Azerbaijan Journal of High Performance Computing, Vol 5, Issue 1, 2022, pp. 94-111
https://doi.org/10.32010/26166127.2022.5.1.94.111

Provide a Method Improving Temperature Control in Smart Buildings Based in Slicing Technique and Clustering IoT Network Based on Composition

Armin Rabieifard[1], Zahra Tayyebi Qasabeh[2], Marzieh Faridi Masulehida[3] and Maryam Abedini[4]

[1] Non-Profit Higher Education Institutions Sardarjangal Branch, Guilan, Iran, Armin.rabieifard@gmail.com, lidanaderlou@gmail.com
[2]Payame Noor university of Guilan, Guilan, Iran, tayyebi.shiva@gmail.com
[3] Department of Computer, Non-profit higher education institutions Sardarjangal Branch, Guilan, Iran, M.faridi@ahrar.ac.ir
[4]PhD, Guest lecturer at Guilan University,Guilan, Iran, Maryam.abedini@gmail.com.

*Correspondence: Zahra Tayyebi Qasabeh, Payame Noor university of Guilan, Iran. tayyebi.shiva@gmail.com

Abstract

Today, energy consumption is important in calculating the heating and cooling loads of residential, industrial, and other units. In order to calculate, design, and select the heating-cooling system, a suitable method of consumption and cost analysis is needed to prepare the required data for air conditioning motors and design an intelligent system. In this research, a method for balancing the temperature of an intelligent building in the context of the Internet of Things is presented based on a combination of network cutting and clustering techniques. In order to achieve the optimization of the algorithm in this method, it is necessary to convert heterogeneous data into homogeneous data, which was done by introducing a complex network and appropriate clustering techniques. In this method, information was collected by the IoT, and a graph matrix of these data was generated, then recorded by an artificial intelligence method and a combination of three methods of hierarchical clustering, Gaussian mixture, and K-means for comparison with the preliminary results. Finally, due to the reliability of the K-means method and the use of majority voting for weights, the K-means method reached 0.4 and was selected as the clustering method. The main part of the proposed method is based on different classifications in Appropriate criteria that were evaluated. Acceptable results were recorded so that with the minimum value of 88% and the highest value of about 100, the results of the proposed method can be confirmed. All hypotheses of the method can be declared possible and acceptable.

Keyword: Smart Building Temperature Equilibrium, IoT, Graph, Grid Cutting Techniques, K-means Gaussian Mixed Hierarchy Clustering.

1. Introduction

Today, managers are faced with a phenomenon called urbanization. Today's cities, where almost half of the world's population lives, are considered complex networks and systems composed of components such as citizens, industries and businesses, transportation, communications, energy infrastructure, Urban services, and other urban subsystems been formed (Albino, V., Berardi, U., & Dangelico, R. M., 2015). The other definition identifies key domains of Smart Cities, such as smart economy, smart mobility, smart environment, smart living, smart people, and smart governance As well Smart Cities can be understood as places generating a particular form of spatial intelligence and innovation based on sensors, embedded devices, large data sets, and real-time information and response (Nowicka, K., 2014). The data produced in a smart city using IoT has several characteristics. The three characteristics of volume, variety, and speed are expected for most data. Data sets are always produced at high speed and are available in a heterogeneous form. The data produced in a smart home of a smart city can be from various services, such as electricity, gas, water, protection, and security (Sharma, V., & Tiwari, R., 2016). In addition to maintaining heterogeneous data, the main challenge is processing such a huge volume of heterogeneous data in a limited time, which is very difficult, and data heterogeneity is still an unsolved challenge (Gohar, M., Ahmed, S. H., et al., 2018).

The main goal of the research is to smooth the temperature of the smart building in the context of the Internet of Things, and since the temperature data collected from the network The Internet of Things of smart buildings are usually heterogeneous data; as a result, data management and converting heterogeneous data into homogeneous is one of the key parts of this research. Since data management programs in cases such as water management, pollution control, and energy consumption do not guarantee a better solution in the entire intelligent environment, one of the potential solutions is the development of an Internet of Things system integrated with new technology, or It is appropriate that this can help to define a proper architecture and reach the definite goal. Therefore, in this research, a new technique for processing heterogeneous data and converting it into homogeneous data is proposed in a smart building. The proposed method combines slicing and graph technologies and clustering techniques based on the Internet of Things. The algorithm is focused on thermal Internet of Things devices, and in smart buildings, Internet of Things equipment is spread throughout the building to manage and monitor it. IoT devices are distributed throughout the building to monitor it in smart buildings.

Since the thermal packaging devices depend on the physical topology of the building and the non-homogeneity of the building temperature, it is therefore assumed that the data collected by the packaging devices will depend on the topology of the building (offices that are closed and with heating, corridors, large common areas, etc.) and Non-uniformity in temperature (i.e., temperature varies between different areas because users can select comfort temperature, doors open or closed, etc.).

They break the network by creating multiple logical networks on a common physical infrastructure. It creates a flexible ecosystem of logical networks within a programmable, isolated, and software-based networking environment, where technical and business innovations are localized and localized in physical networks. Network slicing enables operators to customize and create their networks based on performance, efficiency, and isolation from other networks. In this research, the proposed technique is based on the IoT slicing method because this technique combines complex networks to reduce algorithm input errors and improve the monitoring and control of a smart building. A graph is a mathematical representation of a network and shows the relationship between vertices and edges; in this research, the nodes are the thermal IoT devices, and the edges (edges) of the graph are doors or corridors connected to the IoT nodes 9 Casado-Vara, R., Martin-del Rey, A., Affes, S., Prieto, J., & Corchado, J. M., 2020).

