

Development of a Small Intelligent Weather Station for Agricultural Applications

Yi-Hua Chung¹, Jun-Fu Huang², Yuan-Chen Hu², Chen-Kang Huang^{2,*}

¹Department of Chemistry, National Taiwan University, Taipei, Taiwan

²Department of Biomechatronics Engineering, National Taiwan University, Taipei, Taiwan

Received 29 September 2020; received in revised form 08 December 2020; accepted 01 February 2021

DOI: <https://doi.org/10.46604/aiti.2021.6494>

Abstract

It is known that climate change causes a decrease in the profit gained from agricultural production. This work designs and establishes weather boxes equipped with functions of rainfall prediction, frosting forecast, and lightning detection. With the wireless connection and the build-in decision mode, weather boxes can deliver early-warning by sending texting messages to the users and actuating the corresponding action to response the extreme climate. To implement rainfall and frosting prognostication, two different datasets are analyzed by the technology of data mining. One of the datasets is acquired from the Central Weather Bureau, and the other is from the proposed weather box monitoring the agricultural environment. From the experimental results, the prediction model constructed from the data which is collected by the proposed weather box exhibits a higher accuracy in rainfall forecasting than those based on the Central Weather Bureau.

Keywords: weather box, rainfall, frosting, lightning, early warning

1. Introduction

Climate change and agriculture gain have a reciprocal relationship. Climate change may cause a decrease in the profit gained from agricultural production. Therefore, accurate weather forecasting becomes an important but challenging task to researchers for a long time. In addition, numerous people are aware of climate-changing. For example, orchardists pay attention to the amount of rainfall, which affects the fruit quality. The frosting, which will impact the flavor and aromatics of the tea, is concerned by tea growers. Lightning is responsible for plantation fires and the death of the crops [1-3]. Recently, the continuous advances in sensors, sensing systems, and the Internet of Things (IoT) [4-5] technology significantly impact several fields. For instance, in the agriculture fields, to accomplish the task of the remote environment monitoring for the farmland, developers can utilize single-board microcontrollers or computers to connect with various sensors to collect and accumulate a large amount of local environmental data at first. The data can be analyzed to get a more undiscovered relationship between the data via the technology of data mining [6-12]. After all, those relations can be used to achieve the goal of future predicting.

Currently, some companies and organizations, such as airports and agriculture agencies, collect and accumulate the weather data by their own weather stations to fulfill some of their particular needs. The data, which is for weather monitoring or predictions, become more precise when a greater amount of data are collected and analyzed.

* Corresponding author. E-mail address: ckhuang94530@ntu.edu.tw

Tel.: +886-2-3366-5351

Official weather stations collecting professional data are precious. However, these weather stations are mostly found in the research institutions located in urban area far from the remote agricultural locations, which may cause the grabbed data to have some potentially hidden problems. Fortunately, the concepts, such as free software and free hardware can significantly reduce the cost of a station. To construct a low-cost weather station, the developers can modify the existing version to reduce or increase the functions and costs based on the needs. Sensors are elements that translate a usually non-electrical value to an electrical value, which can be measured, amplified, or even modified. To manage and control these sensors and share the collected data, some new low-cost single board computers may be used, for example, Arduino and Raspberry Pi.

It is commonly known that the Central Weather Bureau (CWB) has already set up numerous weather stations nowadays. Nevertheless, most of the weather stations are built in the megalopolis instead of some distant regions. Furthermore, most of the commercially available weather stations are too expensive to be afforded. Low cost weather stations are required to set up in remote agricultural area. The functions of the weather boxes constructed in this work are shown in Fig. 1. This paper is purposed to construct a weather box with the ability to perform rainfall, frosting prediction, and lightning detection.

For the first aim, rainfall prognostication, a weather box is constructed with the WemosD1 choosing to be the microcontroller. Three sensors, BME280, FC-37, and GY-49 are used to sense the humidity, temperature, pressure, rain, and solar radiation as the local weather parameters to accumulate the amounts of the local weather data in three experimental fields, including Taipei city, Taoyuan city, and Yilan city. After stockpiling the local weather data, data sets from (1) Weather box or (2) the historical dataset released by the CWB, which consists of several atmospheric attributes, are analyzed by the techniques of data mining to extract the hidden relationships among weather parameters. Above all, as the weather box releases a heavy rain alerting by the prediction of the weather forecasting model, it can also send out the alarming message to the users by the communication application.

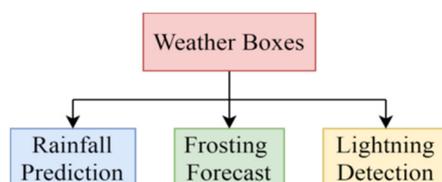


Fig. 1 Different weather boxes with various functions constructed in this work

Second, the frosting module is added into the system with both of the frosting point exploring and recording functions. When in the exploring mode, the testing surface is lowered by the thermoelectric cooler controlled by Raspberry Pi. During the surface temperature drop, photos are taken and analyzed thereafter. When the frost is first observed, the surface temperature is considered as the frosting point for the corresponding environment temperature, pressure, and relative humidity. When in the recording mode, a webcam is set to take photos at the testing surface regularly. The series of photos are analyzed to find out the frosting occurrence and its corresponding surface temperature.

Lastly, regarding the lightning detection, the weather box attained the greatest performance with precision = 66.7% and recall = 22.2% on May 27, 2020, in Taipei, and precision = 71.1% and recall = 45.8% on July 16, 2020, in Yilan. Obviously, the presentation of the detecting results in Yilan is better than those in Taipei as the section of result and discussion shown. The main causation for the poor performance of the recall is because of the small bandwidth for the detecting frequency of the lightning events. Unfortunately, the data released by the CWB is constructed by the CG flash and IC flash events. However, the AS3935 sensor can only detect some of the CG flash events, and all of the IC flash events can not be sensed by the AS3935 due to the narrow bandwidth of it. The principal reason for the subpar precision is on account of the detection method we chose, radio detection. Though we put effort into finding the experimental fields without noises, the system is still easy to be affected by the disturbers in the environment.

2. Literature Review

Low-cost weather stations were developed in previous literature [13]. To establish the weather station, R. C. Brito's team firstly proposed an architecture structured in layers, which considered the scalability of the system and allowed the connection of new sensors to support the needs of monitoring new data. Each feature in this structure was equipped with specific hardware, and the sensors would send the measured data back to the Arduino through different means, such as cable, analogically, or digitally. The Arduino used I²C communication protocol to exchange information with the Raspberry Pi. The Raspberry Pi was connected to the Internet through RJ-45 cable, which sent the data to a web server, the database. It also shared the data in REST format, which could be read by web pages, desktop applications, and mobile devices. The connection to the Internet was done through a WiFi or cellular (GSM, 3G, 4G) network. Proprietary resources weren't used in the creation of this weather station. All the elements, such as hardware, software, and communication protocols were free to use and low-cost to construct.

