Global Transitions Proceedings 2 (2021) 294–298

Contents lists available at ScienceDirect

Global Transitions Proceedings

journal homepage: http://www.keaipublishing.com/en/journals/global-transitions-proceedings/

An efficient algorithm for predicting crop using historical data and pattern matching technique

Anjana, Aishwarya Kedlaya K*, Aysha Sana, B Apoorva Bhat, Sharath Kumar, Nagaraj Bhat

Dept. Computer Science and Engineering, Shri Madhwa Vadiraja Institute of Technology and Management (Affiliated to VTU), Udupi 574115, India

A R T I C L E I N F O

Keywords:
Crop Predictor
Dataset
Parameter preference
Pattern matching
Xarray

A B S T R A C T

As agriculture seems to be a crucial part in food security as well as economic development of a country, selecting crops for cultivation is a most important aspect in an agricultural planning. It relies on variety of parameters which includes weather condition, soil property and government policies. The suggested system helps the farmers to select suitable crop based on season and region of sowing. It will in-turn help the farmers by improving the net profit to them. By considering different datasets with respect to five parameters such as rainfall, temperature, slope, humidity and soil moisture of horticulture data, the system builds a model or method using which can suggest list of crops which is mainly helpful for the farmers in his/her decision making.

1. Introduction

In the economic development of a country like India, agriculture shows a vital role as it provides income and employment to the rural population and it acts as a main source of food. With the time, the need for the crop production is also increasing. In India, it devotes around 20% to the GDP of the country. Today the farmers cultivate the crops depending on the experience picked up from the past age. Because of the old techniques the farmers are not aware about the interest that happens in the current horticultural economy. This results in the misfortune to the farmers. Selection of crop is a most important aspect in agricultural planning. When the farmers know the accurate information on the best crop in their field as per season, it minimizes the loss. The rate of production of a crop relies on many factors [1] like weather specific parameters (e.g. rainfall, temperature, humidity etc.), soil parameters (e.g. soil moisture) and geography of a place (e.g. slope). Different datasets of these attributes are collected and then analysed. Collecting the data from right source plays an important role in building a prediction model as it effects on accuracy of the model.

The process of analysing data using various analytical and logical reasoning to evaluate each component of the data plays a very important role. This type of examining is just one of the many steps that must be performed when conducting a research analysis. Xarray can be used as one of the techniques to analyse and process the data. There are several existing models [2] for crop prediction about which farmers are unaware, may be due to its complexity or cost-effectiveness. Hence there is a need of developing such a model that is simple, user-friendly, and cost-effective and reach desired accuracy. All the existing methods [3] are only region (location) based but, in our algorithm, region is embedded with season so that the accuracy of the prediction can be improved. Here crop yield forecast models are prepared based on crop weather studies for estimating yield much before actual harvest of the crops.

2. Related works

About 70% portion of India's residents are dependent upon farming as its occupation. The paper [4] focuses on predicting and prognosticate the yield of the crop by learning the farming land's past data. Numerous factors such as rainfall, temperature, soil type and supplementary entities are contemplated by the process to develop a forecasting model by utilizing machine learning techniques. Various machine learning algorithms such as Polynomial Regression, Random Forest and Decision Tree is used. The system can be enhanced by combining this with other departments like sericulture, horticulture and many more who work towards the growth of agriculture of a country.

In paper [5] the dataset is built from former historic statistics which incorporates various influencing parameters like rainfall, humidity and temperature. Here Random Forest is used for prediction which is a well reputed machine learning algorithm. The chosen algorithm had capabilities to predict best crop by taking a smaller number of models. The proposed method assists the farmers while selecting which crops to grow in the field. This approach works for organized dataset. As a future enhancement same work can be extended to work with inputs of independent system as well.