Also, clustering is finding similar clusters of data among input samples. Since the temperature data collected from the Internet of Things smart building network is usually heterogeneous, the main purpose of applying this clustering technique is to find the packaging tools with similar data. Then the IoT slicing method converts the heterogeneous data into a group of similar data based on the output of the clustering technique (measurement of the best clustering method between Gaussian mixed model methods, K-means, and hierarchical by Voting technique). In the proposed method, the neighborhood temperature of sensors is selected by IoT nodes to detect errors and increase data quality. IoT packaging devices collect temperature data in a smart building, but there is no way to know if this data can be trusted to monitor and control the temperature of a smart building. As a result, this method will be used to increase the reliability of the collected temperature data for smart building monitoring and control.

2. Related Work

In 2022, Wickramasinghe et al. a paper entitled "Temperature clusters in commercial buildings using k-means and time series clustering." They stated that An efficient building should be able to control its internal temperature in a manner that considers both the building's energy efficiency and the comfort level of its occupants. Thermostats help to control the temperature within a building by providing real-time data on the temperature inside that space to determine whether it is within the acceptable range of that building's control system, and proper thermostat placement helps to better control a building's temperature. More thermostats can provide better building control and a better understanding of the building's temperature distribution. In order to determine the minimum number of thermostats required to accurately measure and control the internal temperature distribution of a building, it is necessary to find locations that show similar environmental conditions. In this paper, we analyzed high-resolution temperature measurements from a commercial building using wireless sensors to assess the performance and health of the building's HVAC zoning and controls system. Then we conducted two cluster analyses to evaluate the existing zoning structure's efficiency and find the optimal number of clusters. K-means and time series clustering was used to identify the temperature clusters per building floor. Based on statistical assessments, we observed that time series clustering showed better results than k-means clustering (Wickramasinghe, A., Muthukumarana, S., Loewen, D., & Schaubroeck, M., 2022).

In 2022, Talei et al. a paper entitled "Smart Building Energy Inefficiencies Detection through Time Series Analysis and Unsupervised Machine Learning. " They stated that The climate of Houston, classified as a humid subtropical climate with tropical influences, makes the heating, ventilation, and air conditioning (HVAC) systems the largest electricity consumers in buildings. HVAC systems in commercial buildings are usually operated by a centralized control system and/or an energy management system based on a fixed schedule and scheduled control of a zone setpoint, which is inappropriate for many buildings with changing occupancy rates. Lately, as part of energy efficiency analysis, attention has focused on collecting and analyzing smart meters and building-related data and applying supervised learning techniques to propose new strategies to operate HVAC systems and reduce energy consumption. On the other hand, unsupervised learning techniques have been used to study different buildings' consumption information and profile characterization after performing cluster analysis. This paper adopts a different approach by revealing the power of unsupervised learning to cluster data and unveil hidden patterns. In this study, we also identify energy inefficiencies after exploring the cluster results of a single building's HVAC consumption and building usage data as part of the energy efficiency analysis. Time series analysis and the K-means clustering algorithm are successfully applied to identify new energy-saving opportunities in a highly efficient office building in Houston (TX, USA). The paper uses 1-year data from a highly efficient Leadership in Energy and Environment Design (LEED)-, Energy Star-, and Net Zero-certified building, showing a potential energy savings of 6% using the K-means algorithm. The results show that clustering is instrumental in helping building managers identify potential additional energy savings (Talei, H., Benhaddou, D., Gamarra, C., Benbrahim, H., & Essaaidi, M., 2021).

In 2021, Floris et al. a paper entitled "An IoT-Based Smart Building Solution for Indoor Environment Management and Occupants Prediction." They stated that Smart buildings use the Internet of Things (IoT) sensors for monitoring indoor environmental parameters, such as temperature, humidity, luminosity, and air quality. Due to the huge amount of data generated by these sensors, data analytics, and machine learning techniques are needed to extract useful and interesting insights, which provide the input for the building optimization in terms of energy-saving, occupants' health and comfort. In this paper, we propose an IoT-based smart building (SB) solution for indoor environment management, which aims to provide the following main functionalities: monitoring of the room environmental parameters; detection of the number of occupants in the room; a cloud platform where virtual entities collect the data acquired by the sensors and virtual super entities perform data analysis tasks using machine learning algorithms; a control dashboard for the management and control of the building. With our prototype, we collected data for ten days. We built two prediction models: a classification model that predicts the number of occupants based on the monitored environmental parameters (average accuracy of 99.5%) and a regression model that predicts the total volatile organic compound (TVOC) values based on the environmental parameters and the number of occupants (Pearson correlation coefficient of 0.939) (Floris, A., Porcu, S., Girau, R., & Atzori, L., 2021).

In 2020, Lymperopoulos et al. a paper entitled "Building Temperature Regulation in a Multi-Zone HVAC System using Distributed Adaptive Control." They stated that In recent years there have been considerable research efforts on improving the energy efficiency of buildings. Since Heating, Ventilation, and Air-Conditioning (HVAC) systems are a big part of energy consumption, developing efficient HVAC control systems is crucial. In most progressive approaches, precise knowledge of system parameters and/or adequate historical data is required. However, these approaches may not perform as well in the presence of dynamic parameter changes due to human activity, material degradation, wear and tear, disturbances, and other operational uncertainties due to occupancy, solar gains, electrical equipment, and weather conditions. In this paper, we consider buildings with several climate zones and propose a distributed adaptive control scheme for a multi-zone HVAC system that can effectively regulate zone temperature by applying online learning and assuming the exchange of information between neighboring zones. The controller of each zone achieves the local objective of controlling zone temperature by compensating for the effects of neighboring zones and possible changes in the system's parameters. Despite the exchange of information, each local controller does not know how a neighboring zone's control actions and temperature affect its zone's temperature. For this reason, each local controller estimates the parameters of the interconnections in real-time and uses them together with the exchanged information to provide a more accurate local zone temperature control. The proposed method is illustrated using an example of temperature control in a six-zone building and a large school building, which is implemented in a Building Controls Virtual Test Bed (BCVTB) environment using EnergyPlus and MATLAB/Simulink (Lymperopoulos, G., & Ioannou, P., 2020)