In the literature and this work, the classification in MATLAB toolbox, including Decision Tree [14], Support Vector Machine [15-17], K Nearest Neighbor [18], and Bagged Tree [19-20], was used to find the best classification to construct the rainfall prediction model of rainfall.

The last case is lightning detection. Discharge activity is a weather phenomenon, including two types. Lightning of cloud to cloud is called IC flash, and lightning cloud to ground is called CG flash. Besides, the CG lightning has a more harmful destructive power to the targets on the ground, including human beings, forest creatures, and so on. Commonly, there are three ways to detect the lightning event, inclusive of radio detection, radar detection, and optical lightning detection [21]. In [22], an intelligent lightning detection system was constructed. The system was built on a lightning signal collection module, central control module, display module, memory module, time module, and host computer module, which is established to cope with the problems of the high cost and the lacking the weather stations for detecting lightning. The 32-bit ARM STM32F103 was used to act as the core, microcontroller, of the system.

As the lightning occurs, the lightning signal acquisition module will output a signal pulse which implies the fast sampling intensity of the lightning in real-time, and the sampled data is transited to the ARM STM32F103. After receiving the data, the microcontroller will count the number of lightning and the strength through TFT color screen display, automatically record the time of lightning occurs, and store these data in the SD card. Meanwhile, via RS485 bus, the ARM STM32F103 will transmit processed data to the PC. monitoring software in the PC which allows the users to get monitoring instantaneously about the number of lightning occurring and the lightning intensity. Moreover, the data can be stored on a computer hard drive, with retrospective queries functions. After owning the ability to detect whenever lightning occurs, another task that needs to be solved is lightning positioning development. As the previous paragraph mentioned, the method of radar to detect lightning is mainly used for thunderstorm activity detection, and it is impossible to detect single lightning and its intensity parameters. Therefore, in order to judge the real placement that the lightning happens, the time-of-arrival (TOA) method and the magnetic direction method (MDF) are used in combination.

Referring to the main purpose of distant regions' remote environmental monitoring, the way of radio detection is chosen to be used to sense the lightning events, and the events of CG flash are mainly focused on. To complete lightning detection, the AS3935 [23] sensor is used to support our decision model to make the arbitration and the Raspberry Pi acts as the core. Two experimental fields, including National Taiwan University in Taipei city and the National Yilan University in Yilan city, are set. When the microcontroller is connected to the Internet, the collected local lightning events data will be uploaded to the Intelligent Biosensing Platform, an online website. Otherwise, the data is accumulated in the Micro SD card slotted on the microcontroller. After piling up the data, the data collected by the weather box is compared with the lightning events data released by the CWB, which is open access, as the benchmark to ensure the precision and the accuracy of the data collected by the weather box.

The results of rainfall prognostication show that the best prediction model that covered the data collected by the weather box is KNN (k=3). As for the data released by the CWB, Bagged Tree is the best. Besides, the Matthews correlation coefficient (MCC) [24] is used as the index to measure the qualities of the classifications used in the experiments. As the experimental result showed in the results section, the best MCC number of the models to predict rainfall before raining for 2-hr of the data collected by the weather box is 0.925 (KNN), which is better than 0.596 (Bagged Tree), the outstanding MCC of the data released by the CWB. Compared to the CWB data, the reasons why the data collected by the weather box, lead to more precise and instantaneous prediction models are on account of the higher location congestion of the weather stations and the more frequent interval of the detecting time. For the frosting prediction, the alarming would take the corresponding actions when the temperature < 5°C, less cloud cover, and the relative humidity greater than 60% were occurred simultaneously.

3. Material and Methods

3.1. Rainfall forecast

(1) Weather box construction

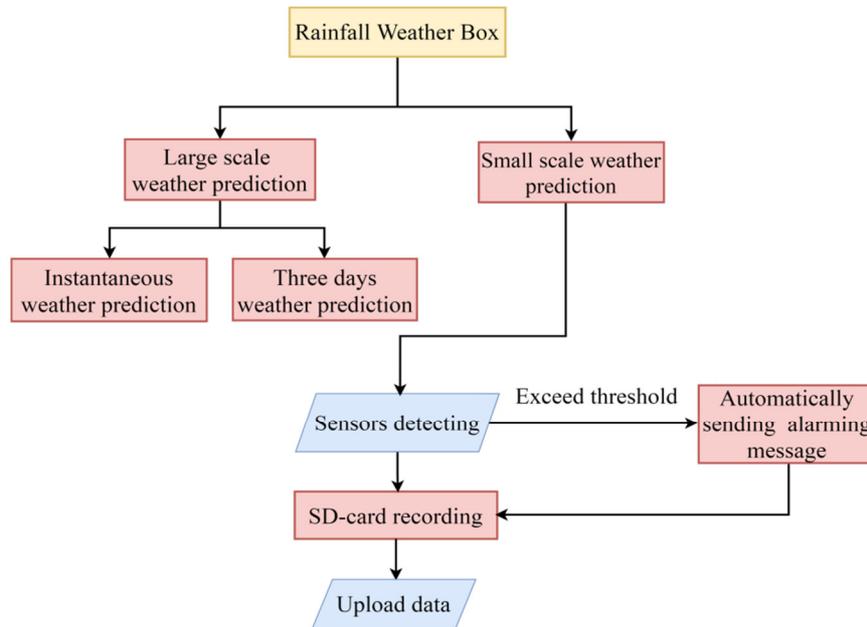


Fig. 2 The rainfall weather box designing flow chart

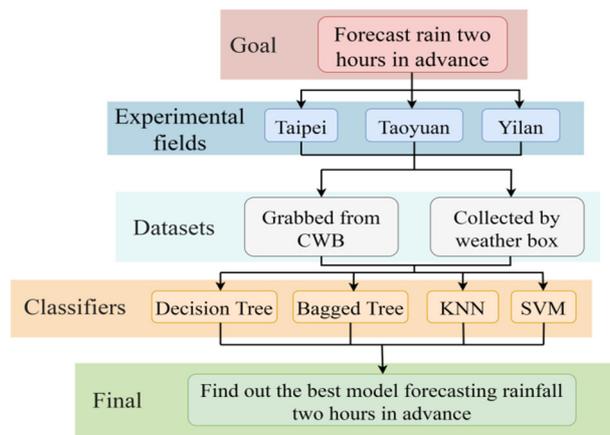


Fig. 3 The rainfall weather box experimental set-up diagram

To reach the goal of rainfall forecast, a weather box is constructed to collect the data, which has a coverage of a small area, in order to provide an instantaneous and short-range weather prediction that reflects the real local weather accurately. Considering the problem of accessing Internet [20] in the remote region and the cost of the weather box, WemosD1 is chosen

to be the core of the weather box for its' low-cost and self-contained Wi-Fi networking capabilities. The weather box of rainfall prediction was equipped with three different detecting sensors, including BME280, FC-37 and GY-49. They are used to sense the humidity, temperature, pressure, rain and radiation and used four li ion 186500 batteries in parallel as its' power supply. The weather box designing flow chart is shown in Fig. 2 and Fig. 3. Besides, the voltage given by the 3.7V batteries does not meet with the specification of the WemmosD1.

