Integrating random forest model and internet of things-based sensor for smart poultry farm monitoring system

Imam Fahrurrozi¹,², Wahyono¹, Yunita Sari¹, Anny Kartika Sari¹, Ilona Usuman¹, Bambang Ariyadi³

¹Department of Computer Science and Electronics, Faculty of Mathematics and Natural Sciences, Universitas Gadjah Mada, Yogyakarta, Indonesia
²Department of Electrical Engineering and Informatics, Vocational College, Universitas Gadjah Mada, Yogyakarta, Indonesia
³Department of Animal Production, Faculty of Animal Science, Universitas Gadjah Mada, Yogyakarta, Indonesia

ABSTRACT

The global poultry industry has encountered growing concerns related to foodborne illnesses, misuse of antibiotics, and environmental impacts. To tackle these issues, this study aims to develop an intelligent poultry farm with real-time environmental monitoring and predictive models. The primary objective is to combine a machine learning-based prediction model with internet of things (IoT) devices to gather and analyze environmental data, such as temperature, humidity, and ammonia levels, to forecast the conditions within poultry houses. These sensor data and additional information, such as feed consumption, water consumption, poultry weight, capacity, and poultry house dimensions will serve as inputs for supervised machine learning models. Among these models, the proposed random forest (RF) model, when augmented with timestamp features, achieves the highest accuracy rate of 96.665%, surpassing other models such as logistic regression (LR), k-nearest neighbor (KNN), decision tree (DT), adaptive boosting (AdaBoost), extreme gradient boosting (XGBoost), support vector machine (SVM), and multi-layer perceptron (MLP) in identifying poultry house conditions. Additionally, this study demonstrates how the trained model can be effectively applied in a web-based monitoring system, delivering real-time data to farmers for well-informed decision-making and ultimately enhancing productivity in smart poultry farming.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Wahyono
Department of Computer Science and Electronics, Faculty of Mathematics and Natural Sciences
Universitas Gadjah Mada
North Sekip, Bulak Sumur, Yogyakarta 55281, Indonesia
Email: wahyo@ugm.ac.id

1. INTRODUCTION

Modern technology, including internet of things (IoT), has improved farming in various fields, especially agriculture [1]–[5]. IoT has been especially beneficial in poultry farming, enabling comprehensive monitoring and decision-making through sensors, cloud technology, and innovations [6]–[10]. This allows farmers to access real-time data about their poultry houses via sensors and gadgets, ultimately improving the reliability and adaptability of farming systems [11]–[13]. Furthermore, IoT’s continued exploration and adoption in agriculture signify ongoing advancements in optimizing farm management practices.

Controlling temperature, humidity, and ammonia levels in broiler chicken production aims to create healthier chickens that thrive and adapt to their environment. Maintaining the right temperature is essential for the comfort and well-being of broiler chickens, ensuring their health from when they are placed in the

Journal homepage: http://ijee.cs.iaescore.com
In our study, we designed an IoT sensor device for monitoring temperature, humidity, and ammonia levels within the broiler house. This IoT device comprises a Mappi32 microcontroller, DHT22 for recording humidity and temperature, and MQ137 as a sensor for measuring ammonia levels. Figure 2 shows IoT devices installed inside the broiler house to collect environmental data, including temperature, humidity, and ammonia levels.

2. **METHOD**

The model illustrated in Figure 1 utilizes IoT sensors such as temperature, humidity, and ammonia levels to forecast the conditions within the poultry house, determining whether they are normal or abnormal. We removed inconsistent entries through preprocessing and resampled the data at 5-minute intervals to create the time-series dataset. The dataset collected underwent a feature extraction stage, which involved the addition of timestamp features. We employed the random forest algorithm for prediction purposes and evaluated the model's effectiveness by comparing it to alternative machine learning models. The model assessment followed a hold-out method, where the dataset was split into training and testing sets in a 70:30 ratio. Finally, the trained model was integrated into a web-based application at the deployment stage, enhancing accessibility for end users.