In 2020, Casado-Vara et al. a paper entitled "IoT network slicing on virtual layers of homogeneous data for improved algorithm operation in smart buildings." They stated that With its strong coverage, low energy consumption, low cost, and great connectivity, the Internet of Things technology had become the key to smart cities. However, faced with many terminals, the rational allocation of limited resources, and the topology and non-uniformity of smart buildings, the fusion of heterogeneous data has become an important trend in the Internet of Things research. As a result, this paper proposes a novel technique for processing heterogeneous temperature data collected by an IoT network in a smart building and transforming them into homogeneous data that can be used as an input for monitoring and controlling algorithms in smart buildings, optimizing their performance. The proposed technique, IoT slicing, combines complex networks and clusters to reduce algorithm input errors and improve the monitoring and control of a smart building. To validate the algorithm's efficiency, it is proposed as a case study using the IoT slicing technique to improve the operation of an algorithm to self-correct outliers in data collected by IoT networks. The results of the case study confirm, irrefutably, the effectiveness of the proposed method (Casado-Vara, R., Martin-del Rey, A., Affes, S., Prieto, J., & Corchado, J. M., 2020).

In 2019, Casado-Vara et al. a paper entitled "Improving Temperature Control in Smart Buildings Based in IoT Network Slicing Technique." They stated that In smart buildings, many different types of IoT devices collect measurements of the environment. These sensors can vary in their characteristics and can also influence the topology of the smart building. For this reason, IoT devices collect heterogeneous measurements. Using complex network and clustering techniques, we have designed a new technique to transform heterogeneous data into homogeneous data; this technique is called IoT slicing. This technique consists of creating a graph with the measurements of the IoT network and virtualizing layers based on the clustering of the graph. To validate this new technique's efficiency, we present a case study's results using a smart building temperature control algorithm (Casado-Vara, R., De la Prieta, F., Prieto, J., & Corchado, J. M., 2019, December)

In 2018, Paul et al. a paper entitled "IoT and Machine Learning Based Prediction of Smart Building Indoor Temperature." They stated that The demand for helpful energy had increased astronomically over the past few decades, especially in the building sector, due to rapid development and enhanced lifestyle. The energy performance of the building is reliant on several parameters like surrounding weather variables, building characteristics, and energy usage patterns. This literature highlights a mechanism integrating the Internet of Things (IoT) and some widely used machine learning algorithms to create a predictive model that can be used for forecasting smart building indoor temperature. This predictive model has been trained with an online learning methodology for developing viability to a completely unfamiliar dataset. The paper carries out a Machine Learning based experimentation on recorded real sensor data to validate the approach. Following that, the paper suggests integrating the following strategy into an Edge Computing based IoT architecture to enable the building to work in an energy-efficient fashion (Paul, D., Chakraborty, T., Datta, S. K., & Paul, D., 2018, August).

In 2018, Vara et al. a paper entitled "Fault-Tolerant Temperature Control Algorithm for IoT Networks in Smart Buildings." They stated that The monitoring of the Internet of things networks depends to a great extent on the availability and correct functioning of all the network nodes that collect data. These network nodes must correctly satisfy their purpose to ensure the efficiency and high quality of monitoring and control of the Internet of things networks. This paper focuses on the fault-tolerant maintenance of a networked environment in the Internet of things domain. Based on continuous-time

Markov chains, together with a cooperative control algorithm, novel feedback model-based predictive hybrid control algorithm is proposed to improve the maintenance and reliability of the Internet of things network. Virtual sensors are substituted for the sensors that the algorithm predicts will not function properly in future time intervals; this allows for maintaining reliable monitoring and control of the Internet of things network. In this way, the Internet of things network improves its robustness since our fault tolerant control algorithm finds the malfunctioning nodes that are collecting incorrect data and self-corrects this issue by replacing malfunctioning sensors with new ones. In addition, the proposed model is capable of optimizing sensor positioning. As a result, data collection from the environment can be kept stable. The developed continuous-time control model is applied to guarantee reliable monitoring and control of temperature in a smart supermarket. Finally, the efficiency of the presented approach is verified by the results obtained in the case study (Casado-Vara, R., Vale, Z., Prieto, J., & Corchado, J. M., 2018).

3. The Proposed Method

IoT devices are spread all over the building to monitor it in smart buildings. This research focused on temperature IoT devices. Notice that temperature IoT devices depend on the physical topology of the building and the non-uniformity of the building temperature. Thus, we will assume that the data collected by IoT devices are heterogeneous due to the topology of the building (offices that are closed and with their heating, corridors, large common areas, etc.) and the non-uniformity of the temperature (i.e., the temperature vary between different zones because users can choose their comfort temperature, open or close doors, and windows, etc.). Due to the reasons mentioned above, heterogeneous data are not optimal inputs for temperature control algorithms; as a result, the precision of control algorithms is lower than if homogeneous data were used. Addressing the inefficiencies mentioned above due to the lack of a control algorithm that uses heterogeneous data, one could consider using homogeneous data to improve their accuracy and, in particular, homogeneous data collected by IoT networks. This research has the following challenges in the field of monitoring and control in smart buildings with IoT networks. IoT networks usually have heterogeneous data, but most algorithms do their best with homogeneous data. Therefore, there is a need to be able to make the algorithms we apply for monitoring and control of IoT networks can have as homogeneous input data. This requires models, theories, methodologies, tools, and mechanisms to develop a system that can adapt and organize itself to possible changes in its surroundings (Casado-Vara, R., Martin-del Rey, A., Affes, S., Prieto, J., & Corchado, J. M., 2020). The purpose of the proposed method is temperature control with heterogeneous data collected from the IoT network of smart buildings.