We cope with the dilemma by assembling a step-up transformer, which is constructed with a boost convertor, LM-2577, and a protection circuit module, XD-58A, which can increase the voltage to the scope of application. Furthermore, SSD1306-OLED is selected as the display module to be the monitor for our climate box. There are five layers for the module to set out distinct information. The first layer shows the time and date instantly. The second and third layers illustrate two kinds of datasets that are both grabbed from the OpenWeatherMap API, a website provided by the CWB open information. Primarily, the second layer, which presents the instantaneous rainfall prediction, indicates some messages, including the character to judge the weather today, T symboling the temperature, H representing the humidity, P designating as the pressure, and W stood for the wind speed. Then, the third layer denotes weather forecasting with three days in succession, which helps us to predict the climate during the upcoming three days. The fourth layer, which demonstrates the dataset supervised by our sensors straight away. Additionally, the data of the instantaneous rainfall prediction provided by the CWB open information is used to compare with the dataset supervised by our sensors. After gripping the data, there are two different ways for data storage. Ultimately, the fifth layer acting as the recording layer aims to determine which way is used to store the data and checked whether the SD card works correctly. Cooperating with the Internet accessibility, if the Internet is connected, the data sensed will be sent to the cloud, ThingSpeak, or the data will only be stored in the SD card module. Notably, when the reading of the FC-37, the rain sensor exceeds the setting threshold. A text message will automatically be sent to the users for the heavy rain advisory by using IFTTT, a freeware web-based service that creates chains of simple conditional statements [25], or a communication application.

(2) Data gathering from weather box

As completing to manufacture the weather box, the experimental fields are set in the National Taiwan University in Taipei, Taoyuan city, and pear orchard located at Yilan city. Since the BME280 sensor needs to be situated in the outside of the weather box to get a precise record. Therefore, all of the weather boxes for experimental fields are located at the places that sheltered from the storm to satisfy the needing and prolong the service life of the sensor at the same time, as shown in Fig. 4. Nevertheless the rain sensor, the FC-37, demands to be exposed under the rain to catch the information of the instant raindrop so that the FC-37 sensor are extended out of the shelter.



Fig. 4 The rainfall prediction weather box

The first experimental field is National Taiwan University. It is worth mentioning that the rain detector is leaned in a slanted position to let the water roll down through the edge to avoid having puddles on the sensor, as shown in Fig. 5. In addition, the second experimental field is set up in Taoyuan city to be a comparison to the first field to find out whether the

experimental location will impact our rainfall weather box or not shown in Fig. 6. The last field is located in the pear orchard in Yilan city. Apart from the other two fields, this field grabs another two specific weather parameters, inclusive of soil humidity and soil electrical conductivity. The method of supervised learning is used and corresponding class labels are given as either sunny or rainy to construct a new rainfall prediction model, which is distinguishing from the previous weather boxes.



Fig. 5 The rain detector was leaned in a slanted position (Taipei)

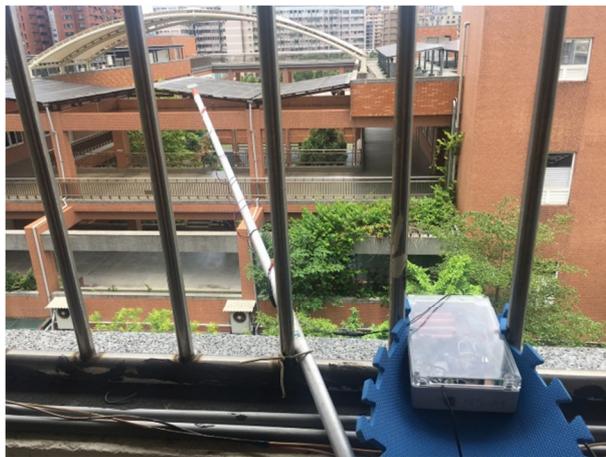


Fig. 6 The rain detector was leaned in a slanted position (Taoyuan)

(3) Data pre-processing

After accessing two datasets (the daily meteorological report provided by the CWB and the data accumulated by our rainfall weather box) the data must have been rectified by data cleaning for the following three different cases. Firstly, for the data collected by our weather box every ten minutes, the anomalous data caused by the problems of instrument and data transfer should be disposed of. The reason why those data can be abandoned is that it will not have a significant change of weather status in ten minutes (the time interval that we collected one piece of data). For the data provided by the CWB, there are two parameters that need to be processed. That is, if the symbol of “V” is presented in wind direction, the current value of the wind direction will be replaced by taking the average of the previous data and the next one. Likewise, if the symbol of precipitation is “T”, which implies the data is less than 0.1mm, the precipitation value will be altered by replacing it with 0.1mm. In order to design a rainfall-forecast model which could give out the alarm two hours earlier, the method of supervised learning is used. A rainfall label is constructed for the labeling schema, which is defined as follows:

$$label = \begin{cases} 1, & \text{if rain or two hours before rain} \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

The methods used for feature scaling are Min-Max Normalization and Standardization (Z-Score Normalization) in this paper. Although there are 17 weather parameters, some of them are lost. Thus, only eight of them are chosen to be the training features, including temperature, dew point, relative humidity, atmospheric pressure for weather station and sea level, wind

speed, wind direction, and solar radiation. After that, the 10-fold cross validation measurement is used to test and improve those constructed models. All of the machine learning model is developed in MATLAB.

3.2. Frosting prediction

The Raspberry Pi controls two sensors, including the BME280 and DS18B20. The BME280 detects temperature, humidity, and pressure in the air; the DS18B20 measures the temperature of the surface. In the hoarfrost point finding mode with the temperature controller and the thermoelectric cooler module the hoarfrost point is measured in any environment by gradually decreasing the surface temperature. In normal mode, the thermoelectric cooler module only have the function of heat dissipation by running fans. By the heat pipe in the thermoelectric cooler module, the surface temperature is close to the atmospheric temperature to simulate the leaf in the outdoor environment for a long time. The picture of the frosting prediction weather box is shown in the Fig. 7. The experimental field is set in the tea plantation in Fushoushan Farm in Taichung.