Understanding spatiotemporal variability of soil moisture is important for amplifying the forecasting power of hydrological models, cli-

* Corresponding author.
E-mail addresses: aishwaryakedlayak@gmail.com (A.K. K), sharathkumar.cs@sode-edu.in (S. Kumar), nagarajbhat.cs@sode-edu.in (N. Bhat).

https://doi.org/10.1016/j.gltp.2021.08.060
Received 29 May 2021; Accepted 28 June 2021
Available online 12 August 2021

mate and land surface. In paper [6] soil moisture dynamics were inspected in response to irrigation and rainfall events by employing statistics received from non-discrete point quantification regulated below wheat and rice yield for an agricultural area situated within the Ganga Basin, India. Soil moisture control is a task of crop administration exercise and soil intensity in an agricultural atmosphere. The investigation spot was separated into 24 sub spots and an analysis was done by considering 18 plots by collecting continuous data for two seasons and the study was mainly done for rice and wheat crops. It was found that for rice crop cover soil moisture distribution was largely monitored by time fixed ingredients and absolute soil moisture and temporal abnormality showcased alike sample with time at all the depths. However, for wheat crop cover, it varied with depth tanks to regular sprinkling and dehydrating state and temporally changing atmospheric needs.

In paper [7] they developed an application with an easy interface to farmers of Andhra Pradesh for fetching stronger yield recommendation based on the investigation of historical agriculture statistics, matching for a season, type of soil, rainfall and place type. This work furnishes a good recommendation for the farmers to choose a crop. The shortcoming of this approach is, that it is third-party software it may not work as expected in some of handhold devices such as Realme, Redmi etc.

In the paper [8], the effect of heat stress and slippage of rainfall on agricultural fertility engaging a zestful panel data resemble is estimated and potency of crop diversification in diminishing their unfortunate effect is also assessed. Farmer's decisiveness to embrace productivity amplifying inputs are largely affected by climatic irregularity, which successively influence agriculture fertility. There are dual intelligible inferences for strategies aiming at building agriculture climate strong. One, the climatic irregularity or disturbance are location-specific; so there is a demand to build up location-specific preliminary caution mechanisms so that well timed details on weather are supplied to the farmers so that they are prepared to resolve on yield and other agronomic exercise in prediction of an irregularity.

The paper [9] aims to increase productivity and reduce the wrong selection of a crop. It proposed a recommendation system through an ensemble model using KNN, Random tree and Naive Bayes as learners to put forward a harvest for the site-distinct parameters. As a future enhancement increased number of attributes will improve the quality of the data set.

The paper [10] made use of data mining proficiency to take out the required details from agricultural statistics to calculate harvest yield for cereal crops in regions of Bangladesh. The methodology comprises of several clustering techniques using KNN, Linear regression, ANN. The dataset included five environmental parameters three biotic parameter and two related variables to ascertain the crop production in different regions. Here the efficiency of forecasting reclines in the range 90-95. The future work stated that adding geospatial analysis to the proposed data processing model improves the accuracy.

If there are numerous possibilities to grow a crop using restricted land resource, Choice of right harvest at the right time is a biggest challenge. The paper [11] articulated a mechanism named Crop Selection Method (CSM) which solves the above problem and put forward plenty of yield to be considered based on numerous parameter which includes water density, type of the soil, crop type and weather. The efficiency of the proposed method CSM purely depend upon how well and accurately influencing parameters are predicted. There is a necessity of identifying a prediction paragon with increased efficacy and accuracy.

3. Methodology

3.1. Dataset

The dataset [12] consisting of the soil specific attributes (e.g. soil moisture) and weather specific attributes (e.g. rainfall, temperature, humidity etc.) is collected from IMD Pune. Crop dataset is prepared manually which contains 20 major crop information. The datasets contain 0, NaN, floating point type of data values.

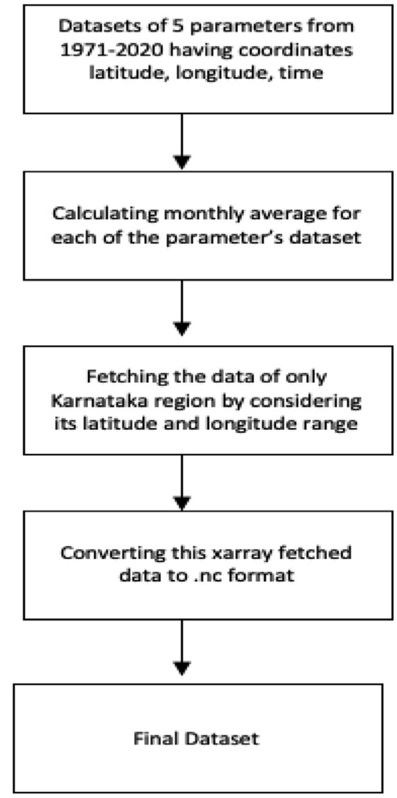

Fig. 1. Final dataset collection.