A combination of mathematical and artificial intelligence techniques has been proposed to deal with the problem of data heterogeneity and the differences in smart buildings in different regions. In this model, an intelligent and self-adaptive framework has been developed that allows the use of temperature control algorithms in all types of Internet of Things networks. The operation of this model starts with data collection by IoT nodes. These data are usually heterogeneous (for example, the temperature collected by the IoT network in a smart building varies greatly depending on the area of the smart building where they are collected). The difference is not in the type of data, it is only the volume and number that forms this heterogeneity. The proposed method includes the following methods or algorithms:

- First, a graph is constructed where the nodes can be IoT nodes and the edges (edges) of the graph are doors or corridors that connect to mobile (IoT) nodes. Since a graph is built with an IoT network, a complex network can be built with this graph.
- The most important and main part, which is clustering and basically deciding on the data collected in the previous step, is done in this step by applying a combined clustering algorithm consisting of three Gaussian, K-means, and hierarchical algorithms on the different and heterogeneous data of the previous step. And in addition to homogenization, it specifies the desired data cluster.

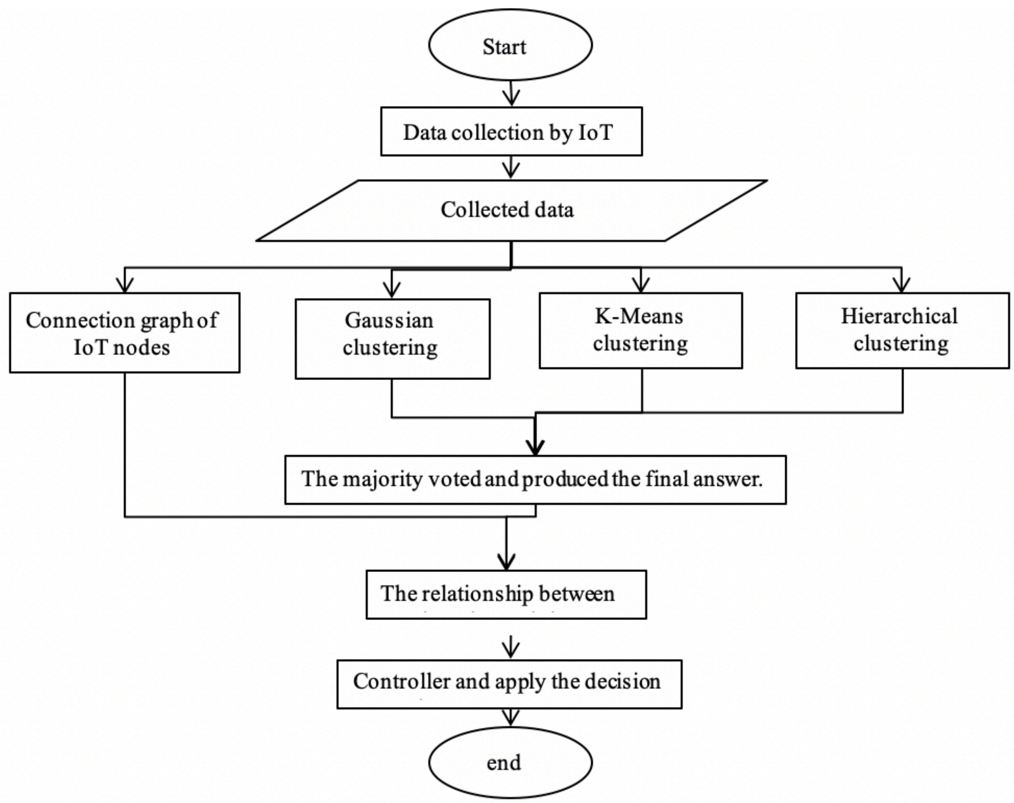

Fig.1. Block diagram of the proposed method

• Specifying the collected data cluster makes a multiplex in which each layer represents a cluster.

• Multiplex layers that do not have communication networks use virtual nodes to build a connected network. • The control algorithm required for use will be applied to homogeneous data in each layer, depending on the purpose. This way, any decision can be applied right in the next step.

• Each layer is placed on the complex network, and the control signal is sent to each operator assigned to the IoT nodes. The flowchart of the proposed method is shown in Figure 1.

Graph design module: the graph is constructed from the topology of the IoT nodes in the intelligent building. In this way, a graph is formed where the vertices of the graph are the IoT nodes, and the edges of the graph are the physical connections between the rooms where the IoT nodes are located (i.e., there is no obstacle on the way from one node to another). That is, if between the rooms in which the IoT nodes are located, there is a door and a corridor, then there is an edge in the graph. The graph represents the heat transfer between the physically connected rooms.

Clustering module: temperature data collected by the IoT network of the smart building are usually heterogeneous data (taking into account the topology and the non-uniformity of the smart building). The main purpose of applying the clustering technique is to find IoT devices with similar data (regarding topology and non-uniformity of the smart building). Thus, we can create a "similar group of IoT devices" based on their collected data in the clusters. So, the IoT slicing method transforms heterogeneous data into groups of similar data based on the cluster technique output. We call these groups of similar data homogeneous temperature data groups. IoT devices from these IoT devices with homogeneous data with similar smart building topology (devices in corridors, devices in common areas, etc.) and the same smart building uniformity (similar heating/cooling systems, similar temperature conform choices, etc.). Then, these groups of homogeneous temperature data are different layers in the multiplex, this way, we can apply the algorithm in each layer, and the algorithm's input is the homogeneous temperature data.