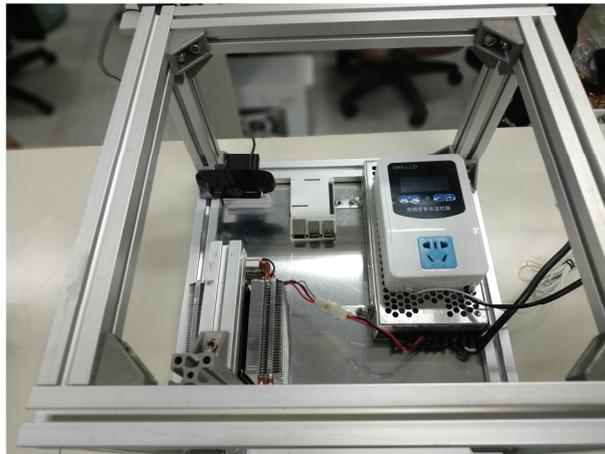


Fig. 7 The frosting prediction weather box

3.3. Lightning detection

For lightning detection, a lightning detecting weather box is constructed with Raspberry Pi acting as the microcontroller. The Raspberry Pi is connected to two sensors, AS3935 and BMP180. The weather box is illustrated in Fig. 8. Following ensuring the connection of the hardware, the first case is tuning the loop antenna. The loop antenna is designed to have a resonance frequency at 500kHz, which is the same as the lightning frequency of most lightning events of CG flashes, by adding or removing the tuning capacitors from 0pF to 120 pF in the step of 8pF to optimize the performance of the signal validation and distance estimation. The embedded algorithm will check the incoming signal pattern to reject the potential man-made disturbers [26].

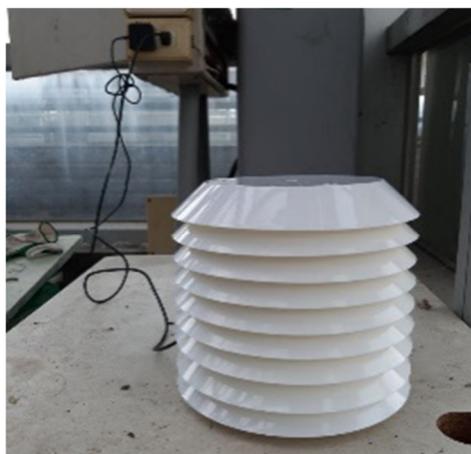


Fig. 8 The lightning detection system in a thermometer shelter in Yilan

In this paper, the sensor interacts with the Raspberry Pi via SPI protocol, and Python2 is chosen as the developing language with package “Spidev” (a Python module for interfacing with SPI devices from user space via the Spidev Linux kernel driver). There are four focused programmable settings in this paper. The first one is Analog-Front-End (AFE), which is a setting that can amplify and demodulate the AC-signal picked up by the antenna. The second one is mask disturber (MASK_DIST), which can shield the man-made disturber from the AS3935. The third one is spike rejection (SREJ), which is used to increase the robustness against false alarms from such disturbers with its range between 0 and 15. The last one is noise floor level threshold (NF_LEV). The output signal of AFE is also used to measure the noise floor level, which is continuously compared to the noise threshold. Whenever the noise floor level exceeds the noise threshold, AS3935 issues an interrupt, INT_NH, to inform the Raspberry Pi that it cannot operate properly due to the high input noise received by the antenna in this work. After the pre-work of the system is finished, whenever an event happened, the INT pin will go high, which implies that an event had happened. If the event is judged as a lightning event, the estimated distance of the lightning will be recorded in the register and stored in the Micro SD card, as the algorithm flow chart shown in Fig. 9. The collected data will be sent to the Intelligent Biosensing Platform within 6 hours, and the lightning detection system status can be monitored remotely with the weather boxes working automatically.

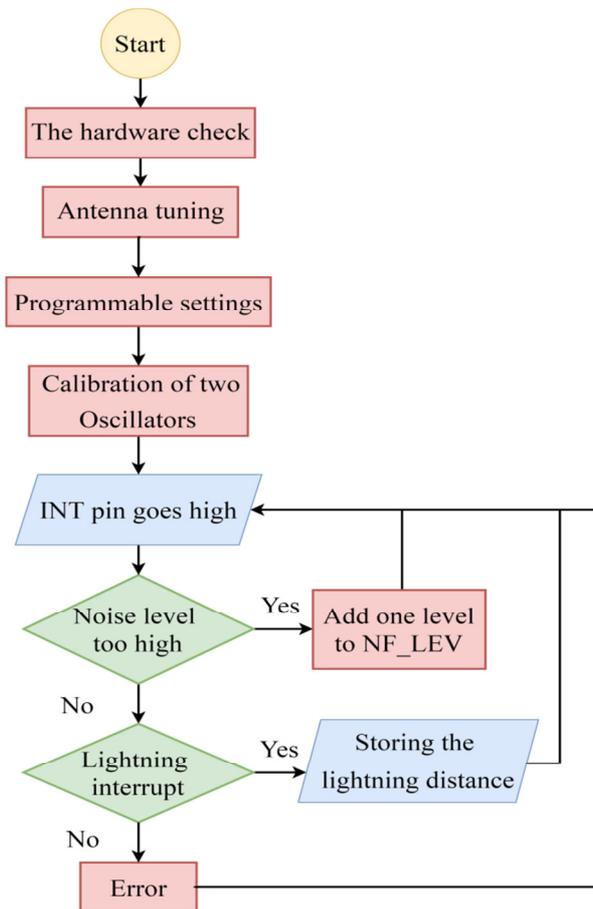


Fig. 9 The algorithm works flow chart of lightning detection system

4. Results and Discussion

4.1. Rainfall forecast

(1) Taipei, data collected by weather box

The instantaneous rainfall prediction model uses four parameters as inputs, including temperature, humidity, pressure, and solar radiation at that time to determine whether it is raining or not. As Table 1 shows, all of the models used in this

research reaches an accuracy of over 90%, and the model using KNN is the best. For constructing a model to predict raining before 2 hours, the extra models are built in two cases, inclusive of “before raining for 1-hr prediction model” and “before raining for 2-hr prediction model”. As the results portrayed in Table 2 and Table 3, both of the outstanding models for the two cases are the models utilizing KNN classifiers.