The Fig 1 shows the flow of final dataset collection. As initial step, different datasets are collected with respect to five parameters such as rainfall, temperature, slope, humidity and soil moisture for 50 years i.e., from 1971-2020 having coordinates latitude, longitude and time. Then monthly average is calculated for each of the parameter's dataset. In case of calculating monthly average, first step is to resample the time series data and calculate monthly average for each year. Then group this data by month and take the mean. That is average rainfall in each of the twelve months for past 50 years.

Next step is about fetching the required amount of data. Since focus is only on Karnataka State, the data of Karnataka region is fetched [13] by considering its latitude and longitude range i.e., latitude from 10 to 20 and longitude from 70 to 80. And then the fetched Xarray is converted to NetCDF file. This way the final dataset is prepared for all five parameters.

3.2. Pseudo code

The entire process is represented in the pseudo code shown below.

1 Loading parameter dataset having coordinates: Latitude, Longitude, Time
2 Calculating monthly average for each parameter's dataset
For example, for Rain data the average is calculated as follows,
```
Data_rain_monthly=Data_rain.resample(datetime='M').
mean(dim='datetime')
Data_rain_avg= Data_rain_monthly.groupby('datetime.month').mean()
```
3 Fetching the data of only Karnataka region
```
Data_rain_avg.precip.sel(lat=np.arange(10,20),lon=np.
arange(70,80),method='nearest')
```
4 Converting this Xarray fetched data to
NetCDF (.nc) format

Data_rain_avg.to_netcdf ("Rain.nc")

5 Similarly, weather dataset is prepared by considering 5 parameters namely: Rain, Temperature, Soil moisture, Humidity, Slope
6 Crop dataset is prepared manually by considering range of all 5 parameters as in weather dataset
7 Once the user enters month and region in the user interface, weather dataset is loaded and then latitude and longitude of the region is fetched using Geolocator.
8 Obtained latitude and longitude along with month that is entered by user is used to fetch the various parameter details from weather dataset and those are stored in corresponding variables.
9 Two result lists are created namely: res1 and res2, for storing crop names that satisfy preference conditions
10 For each crop in crop dataset "crop.csv" do

```
If preference 5 is satisfied then
  If preference 4 is satisfied then
        Append crop name to list "res1"
            Else if preference 3 is satisfied then
                Append crop name to list "res2"
      End if
Else
Reject the crop
            End if
      End for
```

Finally, displaying predicted crop list along with its details to the user.

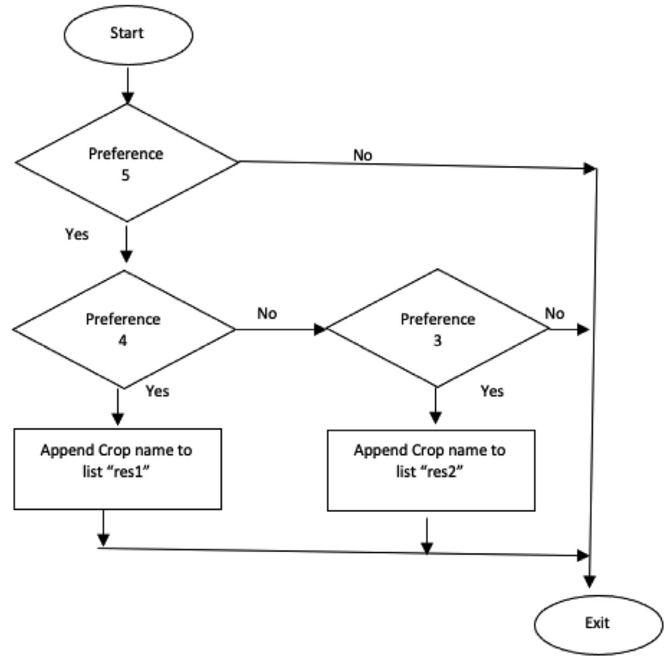

Fig. 2. Preference requirement.

3.3. Implementation

Since the data for only Karnataka region is considered, the user can enter one of the districts of Karnataka in the frontend of the application. From the input given by the user, the latitude and longitude of that district is fetched. Since there may not be the value for exact same latitude and longitude obtained by geolocator in the parameters' dataset, the nearest parameter value is fetched by using method called "nearest". This way the value is fetched of all five parameters for a given input district and month.

