the current rancidity, the average rancidity for day D-1, the weekly average rancidity, and the monthly average rancidity. Each training set is constructed from data for 100 days to generate an SVM model. When the data of a new time window is inputted from the generated model, quality levels are classified and predicted.

3.2 Prediction model for short-term quality patterns

In the prediction function for short-term patterns, patterns for a given time section are identified to determine the quality change times in the levels according to quality, and the quality change pattern for the next time section is predicted. Through the prediction of quality changes, quality management can be performed by predicting replacement times for the raw materials, the state changes of the original material, and so forth. To generate a prediction model, an autoregressive integrated moving average (ARIMA)\(^{(14)}\) was applied.

Real-time series data are frequently nonstationary. However, the AR\((p)\), MA\((q)\), or ARMA\((p, q)\) model combining these two cannot explain such a nonstationary aspect. Therefore, the model used to remove such a nonstationary aspect from the model is the ARIMA model itself, which is expressed as ARIMA\((p, d, q)\)\(^{(15)}\).

In the prediction module, the ARIMA model has been designed using actually measured data in the unit of 1 over one day and then changes in the data for 1 h are predicted later. Since rancidity usually increases over time then decreases again at the time of material replacement, the ARIMA model is suitable for our application.

4. Experiments and Results

4.1 Environment of experiment

We performed various experiments. The environment used in the experiment is shown in Table 1. The rancidity sensor used a sensor we developed in another study\(^{(16)}\). Because we implemented it on the web, we used a server/client model.

4.2 Experiment on quality management level classification

In the classification experiments on quality management labels, actually measured data were generated to produce a measured data set for 500 days, using feature sets prepared with the current rancidity, yesterday average rancidity, weekly average rancidity, and monthly
average rancidity for the construction of the training sets. By constructing a total of 10 training sets with 100 cases of data each composed of feature sets, as shown in Table 2, a model was produced by using the SVM from each learning set, and the performance of the model was verified using the test set.

Experimental results outputted by inputting the test set in the model made from individual training sets are shown below. The total number of data for the data set is 101, with the data level composition of 16 for Good, 57 for Normal, 6 for Poor, and 22 for Very Poor. Measurements were made to determine whether the existing label and the label results predicted by the models were accurate. According to the experimental results with the same test sets used as 10 training sets (Table 3), the prediction accuracy of the final average classification was found to be 91.3%. From the experimental results obtained using the actually measured data, it was found that the prediction model of the SVM quality management label showed excellent performance.

4.3 Experiment on prediction of short-term quality

In the next experiment, short-term quality was predicted using the ARIMA model for the data collected through the developed rancidity sensor. The accuracy was calculated by determining whether the predicted values were similar to the rancidity values actually measured.
in the field using edible oils. We collected and predicted data using the following order.
1. Receive actually measured values of rancidity in the field.
2. Generate a prediction and analysis model for the rancidity values collected thus far through
data learning and predict the next values whenever new data are received.
3. Predict rancidity values on previous dates and determine the prediction accuracy by
comparison with actually measured values.

The data are configured as follows.
- Data set 1: Measured data of rancidity (04/05 21:10–04/06 15:27 in 2018) (Fig. 7 and Table 4)
- Data set 2: Measured data of rancidity (04/06 01:57–04/06 20:49 in 2018) (Fig. 8 and Table 5)
- Data set 3: Measured data of rancidity (04/06 16:28–04/07 01:43 in 2018) (Fig. 9 and Table 6)
- Data set 4: Measured data of rancidity (04/06 21:49–04/07 17:31 in 2018) (Fig. 10 and Table 7)
- Data set 5: Measured data of rancidity (04/07 02:43–04/07 23:39 in 2018) (Fig. 11 and Table 8)

Experimental results for each data set are as follows.

We repeated each experiment 10 times and measured the average of the accuracy results.
According to the standard deviation and dispersion results, individual error values do not show

![Fig. 7](Color) Results of comparison using data set 1.

![Fig. 8](Color) Results of comparison using data set 2.