The distribution of zones with topological and uniform characteristics in smart buildings are usually: (1) Rooms or offices. In these rooms, the users can choose their comfort temperature; therefore, they will have slightly different temperatures from the rest of the building. (2) Open-plan zones. These common areas, such as the reception, the rooms where the offices are, etc., have different temperatures because the doors or windows can be open, which can modify the temperature of the smart building in relation to what is outdoors. (3) Corridors. The corridors are areas that are open and communicate the different areas of the building that may be at different temperatures.

The proposed clustering method uses a two-dimensional temperature array and graph output with clusters. Gaussian mixture model selective clustering method that gives good flexibility. K-means clustering is the second cluster used for clustering and determining the appropriate temperature of building parts. Another clustering method is the hierarchical method, which, unlike discriminative clustering, divides objects into separate groups; hierarchical clustering shows the result of clustering at each level of distance. These levels are hierarchical. "Tree" is used to display the clustering results hierarchically. This method is an effective way to display the results of hierarchical clustering.

Adapting and applying control: Many real-world systems are simultaneously characterized by multiple interactions. Human beings are involved in many social interactions at the same time [44]. The need for networks that allow different types of interaction has long been recognized in sociology. The quality of a social bond can be completely different; transport networks have a multi-relational structure, where different types of links correspond to different modes of transport. Multiplex networks allow us to represent different types of interaction between nodes. Multi-layer or multiplex networks can be defined as those that incorporate different connectivity channels and describe systems that are interconnected with different categories of connection: each channel is represented by a sub-network or layer, and the same node can have different types of links and different neighborhood in each layer.

The technical problem in this research is that we have a graph with the temperatures of a smart building that are interrelated and cannot be analyzed in isolation. Therefore, the solution is that an approach is made using the multiplex technique. In this case, it can be modeled as a multiplex network where the nodes (i.e., IoT devices) are the sensors, and the edges are the physical connections between the different IoT devices. Each layer of this multiplex network will correspond to a temperature cluster. Each layer node has a temperature similar to the other nodes of the same layer, and the edges are the physical connections between them. The importance of using the multiplex approach to solve this problem is that each multiplex layer will have homogeneous temperatures, improving the functioning of the algorithms applied to the data.

4. Results

In this section, the proposed method was tested, and the results of different stages of evaluation are displayed. To simulate the method, Matlab and Weka software was used, and to evaluate the proposed method, the numerical temperature prediction data set available in the data source was used. This set contains fourteen in situ temperature forecast data and five geographic covariates in Seoul, South Korea, in summer. These data are for bias correction of next-day maximum and minimum air temperature forecast administered by the Korea Meteorological Administration in Seoul, South Korea. This data includes summer data from 2013 to 2017. The input data mainly consists of multidimensional model prediction data, maximum and minimum temperature of the location, and geographic covariates. This data has two outputs (i.e., maximum and minimum air temperature for classification). Data validation was done for the period 2015 to 2017. The dataset contains 7752 records and 15 features. The characteristics of this dataset are shown in Table 1.

Table 1. Characteristics of the dataset

1. station - used weather station numbers: 1 to 25
2. Date - Present day: yyyy-mm-dd ('2013-06-30' to '2017-08-30')
3. Present_Tmax - Maximum air temperature between 0 and 21 h on the present day (Â°C): 20 to 37.6
4. Present_Tmin - Minimum air temperature between 0 and 21 h on the present day (Â°C): 11.3 to 29.9
5. LDAPS_RHmin - LDAPS model forecast of next-day minimum relative humidity (%): 19.8 to 98.5
6. LDAPS_RHmax - LDAPS model forecast of next-day maximum relative humidity (%): 58.9 to 100
7. LDAPS_Tmax_lapse - LDAPS model forecast of next-day maximum air temperature applied lapse rate (Â°C): 17.6 to 38.5
8. LDAPS_Tmin_lapse - LDAPS model forecast of next-day minimum air temperature applied lapse rate (Â°C): 14.3 to 29.6
9. lat - Latitude (Â°): 37.456 to 37.645
10. lon - Longitude (Â°): 126.826 to 127.135
11. DEM - Elevation (m): 12.4 to 212.3
12. Slope - Slope (Â°): 0.1 to 5.2
13. Solar radiation - Daily incoming solar radiation (wh/m2): 4329.5 to 5992.9
14. Next_Tmax - The next-day maximum air temperature (Â°C): 17.4 to 38.9
15. Next_Tmin - The next-day minimum air temperature (Â°C): 11.3 to 29.8

Simulation evaluation: To evaluate the method presented in the previous section, first, using the introduced data set, the three clusters introduced in the proposed method were implemented on it. This action is the first step of the proposed method, which is the clustering of the collected data, and the results of the proposed method were compared with other classification algorithms. The results of the K-means algorithm, the Gaussian mixture method, and the results of the hierarchical method are shown in Table 2. As it is unclear, the K-means method has an imbalance in the number of K equal to 2, and cluster one has 51% of the members, and with the increase of K value, this process becomes more balanced. In the number 6, this value reaches an acceptable relative balance. Of course, considering that the proposed method works on three fixed clusters, the number of three clusters is used in all clustering. The results related to the Gaussian mixed method have an imbalance in the number of clusters equal to 2, and cluster one has 51% of the members, and this trend becomes more balanced with the increase of the K value. In the number 6, this value reaches an acceptable relative balance. Also, according to the results table, the hierarchical method has not achieved the balance between the clusters in any of the divisions. Based on these results, the Gaussian mixture method has results close to the K-means method, so that we can accept these results' acceptability.