2.1. **Dataset**

In our study, we designed an IoT sensor device for monitoring temperature, humidity, and ammonia levels within the broiler house. This IoT device comprises a Mappi32 microcontroller, DHT22 for recording humidity and temperature, and MQ137 as a sensor for measuring ammonia levels. Figure 2 shows IoT devices installed inside the broiler house to collect environmental data, including temperature, humidity, and ammonia levels. Inadequate humidity can cause dehydration and respiratory problems [18], [19], while excessive humidity can lead to heat prostration [20]. Additionally, monitoring ammonia levels in the broiler chicken house is crucial, as it can stem from gases produced by chicken waste in the coop [21]–[23]. IoT technology can be incorporated into a monitoring system to gain insights into a poultry house's temperature, humidity, and ammonia levels.

Previous studies have demonstrated the successful application of IoT technology in poultry farming, yielding noteworthy outcomes. Hambali et al. [24] focus on leveraging the IoT and wireless sensor networks (WSN) to automate the monitoring and management of key factors like temperature, humidity, air quality, and food feeding in poultry farms, aiming to improve chicken health and reduce mortality rates. By implementing a prototype system that initiates corrective actions when parameters exceed threshold values, along with automatic alert notifications through SMS, Email, and WhatsApp, the study seeks to enhance poultry productivity and overall farm efficiency through machine learning and IoT technologies. Zheng et al. [25] focus on developing an information management system for poultry farming, integrating IoT technology to collect, transmit, store, and manage data, ultimately enhancing production efficiency and forming a comprehensive data chain. The system supports production management and includes an office management module with disease detection capabilities, laying the groundwork for potential future intelligent poultry farming management systems based on cloud services and big data technology.

Machine learning-based prediction models have found application in various domains, including their use in identifying human loitering through vision sensor technology within surveillance systems [26], to predict gas-fired boiler flue gas oxygen content [27] and human activity recognition [28], [29]. Moreover, the use of IoT sensor data as input for machine learning models has been previously utilized and shown favorable results in predicting the environmental conditions within poultry houses. Using machine learning, Küçüktopcu and Cemek [30] aimed to create an efficient model for predicting ammonia (NH3) concentration in poultry farms. Four different models, including multi-layer perceptron (MLP), adaptive neuro-fuzzy inference systems with grid partitioning (ANFIS-GP) and subtractive clustering (ANFIS-SC), and multiple linear regression analysis (MLR), were applied using easily obtainable climatic variables and litter quality properties. The results revealed that the ANFIS-SC model, utilizing air temperature, air relative humidity, and airspeed as inputs, performed the best in predicting NH3 concentration, demonstrating its potential as a valuable tool for rapid and accurate estimation in poultry farm management. Liu et al. [31] explored the use of machine learning models, including extreme gradient boosting (XGBoost), support vector regression (SVR), and back-propagation neural networks (BPNN), to predict odor concentration in laying hen houses. The XGBoost model demonstrated the highest predictive accuracy (R2 = 0.88), indicating its potential for timely and moderately accurate odor monitoring.

Nevertheless, there needs to be more research regarding the application of random forest, a supervised machine learning algorithm, for poultry house detection using IoT sensor data. Consequently, this study proposes a predictive model that employs random forest to forecast poultry house conditions based on IoT sensor data, encompassing factors such as temperature, humidity, and ammonia levels. Additionally, we integrate supplementary data, including feed and water intake, broiler weight, allocation capacity, and house size. We introduce timestamp features as additional attributes for the random forest model to enhance prediction accuracy. Moreover, by incorporating this proposed predictive model into a web-based system, farmers can improve their decision-making processes and optimize their production strategies.
ammonia levels, at five-minute intervals throughout a production season, spanning approximately one month starting from January 5, 2023. The device was set up in a broiler house in the Sleman District, Yogyakarta Province, Indonesia.

Figure 1. Proposed ML model to detect broiler house conditions

Figure 2. IoT sensor setup installation inside broiler house

The IoT device gathers sensor data, sends this information to a REST API on the server side, and stores it in a MySQL database. The farmer manually entered supplementary data, including feed and water intake, broiler weight, allocation capacity, and house size. These manually inputted data were then integrated with the IoT dataset for subsequent analysis. Finally, our dataset contains a total of 9196 records, with average values of 26.86 for temperature (celsius), 83.46 for humidity (%), 6.40 for ammonia levels (ppm), 659.99 for feed intake (kg), 346 for water intake (liter), 801.64 for broiler weight (gram), 8000 for allocation capacity (broilers quantity), and 336 for house size (square meter).