Table 1 The result of instantaneous rainfall prediction model (Taipei)

Classifier	Accuracy	Precision		Recall		F1 - Measure		MCC
		No Rain	Rain	No Rain	Rain	No Rain	Rain	
Decision Tree	94.3%	94.9%	93.9%	97.2%	89.0%	95.9%	91.3%	0.867
Bagged Tree	96.9%	97.0%	97.7%	99.0%	93.8%	97.9%	95.7%	0.928
KNN(K=3)	97.4%	98.3%	96.4%	98.7%	96.0%	98.5%	96.1%	0.941
SVM	90.8%	92.1%	88.3%	95.1%	83.1%	93.4%	85.6%	0.786

Table 2 The result of before raining for 1-hr prediction model (Taipei)

Classifier	Accuracy	Precision		Recall		F1 - Measure		MCC
		No Rain	Rain	No Rain	Rain	No Rain	Rain	
Decision Tree	94.0%	93.8%	93.7%	96.9%	87.7%	95.4%	90.9%	0.863
Bagged Tree	96.7%	96.7%	97.5%	98.4%	93.5%	97.4%	95.4%	0.925
KNN(K=3)	97.1%	98.1%	96.1%	98.5%	95.7%	98.3%	95.9%	0.934
SVM	90.4%	91.0%	88.0%	94.5%	82.3%	92.7%	84.9%	0.782

Table 3 The result of before raining for 2-hr prediction model (Taipei)

Classifier	Accuracy	Precision		Recall		F1 - Measure		MCC
		No Rain	Rain	No Rain	Rain	No Rain	Rain	
Decision Tree	91.7%	91.5%	92.2%	96.3%	84.1%	93.8%	87.8%	0.821
Bagged Tree	96.4%	96.1%	97.3%	98.2%	93.4%	97.1%	94.9%	0.922
KNN(K=3)	96.6%	97.8%	96.0%	98.3%	94.5%	98.0%	95.2%	0.925
SVM	88.6%	89.9%	88.1%	94.1%	80.5%	91.9%	83.8%	0.750

(2) Taoyuan, the data collected by weather box

Similar to the method used to predict rainfall in Taipei, there are also three models constructed in this part, inclusive of the “Instantaneous rainfall prediction model”, the “before raining for 1-hr prediction model”, and the “before raining for 2-hr prediction model”. As the results showed in Table 4, Table 5, and Table 6, the models using KNN as a classifier have an exceptional outcome. It is worth mentioning that the models using Bagged Tree as their classifier in Taoyuan are better than those in Taipei.

Table 4 The result of instantaneous rainfall prediction model (Taoyuan)

Classifier	Accuracy	Precision		Recall		F1 - Measure		MCC
		No Rain	Rain	No Rain	Rain	No Rain	Rain	
Decision Tree	94.3%	96.2%	86.1%	97.0%	82.5%	96.4%	83.9%	0.804
Bagged Tree	96.8%	97.2%	94.3%	99.0%	88.1%	97.9%	90.9%	0.891
KNN(K=3)	97.1%	97.4%	94.4%	99.1%	88.9%	98.1%	91.6%	0.899
SVM	89.9%	92.9%	73.5%	94.2%	70.0%	93.4%	71.7%	0.651

Table 5 The result of before raining for 1-hr prediction model (Taoyuan)

Classifier	Accuracy	Precision		Recall		F1 - Measure		MCC
		No Rain	Rain	No Rain	Rain	No Rain	Rain	
Decision Tree	93.2%	95.1%	86.0%	97.0%	80.1%	95.9%	82.8%	0.783
Bagged Tree	96.2%	97.0%	92.1%	99.0%	87.3%	97.9%	89.6%	0.877
KNN(K=3)	96.4%	97.2%	93.1%	98.1%	88.2%	97.4%	90.5%	0.884
SVM	88.8%	91.5%	72.4%	94.1%	68.1%	92.9%	71.0%	0.639

Table 6 The result of before raining for 2-hr prediction model (Taoyuan)

Classifier	Accuracy	Precision		Recall		F1 - Measure		MCC
		No Rain	Rain	No Rain	Rain	No Rain	Rain	
Decision Tree	93.2%	92.9%	85.0%	96.1%	76.1%	94.4%	80.2%	0.753
Bagged Tree	96.0%	96.8%	92.1%	98.2%	88.1%	97.4%	90.1%	0.881
KNN(K=3)	96.0%	96.9%	92.0%	98.2%	88.0%	97.5%	90.0%	0.882
SVM	87.7%	91.0%	74.6%	93.8%	65.6%	92.4%	70.0%	0.629

(3) Yilan, the data collected by weather box

Apart from the models used in the previous fields, models in this part are constructed differently for comparison to the previous ones. The models constructed in this part use six parameters as inputs, including temperature, humidity, soil humidity, soil electrical conductivity, wind speed, and solar radiation. As the results illustrated in Table 7, the performance in Yilan is not as good as those in other experimental fields. Besides, the model using SVM as a classifier have a recall of zero to predict rainfall.

Table 7 The result of instantaneous rainfall prediction model (Yilan)

Classifier	Accuracy	Precision		Recall		F1 - Measure		MCC
		No Rain	Rain	No Rain	Rain	No Rain	Rain	
Decision Tree	87.9%	93.1%	71.1%	91.8%	73.2%	92.4%	72.1%	0.646
Bagged Tree	91.3%	90.2%	86.3%	96.7%	72.9%	93.3%	79.0%	0.736
KNN(K=3)	87.3%	90.8%	72.0%	92.1%	69.3%	91.4%	70.6%	0.626
SVM	77.9%	78.0%	22.0%	100%	0%	87.6%	0%	0

(4) Taipei, the data provided by Central Weather Bureau

As shown in Tables 8 to 10, it can be found that models trained by the data collected by the weather boxes in Taipei have lower accuracies. The models using Bagged Tree as their classifier perform best in this part. Moreover, the accuracy and recall of predicting rainfall decrease for longer predictions.

Table 8 The result of instantaneous rainfall prediction model (CWB_Taipei)

Classifier	Accuracy	Precision		Recall		F1 - Measure		MCC
		No Rain	Rain	No Rain	Rain	No Rain	Rain	
Decision Tree	86.7%	90.8%	69.2%	94.1%	55.3%	92.4%	61.4%	0.537
Bagged Tree	87.8%	90.2%	74.4%	95.7%	56.3%	92.8%	64.0%	0.574
KNN(K=5)	84.8%	90.5%	61.1%	90.3%	60.4%	90.3%	60.0%	0.508
SVM	85.7%	88.2%	69.3%	95.4%	46.6%	91.6%	55.7%	0.481

Table 9 The result of before raining for 1-hr prediction model (CWB_Taipei)

Classifier	Accuracy	Precision		Recall		F1 - Measure		MCC
		No Rain	Rain	No Rain	Rain	No Rain	Rain	
Decision Tree	84.6%	90.7%	69.8%	94.1%	54.1%	91.9%	60.9%	0.523
Bagged Tree	86.3%	90.0%	74.3%	95.4%	56.4%	92.4%	64.1%	0.575
KNN(K=5)	83.3%	90.2%	61.4%	91.3%	58.2%	90.7%	60.0%	0.501
SVM	83.8%	88.3%	69.5%	95.2%	46.3%	91.6%	55.5%	0.487

Table 10 The result of before raining for 2-hr prediction model (CWB_Taipei)

Classifier	Accuracy	Precision		Recall		F1 - Measure		MCC
		No Rain	Rain	No Rain	Rain	No Rain	Rain	
Decision Tree	83.3%	86.1%	72.3%	93.5%	54.1%	89.3%	61.8%	0.523
Bagged Tree	85.7%	88.1%	78.3%	94.0%	60.4%	90.9%	68.1%	0.596
KNN(K=5)	82.4%	86.7%	63.4%	88.7%	60.2%	87.4%	61.7%	0.495
SVM	82.3%	84.8%	72.1%	94.6%	49.1%	89.2%	58.4%	0.487