Table 2: Clustering results of three K-means algorithm techniques, Gaussian mixture method, and hierarchical method

K-means algorithm clustering results												
	K=2		K=3		K=4		K=5		K=6		K=7	
Cluster number	Number of Members	Percentage of members	Number of Members	Percentage of members	Number of Members	Percentage of members	Number of Members	Percentage of members	Number of Members	Percentage of members	Number of Members	Percentage of members
0	4574	59	4042	52	2224	29	2447	32	1524	20	1528	20
1	3178	41	2775	36	2369	31	2084	27	998	13	998	13
2			935	12	934	12	623	8	1389	18	1391	18
3					2225	29	2281	29	1686	22	1680	22
4							317	4	1225	16	1222	16
5									930	12	621	8
6											312	4
Clustering results of Gaussian mixture method												
	nC=7		nC=6		nC=5		nC=4		nC=3		nC=2	
Cluster number	Number of Members	Percentage of members	Number of Members	Percentage of members	Number of Members	Percentage of members	Number of Members	Percentage of members	Number of Members	Percentage of members	Number of Members	Percentage of members
0	3961	51	1406	18	1231	16	932	12	1595	21	1591	21
1	3791	49	3053	39	2175	28	1498	19	1461	19	1461	19
2			3293	42	1853	24	2027	26	485	6	1400	18
3					2493	32	1523	20	932	12	310	4
4							1772	23	1399	18	485	6
5									1880	24	620	8
6											1885	24
Hierarchical clustering results												
	nC=7		nC=6		nC=5		nC=4		nC=3		nC=2	
Cluster number	Number of Members	Percentage of members	Number of Members	Percentage of members	Number of Members	Percentage of members	Number of Members	Percentage of members	Number of Members	Percentage of members	Number of Members	Percentage of members
0	7251	94	6960	94	6735	87	6728	87	4532	58	2825	36

1	501	6	351	4	187	2	187	2	34	0	34	0
2			441	6	220	3	220	3	102	1	102	1
3					610	8	610	8	422	5	384	5
4							7	0	7	0	7	0
5									2655	34	1937	25
6											2463	32

After the clustering, three clustering methods are used in a voting system to continue the proposed method. Due to the reliability of the K-means method and the use of majority voting, the weights of the K-means method are equal to 0.4. Gaussian 0.3 and hierarchical method is also 0.3. These weights are obtained based on the error test. An example of this selection and formation of the final cluster is shown in Table 3

Table 3. An example of weighted voting of the proposed method

Kmeans	hierarchical	Gaussian	Final answer
2	0	1	2
1	2	0	1
1	1	0	1
0	1	1	1
1	2	0	1
1	1	0	1
1	1	0	1
0	1	0	0
0	2	0	0
0	0	0	0

As shown in Table 3. The proposed method selects the K-means method as the final vote in cases where all three methods differ. In the following, the clusters obtained from the number equal to 3 were compared with different classification methods to obtain the power and improvement of the proposed method compared to the known and widely used methods. To evaluate the important criteria of accuracy, TP Rate, FP Rate, accuracy and readability, and F-Measure, and in all experiments, an accuracy test by K-Fold method with K=10 was used. In this type of validation, the data is divided into K subsets. Of these K subsets, each time, one is used for validation, and another K-1 is used for training. This procedure is repeated K times, and all data are used precisely once for training and once for validation. Finally, the average result of these K validation times is chosen as a final estimate. Of course, other methods can be used to combine the results. Normally, 10-fold is used. K-nearest neighbor, J48, SMO, Decision Table, algorithm based on Bayes theory, and Bagging have been used for evaluation. In this part, the evaluation of the proposed method is done. For this purpose, the proposed method was evaluated with the widely used and basic classifications used in most studied articles. The required criteria are categorized in the following section:.

Table 4. The class assigned by the model

Negative	Positive		
FN Rate: It represents the number of records whose true category is positive, and the classification algorithm has mistakenly recognized their category as negative.	TP Rate: It represents the number of records whose real category is positive, and the classification algorithm has recognized their category as positive.	Positive	Real class
TN Rate: It indicates the number of records whose real category is negative, and the classification algorithm has correctly recognized their category as negative.	FP Rate: It represents the number of records whose actual category is negative, and the category classification algorithm mistakenly recognizes them as positive.	Negative	

Accuracy: refers to the ratio of the total number of samples correctly classified by the classifier to the total number of samples (Lakshmi, K., Visalakshi, N. K., Shanthi, S., & Parvathavarthini, S., 2017; Tatbul, N., Lee, T. J., Zdonik, S., Alam, M., & Gottschlich, J., 2018).

$$Accuracy = \frac{TN+TP}{TN+FN+TP+FP} \quad (1)$$

Precision and Recall: refers to the TP and is expressed as 2 and represents the TN and is expressed as 3 (Lakshmi, K., Visalakshi, N. K., Shanthi, S., & Parvathavarthini, S., 2017; Tatbul, N., Lee, T. J., Zdonik, S., Alam, M., & Gottschlich, J., 2018)

$$Recall = \frac{TP}{FN+TP} \quad (2)$$

$$Precision = \frac{TP}{FP+TP} \quad (3)$$

F- Measure: is the one that has the combination of both precisions and Recall, which is used to compute the score (Lakshmi, K., Visalakshi, N. K., Shanthi, S., & Parvathavarthini, S., 2017; Tatbul, N., Lee, T. J., Zdonik, S., Alam, M., & Gottschlich, J., 2018).

$$F - Measure = 2 * \frac{Precision*Recall}{Precision+Recall} \quad (4)$$

Based on the mentioned criteria, the efficiency of the proposed algorithm has been evaluated based on K-nearest neighbor classes, Decision Table and SMO.