### Table 4
Results of comparison using data set 1.

| Actual value | Prediction value | Error  |
|--------------|------------------|--------|
| am 2:43:21   | 8                | 7.867626 | 0.132374 |
| pm 2:31:03   | 8.7              | 8.527008 | 0.172992 |
| pm 3:31:03   | 8.7              | 8.664343 | 0.035657 |
| pm 4:20:03   | 8.7              | 8.692595 | 0.007405 |
| pm 5:31:03   | 8.9              | 8.856986 | 0.043014 |
| pm 6:24:03   | 9.1              | 9.049657 | 0.050343 |
| pm 7:49:24   | 9.9              | 9.725352 | 0.174648 |
| pm 8:39:24   | 11               | 10.75061 | 0.249393 |
| pm 9:49:24   | 12.8             | 12.44614 | 0.353858 |
| pm 10:39:24  | 13.5             | 13.33842 | 0.161585 | 0.138127 |

### Table 5
Results of comparison using data set 2.

| Actual value | Prediction value | Error  |
|--------------|------------------|--------|
| pm 9:49:24   | 13.1             | 13.09806 | 0.001939 |
| pm 10:39:24  | 13.5             | 13.49315 | 0.006846 |
| pm 11:43:21  | 14.1             | 14.09212 | 0.007885 |
| am 12:43:21  | 5                | 5.985541 | 0.985541 |
| am 1:43:21   | 7.5              | 7.363382 | 0.136618 |
| am 2:43:21   | 8                | 7.867626 | 0.132374 |
| pm 2:31:03   | 8.7              | 8.527008 | 0.172992 |
| pm 3:31:03   | 8.7              | 8.664343 | 0.035657 |
| pm 4:20:03   | 8.7              | 8.692595 | 0.007405 |
| pm 5:31:03   | 8.9              | 8.856986 | 0.043014 |
| pm 6:24:03   | 9.1              | 9.049657 | 0.050343 |
| pm 7:49:24   | 9.9              | 9.725352 | 0.174648 |
| pm 8:39:24   | 11               | 10.75061 | 0.249393 |
| pm 9:49:24   | 12.8             | 12.44614 | 0.353858 |
| pm 10:39:24  | 13.5             | 13.33842 | 0.161585 | 0.153027 |
a large difference (small dispersion: 0.002448) or a change (small standard deviation: 0.0494).
The accuracy of the error for overall predicted values can be determined by using average errors of the data. Since the deviation and dispersion are small, the accuracy of the model is considered to be excellent.
Table 9
Results of accuracy.

| Accuracy (%) | Scale 0–100 | Scale 1–24 |
|--------------|-------------|-------------|
| 99.862       | 96.668      |

The range of the average error for data sets 1–5 was about 0.138 from the actual values. Since rancidity values have a range of 0–100, the prediction accuracy calculated from the average value of the error is about 99.862% on the relevant scale (Table 9). Apart from that in special cases, the actual rancidity does not exceed 24. If managed normally, the error is about 3.332 when the accuracy is computed by adjustment for a rational range so that the accuracy is determined to be about 96.668%. Consequently, from the accuracy result, the performance of the model for short-term quality prediction in the present system is considered to be excellent.

5. Further Study and Conclusion

In this study, sensors allowing the measurement of rancidity have been developed, and a communication device capable of transmitting measurement data to the network has been proposed to analyze data obtained from the sensors. In addition, a data analysis system has been constructed, and a system where rancidity states are determined and changes are predicted by the analysis of the collected rancidity data by machine learning has been proposed.

The proposed system is an intelligent system that allows the monitoring of the rancidity of raw materials in the field by real-time measurement, the determination of quality states for raw materials through analysis, and the prediction of changes. For the verification of the performance, the performance of the sensors has been evaluated by measuring the rancidity of oils at actual businesses. In addition, classification and prediction experiments have been performed by analyzing the collected data using the proposed technique. The results showed excellent performance levels of 91.3% for classification and 96.668% for prediction.

For further study, the development of a dispersion processing system capable of real-time processing/analysis is planned once multiple data have been collected from experiments using raw materials other than oil and multiple sensors. Figure 12 schematically shows joint cutting using IoT and bigdata analysis in a road pavement quality management system. The proposed system can be applied as a quality control system in various fields when combined with not only saturation sensors but also widely available temperature sensors. As a representative application, data are being collected using IoT for road pavement quality management and building information modeling (BIM) area construction, and raw material and construction quality management is being studied. In the case of asphalt and concrete, rancidity, temperature, and humidity are important quality control factors. Therefore, analyzing and predicting the changes in rancidity, temperature, and humidity can predict the quality characteristics (e.g., strength) of asphalt and concrete. According to the analysis results, it is possible to predict the optimal timing of the construction process such as joint cutting. The training set consists of IoT sensor data, temperature, rancidity, humidity, weather data, and source information data. The data set is redefined as feature vectors based on factors that affect the strength of concrete.
Figure 13 shows a system for further study. We will study a road pavement quality management technology that can provide automated information using IoT and a big data platform. The system collects data from actual sites, such as rancidity using IoT sensors. We will analyze the road paving quality by analyzing the data using the accelerated big data platform and providing the prediction data.
Acknowledgments

This study was conducted under a research project [Development of road pavement on-site quality management system using IoT technology (19POQW-B152690)] funded by the Ministry of Land, Infrastructure and Transport (MOLIT) and the Korea Agency for Infrastructure Technology Advancement (KAIA). The authors would like to thank the members of the research team, MOLIT, and KAIA for their guidance and support throughout the project.