2.2. Feature extraction

The sensor data collected included temperature, humidity, and ammonia levels. In contrast, the manually inputted data by the farmers encompassed feed intake, water intake, broiler weight, allocation capacity, and broiler house size. These two datasets were merged into a unified dataset, each entry accompanied by date and time information at 5-minute intervals. The proposed timestamp features, such as the “hour of the day” and “part of the day,” were derived from the date and time column. The "part of the day" categories were defined as early morning (4 < x ≤ 8), morning (8 < x ≤ 12), afternoon (12 < x ≤ 16), evening (16 < x ≤ 20), night (20 < x ≤ 24), and late night (x ≤ 4). We applied label encoding, representing "early morning" as 0, "morning" as 1, "afternoon" as 2, "evening" as 3, "night" as 4, and "late night" as 5.
The initial processing phase is crucial for converting the gathered dataset into an input matrix $X$ and an output vector $Y$. This transformation allows conventional machine learning models to recognize patterns and make predictions. With a dataset consisting of $m$ distinct sensor data readings, hour of the day ($h$), part of the day ($p$), and a total of 10 features, the input matrix $X$ can be represented as a $[m \times 10]$ matrix.

$$X = \begin{bmatrix} \text{temp}_1 & \text{hum}_1 & \text{ammonia}_1 & \text{feed}_1 & \ldots & h_1 & p_1 \\ \text{temp}_2 & \text{hum}_2 & \text{ammonia}_2 & \text{feed}_2 & \ldots & h_2 & p_2 \\ \vdots & \vdots & \vdots & \vdots & \ldots & \vdots & \vdots \\ \text{temp}_{m-1} & \text{hum}_{m-1} & \text{ammonia}_{m-1} & \text{feed}_{m-1} & \ldots & h_{m-1} & p_{m-1} \\ \text{temp}_m & \text{hum}_m & \text{ammonia}_m & \text{feed}_m & \ldots & h_m & p_m \end{bmatrix}$$

(1)

Each instance corresponds to a broiler house’s condition, categorized as either "normal" or "abnormal," represented as $y_m$. In the end, the target output $Y$ was organized as a $[m \times 1]$ vector.

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{m-1} \\ y_m \end{bmatrix}$$

(2)

We conducted interviews with farmers to determine the criteria for abnormal conditions, defining it as a situation where the number of broiler deaths $>10$ during the hour of the day. Subsequently, we labeled $y$ as "normal" (when broiler deaths are $< 10$) and "abnormal" (when broiler deaths are $\geq 10$). As mentioned earlier, we applied label encoding, representing labels "normal" as 0 and "abnormal" as 1.

### 2.3. Random forest

The random forest (RF) algorithm is a classification method that combines decision trees [32]. Previous research has shown that using a randomization approach, such as bagging or the random space method, can improve the performance of RF. This randomization is achieved using bootstrapped sampling of the original data and randomly selecting a subset of features at each node to determine the best split. The process of generating each tree in an RF model is described in Algorithm 1.

**Algorithm 1. Random forest**

1. **Input**: The dataset used for training $D_t$, ensemble size denoted as $T$, and subspace dimension referred to as $d$
2. **Output**: Consensus decision derived from the majority of tree models
3. for $t = 1$ to $T$ do
4. Create a bootstrapped sample $D_t$ from $D$
5. Randomly choose $d$ features and reduce the dimensionality of dataset $D_t$ accordingly
6. Develop a tree model through training $M_t$ on $D_t$
7. Divide the dataset based on the most suitable feature among the selected $d$ features.
8. Allow the tree $M_t$ to grow without applying pruning techniques
9. End

The random forest technique involves the creation of individual decision trees through the random selection of a subset of attributes at each node for making splits. The procedure functions as follows: given a training dataset ($D$), the quantity of trees ($T$) within the model, a subspace dimension ($d$), and the available features ($F$), a bootstrapped sample ($D_t$) is derived from the original dataset ($D$). This sample incorporates certain records from the original dataset multiple times and omits others. Subsequently, a subset ($d$) of attributes is randomly picked from the bootstrapped sample ($D_t$) to be considered candidates for splits at each node. The decision tree classifier is trained on this bootstrapped sample ($D_t$) along with the selected attributes ($d$), and it is grown to its maximum extent without pruning. This process is iterated for all trees within the forest. During the classification phase, each tree votes, and the most prevalent class is designated as the predicted outcome.