On the other hand, two lists namely "res1" and "res2" are taken. The crop dataset which is prepared manually is processed and result is stored for later comparison.

In proposed model pattern matching technique is used [14]. For each crop one pattern is given i.e., pattern of preference of parameters. The sequence of parameters taken is rainfall, temperature, humidity, soil moisture and slope. So, if the pattern is 5-4-1-3-0, then for that crop rain requirement must satisfy since it has highest preference 5. Then temperature requirement should satisfy, since the second highest preference is 4. If these two parameters satisfy, then that crop will be stored in "res1" list. If not, parameter having preference 3 is checked. In this case, it is soil moisture. If preference 3 satisfies then that crop will be stored in "res2" list. Otherwise, that crop will be discarded, and same thing repeats for the remaining crops in the crop dataset.

As shown in the Fig 2 the crop must satisfy the parameter requirement having highest preference i.e., 5. Then it should satisfy the requirement of the parameter with preference 4 or 3.

Once all the crops are analysed, if "res1" list is not empty, then the crops in that list will be printed. Otherwise, "res2" list will be printed. If both are empty, then an alert message stating no crops will popup. Hence the logic part is mainly based on preferences given to the parameters for a particular crop.

3.4. System design

The Fig 3 represents the flowchart of system design. First the user must enter the location and month in which he/she wants to grow the crop, as input. The system will then fetch the latitude and longitude for that location. Once the latitude and longitude are obtained, then the weather- related data [15] is fetched for that location at given input season. This fetched weather related data is compared with the crop dataset to check which crop is suitable for that location at that season.

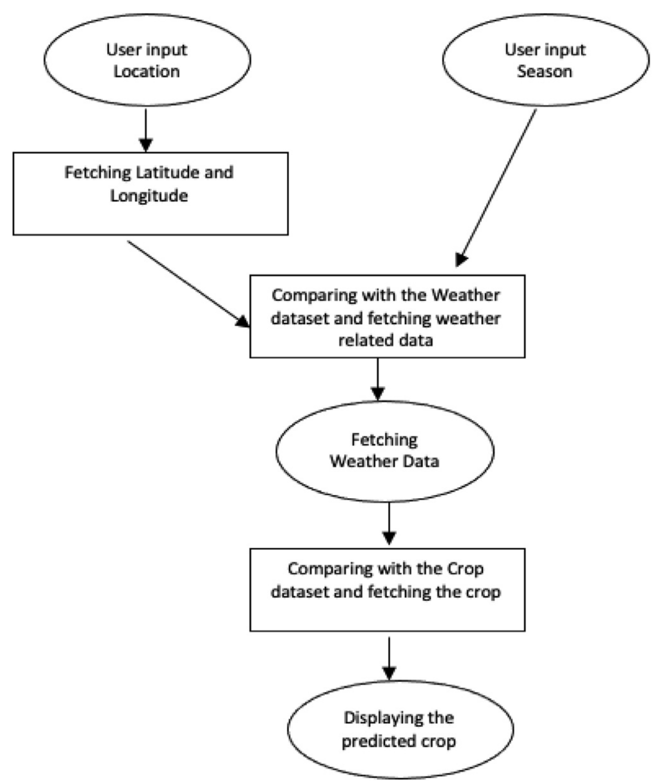

Fig. 3. System design.

4. Result and discussion

A responsive web-platform is developed for this proposed work which has user-friendly interface that can be accessed by the user by

Fig. 4. User interface for crop prediction.

Fig. 5. Output from crop-predictor.

Table 1
Accuracy analysis for 2 districts.

Year City		2016	2017	2018	2019	2020
UDUPI	PR	98.0%	96.5%	97.5%	97.0%	97.65%
	FPR	2.0%	3.5%	2.5%	3.0%	2.35%
DAKSHINA	PR	97.0%	98.5%	96.85%	98.0%	98.5%
KANNADA	FPR	3.0%	1.5%	3.15%	2.0%	1.5%

either laptop or mobile. There is no complexity of registration or login to use this platform.

The Fig 4 shows the user interface for crop prediction. The user can enter one of the districts of Karnataka and month in which he/she is willing to cultivate a crop. Once the weather-related data of the given input region & season is fetched and compared with the crop dataset, a list of crops suitable for given input condition will be displayed.