References

1. A. R. Youn, K. Y. Han, S. Y. Oh, and B. S. Noh: Food Eng. Progress 9 (2005) 309.
2. R. Kulchan, W. Boonsupthip, T. Jinkarn, and P. Suppakul: Int. Food Res. J., 23 (2016) 1092. https://doi.org/10.1016/j.ifrj.2016.01.003
3. S. Das and G. Chatterjee: Int. J. HIT Trans. ECCN 3 (2017) 72.
4. A. Abdelgawad and K. Yelamarthi: Wireless Commun. Mobile Comput. 2017 (2017). https://doi.org/10.10119/wf-iot.2018.355094
5. A. Kumar: Manu. Lett. 15 (2018) 122. https://doi.org/10.1016/j.mlfglet.2017.12.014
6. F. Alam, R. Mehmood, I. Katib, and A. Albeshri: Proc. Comput. Sci. 98 (2016) 437. https://doi.org/10.1016/j.procs.2016.09.068
7. M. S. Mahdavinejad, M. Rezvan, M. Barekatain, P. Adibi, P. Barnaghi, and A. P. Sheth: Digital Commun. Netw. 4 (2018) 161.
8. F. Chen, P. Deng, J. Wan, D. Zhang, A. V. Vasilakos, and X. Rong: IJDSN 11 (2015) 431047. https://doi.org/10.1016/j.ijdsn.2015.04.047
9. S. Capone, M. Epifani, F. Quaranta, P. Siciliano, A. Taurino, and L. Vasanelli: Sens. Actuators, B 78 (2001) 174. https://doi.org/10.1016/s0925-4005(01)00809-7
10. J. E. Haugen, F. Lundby, J. P. Wold, and A. Veberg: Sens. Actuators, B 116 (2006) 78. https://doi.org/10.1016/j.snb.2005.12.060
11. J.-W. Lee, S.-Y. Lee, and J.-S. Moon: JKIISC 24 (2014) 87. https://doi.org/10.13089/JKIISC.2014.24.1.87
12. V. Kotu and B. Deshpande: Predictive analytics and data mining: concepts and practice with rapid miner (Morgan Kaufmann, 2014). https://doi.org/10.1016/b978-0-12-801460-8.00009-4
13. K. Byungsam: Food Preservation and Processing Industry 9 (2010) 32.
14. D. Asteriou and S. G. Hall: ARIMA Models and the Box–Jenkins Methodology (Applied Econometrics, Palgrave MacMillan, 2011) 2nd ed., p. 265. https://doi.org/10.1057/978-1-377-41547-9.13
15. G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung: Time Series Analysis: Forecasting and Control (John Wiley & Sons. Box, George, Jenkins, Gwilym, 1970). https://doi.org/10.1111/j.1467-9892.2009.00643.x

About the Authors

Sung-Sam Hong received his B.S. degree from Gachon University, Korea, in 2009 and his M.S. and Ph.D. degrees from the Gachon University, Korea, in 2011 and 2016, respectively. Since 2016, he has been a research professor at Gachon University, Korea. His research interests are in IoT, bigdata analysis, machine learning, proactive prediction system.
(sungsamhong0@gachon.ac.kr)

Kisoo Chang is a professor in the Department of Computer Engineering at Gachon University. He received his B.E. degree from Yonsei University, and his M.E. and Ph.D. degrees from KAIST, Korea, in 1997 and 2012, respectively. His research interest are in communication networking and electric vehicle charging protocol. (kschang@gachon.ac.kr)
Junhyung Lee is a postdoctoral researcher at the Korea Institute of Civil Engineering and Building Technology. He received his B.E. and Ph.D. degrees from Yonsei University, Korea, in 2013 and 2019, respectively. His research interest are in traffic flow theory, uninterrupted traffic flow, merging bottleneck, and autonomous vehicle driving algorithm.
(takion0424@yonsei.ac.kr)

Byungkon Kim is a research fellow at the Korea Institute of Civil Engineering and Building Technology. He received his B.S., M.S. and Ph.D. degrees from Hanyang University, Korea, in 1991, 1993, and 2013, respectively. His research interest are in real-time systems, construction information, construction automation, and building information modeling.
(bkkim@kict.re.kr)