Random forests (RFs) can tackle various predicaments encountered by decision trees, such as averting overfitting and minimizing variance. The random forest model was trained to comprehend two classes (normal or abnormal) using the prepared training dataset in this investigation. The predictive results of this model were matched against the testing set to assess its accuracy. The input attributes in this study consist of sensor data such as temperature, humidity, ammonia, feed, water, weight, allocation, area size,
Addresses the suggested model's effectiveness and timestamp features' influence on its classification models. The proposed model's classification models displayed an increase of approximately 8.571% compared to incorporating the timestamp feature. Accuracy are provided in Figure 2, achieving improvements of up to 96.665% in accuracy, 94.263% in precision, and 94.793% in specificity, and F1-score metrics.

3. RESULTS AND DISCUSSION

This section addresses the suggested model's effectiveness and timestamp features' influence on its performance. Additionally, we showcase the model's practicality by implementing it within a web-based monitoring system. This implementation underscores the model's versatility and highlights its potential for seamless integration into existing monitoring infrastructures, thereby enhancing its applicability in practical settings.

3.1. Performance of machine learning models

In this study, supervised machine learning techniques were employed to anticipate the environmental conditions of poultry farms as either normal or abnormal. This was achieved using input data collected by an IoT sensor device and other data. The focus was on evaluating the accuracy of these machine-learning models. A comparison was made between a model using a random forest combined with additional timestamp features and other classification models (without additional timestamp features) for predicting poultry farm environmental conditions. Various machine learning algorithms, including logistic regression (LR), K-nearest neighbor (KNN), decision tree (DT), adaptive boosting (AdaBoost), extreme gradient boosting (XGBoost), support vector machine (SVM), and multi-layer perceptron (MLP), were employed as classification models. The outcomes presented in Table 1 demonstrate the diverse model performances regarding accuracy, precision, recall, and f-score. The results indicated that the proposed model exhibited superior performance to the alternative models, achieving improvements of up to 96.665% in accuracy, 94.263% in precision, 94.793% in specificity, 94.793% in recall, and 94.525% in f-score.

### Table 1. Performance evaluation results

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Specificity</th>
<th>Recall</th>
<th>F1 Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logistic regression</td>
<td>81.406</td>
<td>40.703</td>
<td>50.000</td>
<td>50.000</td>
<td>44.875</td>
</tr>
<tr>
<td>KNN</td>
<td>84.487</td>
<td>77.908</td>
<td>63.022</td>
<td>63.022</td>
<td>65.978</td>
</tr>
<tr>
<td>Decision tree</td>
<td>83.762</td>
<td>73.707</td>
<td>65.134</td>
<td>65.134</td>
<td>67.653</td>
</tr>
<tr>
<td>AdaBoost</td>
<td>81.877</td>
<td>69.589</td>
<td>54.651</td>
<td>54.651</td>
<td>54.315</td>
</tr>
<tr>
<td>XGBoost</td>
<td>85.067</td>
<td>78.759</td>
<td>65.033</td>
<td>65.033</td>
<td>68.284</td>
</tr>
<tr>
<td>SVM</td>
<td>81.623</td>
<td>73.372</td>
<td>51.111</td>
<td>51.111</td>
<td>47.353</td>
</tr>
<tr>
<td>MLP</td>
<td>83.870</td>
<td>78.590</td>
<td>59.711</td>
<td>59.711</td>
<td>61.857</td>
</tr>
<tr>
<td>Proposed random forest + timestamps feature</td>
<td>96.665</td>
<td>94.263</td>
<td>94.793</td>
<td>94.793</td>
<td>94.525</td>
</tr>
</tbody>
</table>