Three types of models use four classifiers to find out the best classifier to fulfill the task of rainfall forecast. As the results shown in Table 1 to Table 3, for the three types of the models, it is obvious to recognize that the all of the models using the KNN as a classifier have the best performance, and those who used SVM have the worst. As for the models utilize the data provided by the CWB in Taipei used eight training features, inclusive of temperature, dew point, relative humidity, atmospheric pressure (weather station and sea level), wind speed, wind direction, and solar radiation. The training results of these models indicate that the models using Bagged Tree as their classifier have the top achievement, and those using SVM as their classifier look wretched as the results shown in Table 8 to Table 10. By the comparison of the models between the two different datasets, the prediction results suggest that the models which use the data collected by the weather box rather than the CWB lead to the increase of the predicting accuracies.

Additionally, it is clear to find that the accuracies decrease continuously from the “instantaneous rainfall prediction model” to the “before raining for 1-hr prediction model” until the “before raining for 2-hr prediction model”. It is thought to be because of the wrong prediction of the rainfall. As the high density of the orange dots that denote raining between the humidity of 70% and 80% in the humidity- pressure graph in Fig. 10, this situation implies that the prediction models might have had the wrong prediction in this zone caused by the high density. This might also be the reason why the rainfall prediction results at the Yilan experimental field are worse than those at the Taipei. Because the weather in Yilan is more prone to rain, and the detected humidity maintains at a high level which might cause the misjudgments.

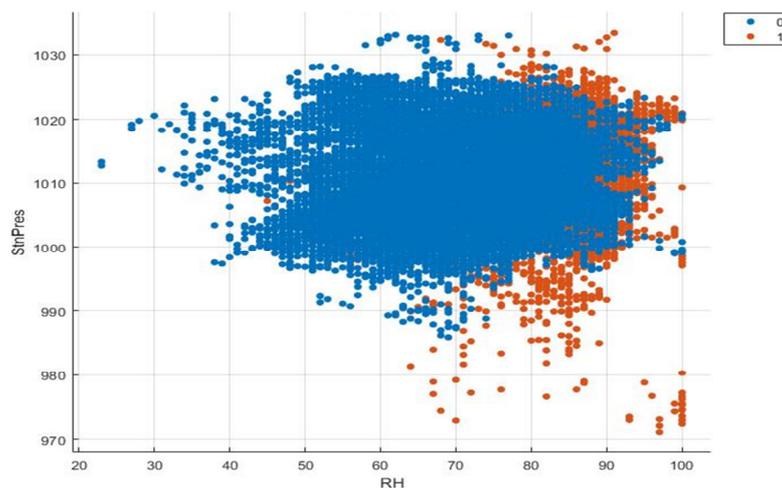


Fig. 10 The humidity- pressure graph provided by the CWB

In order to have a higher accuracy of the rainfall prediction before raining for two hours earlier, the construction of the data is modified to build more accurate prediction models. More data are being labeled as “raining” in the new and “before raining for 2-hr prediction model” in the following extra experiments by adjusting the ratio of the raining data to no-raining data to find out the best constitution of the training datasets for the four classifiers as presented in Table 11, though this action might have led to the ability to forecast no-rain decrease. For the following experimental results, these models of CWB and of our weather box are tested during other raining days. The experimental results are demonstrated in Table 12 and Table 13. It is suggested that because the shorter data collecting period (ten minutes) which is more frequent than an hour, the period time of CWB grabbing data, causes a higher accuracy as the reason in the previous part. Additionally, the models built by the data collected by the weather box are overfitted, which leads to the poor performance as presented in Table 12 and comparing to the results in Table 11.

For the results in the experimental field in Taoyuan, the best recall of the models built to forecast the rainfall in Taoyuan is 88%, which is lower than the best recall among all models in Taipei, 94.5%. This is caused by the different locations where the weather boxes are placed. The weather boxes in the Taoyuan experimental field are placed indoors. Even though we try to extend the weather boxes to the outdoors, they are still placed directly indoors. Nonetheless, the weather boxes in Taipei

experimental fields are placed in the outdoors, which receive the weather parameters directly. Another possible reason is the terrain, since the topography of Taoyuan city is tableland, dissimilar to basin, the terrain of Taipei. For the models in Yilan, though they use more weather parameters than those in Taipei, they perform poorly. For example, as the Table 7 showed, the model using SVM as classifier has no ability to predict rainfall. The reason is because of the lack of the label of the state of weather. To fix up this problem, the corresponding method is taken at the time using the relative humidity to substitute the label of the state of weather. The fixing approach is applicable because it is obvious that as soon as the FC-37 sensor of the weather box detects the raindrop, the reading of the humidity will increase instantly as response. The results reveal that this solution is not a proper way to take. Therefore, there are some recommended improving ways to implement these models. For example, adding some extra indicators such as rain gauge and raindrop sensor for the experimental modules might help predicting the rainfall. Otherwise, the collected data should be analyzed through the differential method to find the trend of the rain point.

Table 11 The result of before raining for 2-hr best-ratio prediction model (Weather box_ Taipei)

Classifier	Accuracy	Precision		Recall		F1 - Measure		MCC
		No Rain	Rain	No Rain	Rain	No Rain	Rain	
Decision Tree	93.7%	94.2%	92.2%	95.8%	90.1%	94.9%	91.1%	0.862
Bagged Tree	93.5%	94.1%	92.3%	96.1%	89.8%	95.0%	91.0%	0.858
KNN(K=3)	93.9%	92.4%	96.0%	98.5%	86.9%	95.3%	91.2%	0.868
SVM	92.0%	91.1%	93.8%	96.5%	83.5%	93.7%	88.3%	0.825

Table 12 Result of before raining for 2-hr best-ratio prediction model (1) (CWB_ Taipei)

Classifier	Ratio	Accuracy	Precision		Recall		F1 - Measure		MCC
			No Rain	Rain	No Rain	Rain	No Rain	Rain	
Decision Tree	75:25	83.7%	88.3%	64.5%	91.1%	57.7%	89.6%	60.9%	0.506
Bagged Tree	50:50	79.7%	92.7%	52.7%	80.3%	77.8%	86.0%	62.8%	0.513
KNN(K=5)	50:50	79.0%	91.7%	51.8%	80.4%	74.2%	85.7%	61.0%	0.487
SVM	50:50	81.1%	92.0%	55.3%	83.0%	74.4%	87.3%	63.4%	0.520

Table 13 The result of before raining for 2-hr best-ratio prediction model (2) (CWB_ Taipei)

Classifier	Ratio	Accuracy	Precision		Recall		F1 - Measure		MCC
			No Rain	Rain	No Rain	Rain	No Rain	Rain	
Decision Tree	75:25	72.6%	91.0%	47.1%	70.4%	79.2%	79.4%	59.1%	0.434
Bagged Tree	50:50	77.4%	92.2%	53.2%	76.4%	80.6%	83.6%	64.1%	0.508
KNN(K=5)	50:50	77.1%	87.1%	53.5%	81.5%	63.9%	84.2%	58.2%	0.429
SVM	50:50	69.8%	92.7%	44.5%	64.8%	84.7%	76.2%	58.3%	0.430

4.2. Frosting prediction

Via some tentative, there are three situations found to be the conditions of frosting, involving “as the temperature is lower than 5°C”, “the cloud cover is less”, and “the relative humidity is higher than 60%”. Under these situations, the weather box will send out a warning message to alarm the users to take some protection procedures, such as spraying water when the phenomenon of frosting is occurring or applying irrigation before the crops are frosted at nights soon.