The Fig 5 represents the output from Crop-Predictor. The user can get more information on a particular crop by clicking on the name of that crop, which includes soil specific requirements, rainfall, temperature and other weather specific requirements for the cultivation of that crop. The user interface also facilitates the users by guiding them how to use the platform. The prediction accuracy of the proposed system is 98%. Hence, higher the accuracy results in more profit to the user [16-18].

As there is no relevant dataset available, 2 districts of Karnataka (Udupi and Dakshina Kannada) and 5 years of actual data from the horticulture department to predict the accuracy of the proposed algorithm are considered [19-21].

The Table 1 shows the accuracy analysis of 2 districts in which PR represents Prediction Rate and FPR represents False Prediction Rate. The accuracy of the prediction is calculated using the following formulas.

$$PR = \frac{TP}{No\ of\ cases\ compared} \tag{1}$$

$$FPR = \frac{FP}{No\ of\ cases\ compared} \tag{2}$$

Where, TP (Truly Predicted): The Prediction that was observed in the algorithm, was matching with that of horticulture department data.

FP (Falsely Predicted): Wrong decision taken by the algorithm in which results are not matching exactly, either some crop was missing, or some crop was printing extra.

No of cases compared is the count of total seasonal results available with horticulture department data [22-24].

The Table 2 shows the comparative analysis of the prediction accuracy with that of existing systems. It is very evident from the table that

Table 2
Accuracy comparison with existing methods

Reference	Algorithm/ Technique Used	Accuracy (%)
P. Patil et al. [18]	Decision Tree Classifier	76.8
S. Pudumalar et al. [9]	Naïve Bayes, K-Nearest Neighbor, CHAID and Random Tree	88
E. Manjula and S. Djodiltachoumy [21]	Association Rule Mining	88
T.M.S. Ahamed et al. [10]	Linear Regression, K-Nearest Neighbor and Artificial Neural Network	90-95
Proposed Method	Pattern Matching	98

accuracy of the proposed method is more when compared with existing methods.

5. Conclusion

This work presents an effective crop prediction system by using past data. Using Xarray functions datasets are analysed and pattern matching technique is used to fetch the crop based on region and season. This proposed system helps the farmers to select suitable crop based on season and region of sowing. It will in- turn help the farmers by reducing the loss faced by them and improve net crop yield. Since this system is implemented only for Karnataka, the future work aimed at considering larger region, i.e., implementation in whole country. Also including the regional languages in the user interface, would help the farmers for better understanding.

References

[1] K.O.G. Dias, H.P. Piepho, L.J.M. Guimarães, et al., Novel strategies for genomic prediction of untested single-cross maize hybrids using unbalanced historical data, Theor. Appl. Genet. 133 (2020) 443–455, doi:10.1007/s00122-019-03475-1.

[2] A. Dahane, R. Benameur, B. Kechar, A. Benyamina, An IoT based smart farming system using machine learning, Int. Symposium Networks, Comput. Commun. (ISNCC) (2020) 1–6, doi:10.1109/ISNCC49221.2020.9297341.

[3] D. Sawant, A. Jaiswal, J. Singh, P. Shah, AgriBot - An intelligent interactive interface to assist farmers in agricultural activities, in: 2019 IEEE Bombay Section Signature Conference (IBSSC), 2019, pp. 1–6, doi:10.1109/IBSSC47189.2019.8973066.

[4] Shruthi Gowda, Sangeetha Reddy, Design and implementation of crop yield prediction model in agriculture, Int. J. Sci. Technol. Res. 8 (2020) 544.

[5] Y.J.N. Kumar, V Spandana, V.S. Vaishnavi, K. Neha, V.G.R.R. Devi, Supervised machine learning approach for crop yield prediction in agriculture sector, in: 5th International Conference on Communication and Electronics Systems (ICCES), 2020, pp. 736–741, doi:10.1109/ICCES48766.2020.9137868.

[6] Ephrem Yetbarek, Richa Ojha, Spatio-temporal variability of soil moisture in a cropped agricultural plot within the Ganga Basin", India, Agric. Water Manage. 234 (2020).