3.2. Impact of timestamps feature on model performance

In this study, we focused on investigating how the inclusion of a timestamp feature influences the accuracy of classification models. Our observations indicate that utilizing these methods led to an enhancement in model accuracy. Upon introducing timestamp features such as the hour of the day and parts of the day, the classification models displayed an increase of approximately 8.571% compared to conventional machine learning models. Detailed insights into the influence of the timestamp feature on classification accuracy are provided in Figure 3. Across our dataset, incorporating the timestamp feature consistently elevated the accuracy of all classification models. To sum up, the integration of timestamp features into classification models has the potential to enhance overall model accuracy.

3.3. Comparison with previous studies

In this section, we performed a comparative examination of our study and previous research that concentrated on monitoring systems utilizing IoT and the detection of broiler house conditions. Within an IoT-based monitoring system framework, Hambali et al. [24] introduced a concept for a smart poultry farm in Brunei that utilizes IoT technology. Data from sensors measuring temperature, humidity, air quality, and food feeding were gathered, revealing that implementing IoT and mobile technology in poultry farming can effectively

---

*Integrating random forest model and internet of things-based sensor for smart ... (Imam Fahrurrozi)*
manage environmental conditions and decrease poultry farm mortality rates in Brunei. Zheng et al. [25] introduced the development and deployment of a poultry farming information management system using a cloud database. They employed IoT sensors within contemporary Chinese poultry farms to gather data on environmental parameters like temperature, humidity, light intensity, and gas levels within the poultry houses. Finally, Mondol et al. [34] introduced an IoT-based smart weather monitoring system tailored for poultry farms in Bangladesh, employing temperature sensors to enhance farm management and efficiency.

Incorporating IoT technology and predictive models has been put into practice and yielded notable outcomes. Küçüktopcu and Cemek [30] aimed to develop a precise and cost-effective model for predicting ammonia (NH3) concentration in poultry farms using machine learning techniques. Among the models tested, the ANFIS-SC model, utilizing inputs such as air temperature, air relative humidity, and airspeed, demonstrated the highest accuracy in estimating NH3 concentration, making it a promising tool for rapidly and accurately assessing NH3 levels in poultry farm management. Liu et al. [31] investigated the application of machine learning techniques, such as extreme gradient boosting (XGBoost), support vector regression (SVR), and back-propagation neural networks (BPNN), to forecast odor levels within laying hen enclosures. The findings showed that the XGBoost model exhibited the most substantial predictive capability (R2 = 0.88), suggesting its suitability for timely and reasonably accurate odor surveillance.

Our study introduced a novel approach for assessing poultry house conditions by integrating IoT technology with machine learning models. Distinguishing from earlier studies, our method employed IoT sensor data as input and harnessed the power of the random forest algorithm to enhance predictive accuracy. Moreover, we successfully deployed our trained model within web-based applications, offering farmers real-time insights into poultry house conditions.

3.4. Practical application

This study aims to develop a web-based system that employs a machine learning model to forecast the environmental status of poultry farms precisely and aid in managerial decision-making. Prior investigations have indicated the utility of such a system in online traceability [35], inventory management [36], and disease prognosis [37], [38]. A machine learning-based predictive model makes it possible to accurately discern whether the conditions are normal or unusual. The web-based monitoring system was constructed using the PHP programming language and a MySQL database on the server side. Python was employed for the REST API and machine learning model. The predictive model was established using the Flask web framework and the Scikit-learn library on the server side. This model was employed to identify the condition of the poultry houses. We designed an IoT sensor device that utilizes an MQ137 sensor to measure ammonia levels (with a range from 5 to 500 ppm), a DHT22 sensor to monitor temperature and humidity levels (within a temperature range of -40 to +125 degrees Celsius, and a humidity range from 0 to 100%). We employed the Mappi32 microcontroller, as depicted in Figure 4. The trained model was employed to forecast whether the environmental conditions were normal or abnormal. The outcomes were then presented to the management through a web-based interface, as shown in Figure 5.
Integrating random forest model and internet of things-based sensor for smart … (Imam Fahrurrozi)