4.3. Lightning detection

After constructing the lightning-detecting weather box, we start to collect the lightning data and optimize the related settings based on the feedback of the collected data during the plum rain season in Taiwan. Besides, we use the lightning data received by the Central Weather Bureau as the standard to evaluate the recall and precision of the data we receive. Moreover, five minutes is set as a time-interval. That is, if the Central Weather Bureau and the weather box we constructed both receive lightning events in the same time-interval, we consider that the data detected from both resources are the same flash event. Some parameters are defined as the followings, the units of them are the number of time-intervals.

True positive : if both resources received data (2)

False positive : if both weather box received data (3)

False negative : if only CWB received data (4)

$$\text{Recall} = \frac{\text{True positive}}{\text{True positive} + \text{False negative}} \quad (5)$$

$$\text{Precision} = \frac{\text{True positive}}{\text{True positive} + \text{False positive}} \quad (6)$$

The reason why we use time-interval to calculate the recall is because the AS3935 sensor will only record the most recent data among all the detected lightning events during one time-interval. The estimated distance does not represent the distance to a single lightning event but the estimated distance to the leading edge of the storm. For instance, if there exist three lightning events which estimated distances are 1-km, 3-km, and 5-km at the same time-interval, the AS3935 will only record the lightning event of the 1-km one during this time-interval.

As the results shown in Table 14, both of the recall and the precision outcome are not good enough, especially the performance of the recall. Besides, compared to the results in the field of Taipei in Table 14, the results, no matter the recall or the precision, in Yilan in Table 15 present better. The reasons of the poor performance of the lightning detection system are discussed in the following.

Table 14 The part of lightning events result in the field of Taipei

Data received during 2020/05/26					
Parameter	True positive	False positive	False negative	Recall	Precision
Counts	2	3	6	25.0%	40.0%
Data received during 2020/05/27					
Parameter	True positive	False positive	False negative	Recall	Precision
Counts	2	1	7	22.2%	66.7%
Data received during 2020/05/28					
Parameter	True positive	False positive	False negative	Recall	Precision
Counts	1	1	9	10.0%	50%

Table 15 Part of lightning events result in the field of Yilan

Data received during 2020/07/01					
Parameter	True positive	False positive	False negative	Recall	Precision
Counts	35	17	78	31.0%	67.3%
Data received during 2020/07/10					
Parameter	True positive	False positive	False negative	Recall	Precision
Counts	17	8	42	28.8%	68.0%
Data received during 2020/07/16					
Parameter	True positive	False positive	False negative	Recall	Precision
Counts	27	11	32	45.8%	71.1%

At the beginning of the experiment, it is found out that the constructed lightning detection system is vulnerable to be interference with external disturbance. The data collected in March fills with a distance of 1 km. There are some possible reasons to discuss. The first one is because of the noise generating by the Raspberry Pi itself. Another is that the NTU's internet base station formed an interference source that is 1-km away from the place where we collect the data. The fixing-up solution we adopt is extending the wires to make the microcontroller become farther away from the AS3935. This demand is due to the detection way of AS3935, radio detection, which is known to be sensitive to any other nearby interference [21].

So during April, we put the effort into finding the proper places with fewer disturbers to collect the lightning data, and the experimental field is finally set on the top floor of the department of Bio-Mechatronics Engineering of NTU for less man-made interference. The power supply used at that time IS solar power, and the weather box is placed under the solar panels to shield from the rainfall. However, after several times failing to receive data, we ultimately find out that the solar panel will block the reception of electromagnetic waves. The lightning detection system is finally placed in a thermometer shelter to avoid the rain from destroying the system. Besides, the system is connected to the power wire of Raspberry Pi to the socket as the power supply.

Additionally, the pieces of lightning events collected by the weather box are far less than those provided by the CWB during one storm in the experimental field of Taipei. It is supposed that the number of weather stations of the weather box in Taipei is far less than the number of weather stations of CWB in Taipei. Furthermore, it is easy to reveal that the recall and the precision of the system perform worse than expected. The first reason causing the poor recall is due to the narrow bandwidth of the AS3935 to collect the lightning events. It is known that the frequency bandwidth of CG flash is 3-3000kHz, and the frequency bandwidth of IC flash is 30,000-300,000KHz. However, according to the datasheet of AS3935, the AS3935 only receive lightning events with frequency in the range of (500 ± 33) kHz. There are large amounts of the lightning events that are not able to be received by the AS3935, which have caused the number of false negatives soared and poor performance of recall. For the precision, though the precision is better than the recall in the Taipei field, it still does not meet with the expected result.

The main reason to be discussed is the disturbance. Raspberry Pi 4 and Raspberry 3B+ are used as the microcontroller at the beginning. However, the experimental data points out that the noises generated by Raspberry Pi 4 will misjudge as the lightning events, which give rise to the number of the true positive and decreased the precision of the detection system. Moreover, as mentioned in the first paragraph of this section, though the chosen experimental field in Taipei is as possible as interference-free, the unexpected man-made disturbance can not be avoided completely, which increases the true positive and the decreased the precision. Additionally, it is worth mentioning that even if the datasheet of AS3935 claims that AS3935 is subjected to lightning events with a radius of 40 kilometers, according to the actual use, the range of lightning events that can be detected as completely as possible is 20 kilometers. Besides, compared to the results of Taipei experimental field, the results in Yilan perform noticeably. Two main differences are responsible for it. Firstly, the disturbance with a radius of 40 kilometers of the Yilan field are lesser than the Taipei field. Secondly, the lightning detection system in Yilan had an automatic scheduling function, and the system will retune the antenna every minute to ensure the accuracy of every moment. Additionally, this system can also be connected remotely to observe and monitor the working status of the entire system at any time.