[7] N. Usha Rani, G Gowthami, Smart crop suggester", advances in computational and bio- engineering, Learning and Analytics in Intelligent Systems, vol 15, Springer, Cham, 2020, doi:10.1007/978-3-030-46939-9_34.

[8] Pratap S. Birthal, Jaweriah Hazrana, Crop diversification and resilience of agriculture to climatic shocks: evidence from India, Agric. Syst. 173 (2019).

[9] S. Pudumalar, E. Ramanujam, R.H. Rajashree, C. Kavya, T. Kiruthika, J. Nisha, Crop recommendation system for precision agriculture, in: Eighth International Conference on Advanced Computing (ICoAC, 2017, pp. 32–36, doi:10.1109/ICoAC.2017.7951740.

[10] A.T.M.S. Ahamed, et al., Applying data mining techniques topredict annual yield of major crops and recommend planting different crops in different districts in Bangladesh, in: IEEE/ACIS 16th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), 2015, pp. 1–6, doi:10.1109/SNPD.2015.7176185.

[11] R. Kumar, M.P. Singh, P. Kumar, J.P. Singh, Crop Selection Method to maximize crop yield rate using machine learning technique, in: International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM), 2015, pp. 138–145, doi:10.1109/ICSTM.2015.7225403.

[12] P.C. Reddy, A.S. Babu, Survey on weather prediction using big data analystics, in: Second Int. Conf. Electr., Comp. Commun. Technol. (ICECCT), 2017, pp. 1–6, doi:10.1109/ICECCT.2017.8117883.

[13] R.V. Kishor, et al., Agromet expert system for cotton and soyabean crops in regional area, in: 2018 International Conference On Advances in Communication and Computing Technology (ICACCT), 2018, pp. 149–154, doi:10.1109/ICACCT.2018.8529664.

[14] G. Ghazaryan, S. Skakun, S. König, E.E. Rezaei, S. Siebert, O. Dubovyk, Crop yield estimation using multi-source satellite image series and deep learning, in: IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium, 2020, pp. 5163–5166, doi:10.1109/IGARSS39084.2020.9324027.

[15] Meeradevi, H. Salpekar, Design and implementation of mobile application for crop yield prediction using machine learning, in: 2019 Global Conference for Advancement in Technology (GCAT), 2019, pp. 1–6, doi:10.1109/GCAT47503.2019.8978315.

[16] P. Karuppusamy, I. Perikos, F. Shi, T.N. Nguyen, Sustainable communication networks and application, Lecture Notes Data Eng. Commun. Technol. (2020) 65–72.

[17] N.T. Le, J.W. Wang, C.C. Wang, T.N. Nguyen, Automatic defect inspection for coated eyeglass based on symmetrized energy analysis of color channels, Symmetry 11 (12) (2019) 1518 p..

[18] P. Patil, V. Panpatil, S. Kokate, Crop prediction system using machine learning algorithms, Int. Res. J. Eng. Technol. 7 (Feb. 2020).

[19] T.N. Nguyen, B.H. Liu, S.Y. Wang, On new approaches of maximum weighted target coverage and sensor connectivity: Hardness and approximation, IEEE Trans. Network Sci. Eng. 7 (3) (2019) 1736–1751.

[20] S. Rajendrakumar, V.K. Parvati, Automation of irrigation system through embedded computing technology, in: Proceedings of the 3rd International Conference on Cryptography, Security and Privacy, 2019, pp. 289–293.

[21] E. Manjula, S. Djodiltachoumy, A model for precision of crop yield, Int. J. Comput. Intell. Inf. 6 (4) (2017) 298–305.

[22] B.D. Parameshachari, R.P. Kiran, P. Rashmi, M.C. Supriya, Rajashekarappa, H.T. Panduranga, Controlled partial image encryption based on LSIC and chaotic map, ICCSP (2019) 60–63.

[23] B.D. Parameshachari, H.T. Panduranga, Secure transfer of images using pixel-level and bit-level permutation based on knight tour path scan pattern and henon map, in: Cognitive Informatics and Soft Computing, Springer, Singapore, 2021, pp. 271–283.

[24] S. Chen, L. Zhang, Y. Tang, C. Shen, R. Kumar, K. Yu, U. Tariq, A.K. Bashir, Indoor temperature monitoring using wireless sensor networks: a SMAC application in smart cities, Sustain. Cities Soc. 61 (July 2020) 102333.