4. CONCLUSION

Farmers commonly engage in manual assessment of the state of poultry farms. The manual evaluation of poultry farm conditions demands a significant amount of time. Conversely, using information technology simplifies farmers’ tasks by expediting the decision-making procedure. Recent progress in monitoring systems for poultry farms has emerged through the integration of machine learning algorithms. In this study, a monitoring system was developed for poultry farms, utilizing a machine learning prediction model to anticipate environmental conditions. The proposed RF model and additional timestamp features successfully determined the farm’s state using IoT-based sensors. It underwent testing across diverse scenarios, encompassing both normal and abnormal conditions. The outcomes indicated the superior performance of the proposed model in comparison to others, such as MLP, LR, KNN, DT, SVM, XGBoost, and AdaBoost, showcasing improvements of up to 96.665%, 94.263%, 94.793%, 94.793%, and 94.525% concerning accuracy, precision, specificity, recall, and f-score, respectively. The trained model could be seamlessly integrated into the server side, receiving data from IoT devices to predict environmental conditions. This data could be leveraged to enhance decision-making for farmers, optimizing their production strategies. In forthcoming endeavors, the effectiveness of IoT sensors will be assessed across varied conditions. The potential for comparing the model’s performance with alternative forecasting methods to predict future poultry farm conditions remains possible. These future endeavors aim to enhance the model’s adaptability and contribute to broader advancements in precision agriculture and farm management practices.

REFERENCES


Integrating random forest model and internet of things-based sensor for smart ...

(Imam Fahrurrozi)

BIOGRAPHIES OF AUTHORS

Imam Fahrurrozi is a Doctoral Student at the Department of Computer Science and Electronics, Universitas Gadjah Mada, and a full-time lecturer at the Department of Electrical Engineering and Informatics, Universitas Gadjah Mada, Indonesia. His research areas are artificial intelligence, IoT, machine learning, and recommender systems. He is a co-founder of KIRANA (Kios Rakyat Indonesia), a technology-based e-commerce company for small and medium enterprises (UMKM). He can be contacted at email: imam.fahrurrozi@ugm.ac.id.

Wahyono received the B.Sc. degree in Computer Science from Universitas Gadjah Mada, Indonesia, and the Ph.D. degree from The University of Ulsan, South Korea. He is an Associate Professor at the Department of Computer Science and Electronics, Universitas Gadjah Mada, Indonesia. His research interests include machine learning, computer vision, and pattern recognition. He actively participates as a member of the societies, a reviewer in reputable international journals, and an editor in several journals. He can be contacted at email: wahyo@ugm.ac.id.

Yunita Sari is a full-time lecturer at the Department of Computer Science and Electronics, Universitas Gadjah Mada, Indonesia. She received the B.Sc. degree in Computer Science from Universitas Gadjah Mada Indonesia, Master degree from Universiti Teknologi PETRONAS (UTP) Malaysia, and the Ph.D. degree from The University of Sheffield UK. Her research interests include natural language processing, authorship attribution, author profiling, adversarial stylometry, authorship verification, and deep learning for text analysis. She can be contacted at email: yunita.sari@ugm.ac.id.

Anny Kartika Sari received Ph.D. degree in Computer Science from La Trobe University, Australia. She is a full-time lecturer at the Department of Computer Science and Electronics, Universitas Gadjah Mada, Indonesia. Her research areas are ontology and knowledge representation, algorithms, and software engineering. She can be contacted at email: a_kartikasari@ugm.ac.id.

Ilona Usuman received a Ph.D degree from Sakarya University, Turkey. Her research areas are robotics, control systems, and smart home. She is a full-time lecturer at the Department of Computer Science and Electronics, Universitas Gadjah Mada, Indonesia. She can be contacted at email: ilona@ugm.ac.id.

Bambang Ariyadi received Ph.D degree from Hiroshima University, Japan. He is a full-time lecturer at the Department of Animal Production, Faculty of Animal Science, Universitas Gadjah Mada, Indonesia. His research areas are agricultural and veterinary sciences, poultry sciences, and animal production. He can be contacted at email: bambang.ariyadi@ugm.ac.id.