Figure 2-a shows the evaluation results between the methods in terms of the TP criteria, and as it is known, the Bagging and SMO method has the highest TP value

compared to other methods. The lowest value of this criterion corresponds to the KNN method, with a value of 3. The closest method to the proposed method is the j48 method, and other methods have significant differences in this regard from the proposed method.

Figure 2-b The proposed, and other methods are compared in terms of FP criteria. As it is clear in the figure, the highest FP detection value corresponds to the KNN method, with a value of 3. After this method, the KNN method with K equal to 5 with a value of 0.097 has the highest value compared to other methods. In this evaluation, the SMO and Bagging method received the lowest value, the closest follower in the TP criterion.

Figure 2-c different methods are compared with each other in terms of Precision criteria. As shown in the figure, in this criterion, the SMO and Bagging method has a better value than the j48 method as the closest follower, and the reason for this is the high value of both TP and FP, that is, according to the proposed method. It correctly recognizes the right and the wrong, which has caused the increase of this criterion because it is obtained from the combination of the two. On the other hand, the SMO method has a very small FP value, and this difference has caused a decrease in the Precision criterion. The lowest value in this criterion also corresponds to the KNN method with K equal to 3.

Figure 2-d different methods were compared with each other regarding recall criteria. As shown in the figure, the SMO and Bagging method has the highest recall value compared to other methods. The lowest value of this criterion corresponds to the KNN method with K equal to 3. The closest method to the proposed method is the j48 method, and other methods have significant differences in this respect from the proposed method.

Figure 2-e The proposed, and other methods are compared in terms of F-Measure. As shown in this figure, in this criterion, SMO and Bagging methods have a higher value than J48 and Decision Table methods as the closest pursuers. The reason for this is the high value of both TP and FP; that is, the proposed method correctly detects both true and false, which has caused an increase in this criterion because this criterion is obtained from the combination of Precision and Recall. The lowest value in this criterion also corresponds to the KNN method with K equal to 3.

Finally, in Figure 2-f, different methods were compared with each other in terms of accuracy criteria. As shown in the figure, the SMO method has higher accuracy than all other methods. The bagging method is after SMO with a difference of 0.05%, and J48 is the following method in this order. KNN methods with K equal to 3 and K equal to 5 are also in the following ranks with a difference of about 11%. They are among the algorithms that have respectively shown the lowest value in accuracy. Since many classifications have been done on the clustering results, it can be assumed that they are reliable.