Into the bargain, from the lightning observation document of the CWB, an experiment is done to count the lightning events collected by different lightning-detecting systems during a storm in [26]. The plotting-results reveal the number of lightning events collected by different systems. For instance, the lightning events collected by the CWB system are 76, which is larger than the number of events detected by the JMA system, 0, at 15:00. However, at 15:22, the lightning events collected by the JMA system are 1137, which is far larger than the number of events detected by the CWB system, which is 76. It is shown from the results that the number of the collected lightning events has a significant relationship with the bandwidth that the system can collect lightning and the mainly-focused detecting frequency, which is 500kHz for AS3935. Also, using the number of lightning events provided by the CWB as the benchmark may not have been a suitable method to examine the performance of the system.

5. Conclusion

A small intelligent weather station was developed, designed and established. It detected and collected rainfall, frosting, and lightning. The microprocessor could send out data via Internet or store data locally. Results showed that prediction model

based on the locally collected data was more accurate than using data from weather bureau. Frosting could be observed and recorded by an interval-shooting webcam. The lightning prediction was limited due to the frequency restriction of currently available sensors. For future work, we hope to increase the amount of lightning data by trying to use multiple weather boxes to collect data in the same area at the same time. Besides, we also aim to find other lightning sensors that detect different bandwidths of the lightning events to complete our lightning detection function.

Conflicts of Interest

The authors declare no conflict of interest.

References

- [1] M. A. Cooper and R. L. Holle, *Current Global Estimates of Lightning Fatalities and Injuries, Reducing Lightning Injuries Worldwide*, Springer, 2019.
- [2] R. L. Holle, "Annual Rates of Lightning Fatalities by Country," 20th International Lightning Detection Conference, April 2008.
- [3] R. S. Cerveny, P. Bessemoulin, C. C. Burt, M. A. Cooper, Z. Cunjie, A. Dewan, et al. "WMO Assessment of Weather and Climate Mortality Extremes: Lightning, Tropical Cyclones, Tornadoes, and Hail," *Weather, Climate, and Society*, vol. 9, pp. 487-497, 2017.
- [4] S. Pattar, R. Buyya, K. R. Venugopal, S. Iyengar, and L. Patnaik, "Searching for the IoT Resources: Fundamentals, Requirements, Comprehensive Review, and Future Directions," *IEEE Communications Surveys & Tutorials*, vol. 20, pp. 2101-2132, April 2018.
- [5] P. Sethi and S. R. Sarangi, "Internet of Things: Architectures, Protocols, and Applications," *Journal of Electrical and Computer Engineering*, vol. 2017, January 2017.
- [6] S. Aftab, M. Ahmad, N. Hameed, M. S. Bashir, I. Ali, and Z. Nawaz, "Rainfall Prediction Using Data Mining Techniques: A Systematic Literature Review," *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 5, pp. 143-150, 2018.
- [7] D. J. Hand and N. M. Adams, "Data Mining," *Wiley StatsRef: Statistics Reference Online*, pp. 1-7, 2014.
- [8] J. Han, J. Pei, and M. Kamber, *Data Mining: concepts and Techniques*, Elsevier, 2011.
- [9] A. Iqbal and S. Aftab, "A Feed-Forward and Pattern Recognition ANN Model for Network Intrusion Detection," *International Journal of Computer Network & Information Security*, vol. 11, no. 4, pp. 19-25, April 2019.
- [10] Z. Chao, F. Pu, Y. Yin, B. Han, and X. Chen, "Research on Real-Time Local Rainfall Prediction Based on MEMS Sensors," *Journal of Sensors*, vol. 2018, 2018.
- [11] S. Zainudin, D. S. Jasim, and A. A. Bakar, "Comparative Analysis of Data Mining Techniques for Malaysian Rainfall Prediction," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 6, pp. 1148-1153, 2016.
- [12] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.
- [13] R. C. Brito, F. Favaram, G. Calin, and E. Todt, "Development of a Low Cost Weather Station Using Free Hardware and Software," *Latin American Robotics Symposium (LARS) and 2017 Brazilian Symposium on Robotics (SBR)*, November 2017, pp. 1-6.
- [14] A. Geetha and G. Nasira, "Data Mining for Meteorological Applications: Decision Trees for Modeling Rainfall Prediction," *IEEE International Conference on Computational Intelligence and Computing Research*, 2014, pp. 1-4.
- [15] P. S. Yu, T. C. Yang, S. Y. Chen, C. M. Kuo, and H. W. Tseng, "Comparison of Random Forests and Support Vector Machine for Real-Time Radar-Derived Rainfall Forecasting," *Journal of Hydrology*, vol. 552, pp. 92-104, September 2017.
- [16] C. Cortes and V. Vapnik, "Support-Vector Networks," *Machine Learning*, vol. 20, no. 3, pp. 273-297, September 1995.
- [17] L. C. Liang and L. T. Chen, "Improved SVM Classifier Incorporating Adaptive Condensed Instances Based on Hybrid Continuous-Discrete Particle Swarm Optimization," *Advances in Technology Innovation*, vol. 1, no. 2, pp. 53-57, September 2016.
- [18] T. Cover and P. Hart, "Nearest Neighbor Pattern Classification," *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21-27, January 1967.
- [19] L. Breiman, "Bagging Predictors," *Machine Learning*, vol. 24, no. 2 pp. 123-140, 1996.

- [20] S. Rana and R. Garg, "Slow Learner Prediction Using Multi-Variate Naïve Bayes Classification Algorithm," *International Journal of Engineering and Technology Innovation*, vol. 7, no. 1, pp. 11, January 2017.
- [21] J. Lai, Y. Liu, J. Du, and Q. Li, "Lightning Detection Technology and Application," *International Conference on Meteorology Observations (ICMO)*, December 2019, pp. 1-5.
- [22] Z. Yang and S. Jiang, "Design of Lightning Detection System Based on ARM," *International Conference on Lightning Protection (ICLP)*, October 2014, pp. 346-350.
- [23] AS3935: Franklin Lightning Sensor IC, <https://www.mouser.tw/new/ams/ams-AS3935/>
- [24] S. Boughorbel, F. Jarray, and M. El-Anbari, "Optimal Classifier for Imbalanced Data Using Matthews Correlation Coefficient metric," *PloS One*, vol. 12, no. 6, pp. e0177678, June 2017.
- [25] Welcome to Collective.Ifittt! <https://collectiveifittt.readthedocs.io/en/latest/>
- [26] Y. Shibai, L. F. Tsai, and Y. Q. Lee, "Analysis of the characteristics of each lightning detection system in Taiwan," 2018, http://photino.cwb.gov.tw/conf/history/108/A3/A3_10_L_%E7%99%BD%E6%84%8F%E8%A9%A9_%E5%90%84%E9%96%83%E9%9B%BB%E5%81%B5%E6%B8%AC.pdf (Chinese)



Copyright© by the authors. Licensee TAETI, Taiwan. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-NC) license (<https://creativecommons.org/licenses/by-nc/4.0/>).