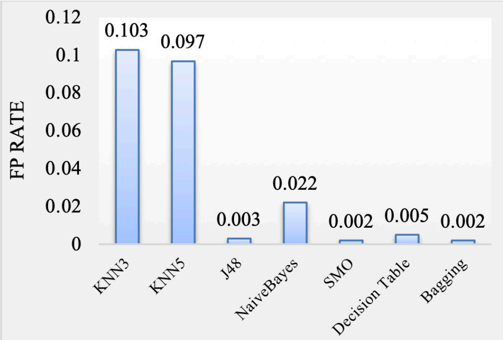

Fig a: Evaluation of FP criterion

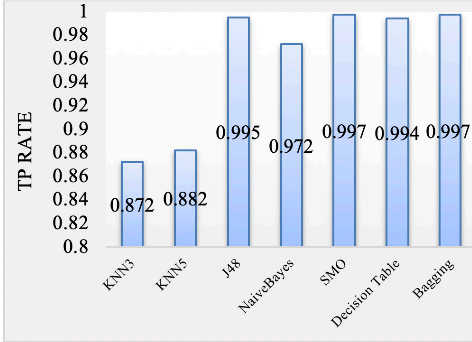

Fig b: Evaluation of TP criterion

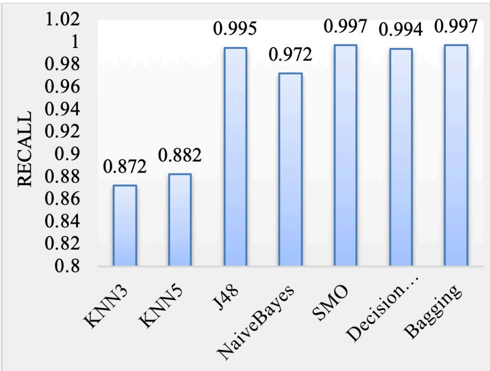

Fig c: Evaluation of Recall criterion

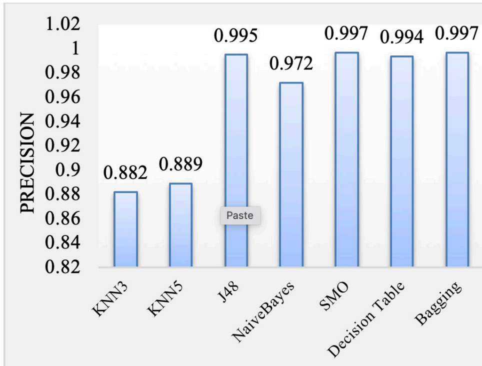

Fig d: Evaluation of Precision criterion

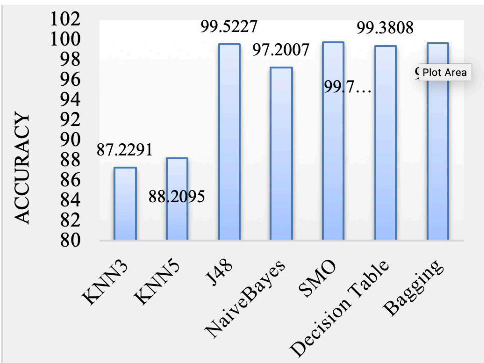

Fig e: Evaluation of Accuracy criterion

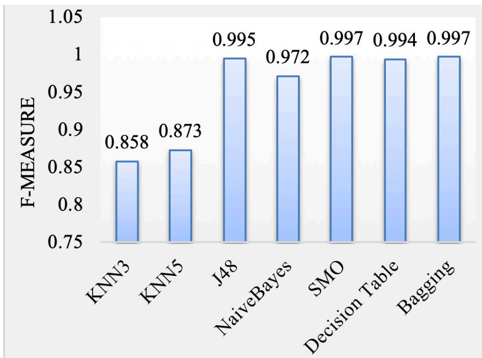

Fig f: Evaluation of F-Measure criterion

Fig 2: Evaluation of classification methods on the clustering dataset with the proposed method in different criteria

5. Conclusions

In this research, a method was used in a smart home, and a graphically oriented program was prepared to achieve the best temperature in the home and its surroundings. IoT collected the information. Based on that, through an artificial intelligence method and the combination of three different hierarchical clustering methods, Gaussian mixture and K-means, which are of different types and each has its unique advantages, the initial results were achieved, and finally, by using a system Maximum weighted voting, the final results were extracted and used. According to the simulations and evaluations, the proposed method has had acceptable results in different classifications. Its results can be confirmed with the lowest value of 88% and the highest value of 100%. All the primary hypotheses of temperature control for smart buildings based on the Internet of Things are possible, the conversion of heterogeneous temperature data into homogeneous data is feasible, and the performance improvement of the proposed method is possible for the correct extrapolation data in the data collected from the IoT network based on the IoT slicing technique. Applying the clustering technique to find IoT devices with similar data is possible. The best clustering method between the Gaussian mixed model, K-means, and hierarchical methods was measured by the Voting technique, and the right method was chosen and, according to the results, achieved the goals.

Among the future works that can be suggested on the subject of this thesis are the full implementation of the method in the real world and the examination of the results obtained in the real world and its comparison with the evaluations made in the simulators, and the elimination of the possible shortcomings of the proposed method in the real world, the combination of exploratory and optimization algorithms. Others are the honey bee and jumping frog algorithm or the use of fuzzy algorithms to achieve balance in decision-making and reduce application time.

References

Albino, V., Berardi, U., & Dangelico, R. M. (2015). Smart cities: Definitions, dimensions, performance, and initiatives. *Journal of urban technology, 22(*1), 3-21.

Casado-Vara, R., De la Prieta, F., Prieto, J., & Corchado, J. M. (2019, December). Improving temperature control in smart buildings based in IoT network slicing technique. In *2019 IEEE Global Communications Conference (GLOBECOM) (pp. 1-6)*. IEEE.

Casado-Vara, R., Martin-del Rey, A., Affes, S., Prieto, J., & Corchado, J. M. (2020). IoT netwo*rk slicing on virtual layers of homogeneous data for im*proved algorithm operation in smart buildings. Future generation computer systems, 102, 965-977.

Casado-Vara, R., Vale, Z., Prieto, J., & Corchado, J. M. (2018). Fault-tolerant temperature control algorithm for IoT networks in smart buildings. *Energies, 11*(12), 3430.

Floris, A., Porcu, S., Girau, R., & Atzori, L. (2021). An iot-based smart building solution for indoor environment management and occupants prediction*. Energies, 14*(10), 2959.

Gohar, M., Ahmed, S. H., et al. (2018). A big data analytics architecture for the internet of small things. *IEEE Communications Magazine, 56*(2*), 128-133.*

Lakshmi, K., Visalakshi, N. K., Shanthi, S., & Parvathavarthini, S. (2017). Clustering Categorical Data Using k-Modes based on Cuckoo Search Optimization Algorithm. *Ictact journal on Soft Computing,* 8(1).

Lymperopoulos, G., & Ioannou, P. (2020). Building temperature regulation in a multi-zone HVAC system using distributed adaptive control. *Energy and Buildings, 215,* 109825..

Nowicka, K. (2014). Smart city logistics on cloud computing model. Procedia-Social and Behavioral Sciences, 151, 266-281.

Paul, D., Chakraborty, T., Datta, S. K., & Paul, D. (2018, August). IoT and machine learning based prediction of smart building indoor temperature. *In 2018 4th International Conference on Computer and Information Sciences (ICCOINS) (pp. 1-6).* IEEE.

Sharma, V., & Tiwari, R. (2016). A review paper on "IOT" & It's Smart Applications. *International Journal of Science, Engineering and Technology Research (IJSETR), 5*(2), 472-476.

Talei, H., Benhaddou, D., Gamarra, C., Benbrahim, H., & Essaaidi, M. (2021). *Smart Building Energy Inefficiencies Detection through Time Series Analysis* and Unsupervised Machine Learning. *Energies, 14*(19), 6042.

Tatbul, N., Lee, T. J., Zdonik, S., Alam, M., & Gottschlich, J. (2018). Precision and recall for time series. *Advances in neural information processing systems,* 31.

Wickramasinghe, A., Muthukumarana, S., Loewen, D., & Schaubroeck, M. (2022). Temperature clusters in commercial buildings using k-means and time series clustering. *Energy Informatics, 5*(1), 1-14.

Submitted: 19.12.2021
Accepted: 20.05.2022