Indoor Air Quality Predictions For Automation

Kadir Gökdeniz

Ankara University

Computer Engineering Department

May 17, 2024

Abstract

This study explores the application of home automation systems for indoor air quality prediction, leveraging real-time data on temperature, humidity, pressure, CO2 levels, occupancy status, and window conditions. Despite potential challenges in data quality and responsiveness, the research employs deep learning techniques to predict air quality. Comparative analysis of GRU and LSTM models reveals GRU's superior performance across multiple metrics, emphasizing its generalization capability. However, model selection requires consideration of all scenarios. Additionally, the study introduces Air-Smart Controller, enabling users to monitor predictions and control home automation systems. In conclusion, the research underscores the promise of home automation in air quality prediction, offering insights into neural network architectures and contributing to advancements in automation technology and air quality management.

Keywords: Automation System, Air Quality, Deep Learning, Machine Learning

1 Introduction

The origins of automation systems can be traced back to before the Common Era. However, the Industrial Revolution is what mass-transformed human life. While the primary domain benefiting from automation systems has been industrial production, as we entered the 2000s, they began to transform our lives in many other areas. Automation systems are reshaping our lives in various fields such as automatic driving systems and transportation, energy management, smart grids, building automation and smart home technologies, telecommunications and communication management, healthcare, and medical technologies.

With the emergence of Covid-19, global quarantine and social distancing measures have forced people to spend more time at home. These prolonged periods spent indoors have brought along comfort and health problems. Smart home technologies promise to address these comfort and health issues. Heating, ventilation, and air conditioning (HVAC) systems work seamlessly with smart home systems to automate indoor air quality. Using microcontrollers, information such as indoor and outdoor temperature, relative humidity, pressure, CO^2 levels, occupancy status, and window status (open/closed) is measured in real-time.

The data obtained from sensors may encounter some potential problems. For example, sensors may produce erroneous results for any reason. The demands of homeowners may vary temporarily, and the automation process may struggle to respond to these demands. On the other hand, measurements of indoor air quality may provide information about potential disasters in the future, but the automation system may be inadequate in responding to them. For instance, high temperature may indicate a fire while high humidity levels may indicate a water leak or flooding. A simple automation system lacks foresight about future situations, hence it is insufficient in providing proactive solutions.

Model Predictive Control (MPC) is a structure developed to predict and analyze future states of automation systems and to provide energy savings. MPC systems integrate automation systems and prediction models to form a hybrid system. The mentioned system design necessitates a Human-Machine Interface (HMI) that enables users to understand how automation systems operate and what mode they are in. HMI simply refers to a software platform where people can intervene in the operation of automation systems and view data.

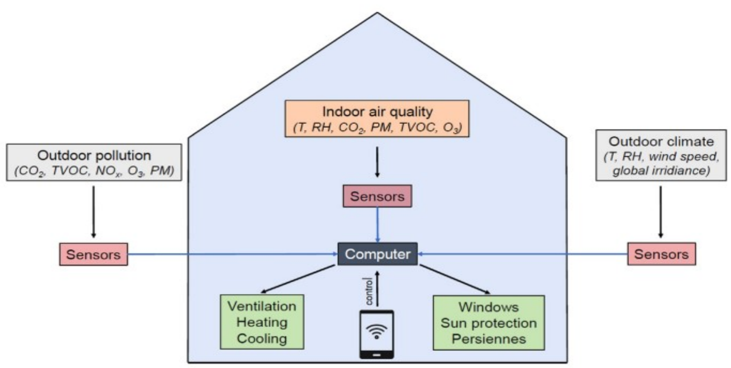

Figure 1: Indoor Air Quality Automation.

Dynamics systems emerge as structures that enable systems to better understand user needs and facilitate the creation of smarter strategies. This approach allows for the development of faster and more effective strategies towards goals, which can improve living standards. However, during the installation phase of automation systems, they are tested, and test engineers may find the data quality low due to insufficient data and the presence of outlier values. Prediction data obtained with machine learning methods can help meet such needs. Additionally, the operation of the automation system can be automated with new data using prediction data. This is another benefit of generating prediction data.

Human-Machine Interface (HMI) systems developed in smart home technologies promote low energy consumption and utilize specialized technologies to enhance indoor air quality [1]. These systems also aid in developing new strategies to protect against the adverse effects of climate change. As noted in study [2], modern sensor technologies not only diagnose climatic characteristics but also detect air pollutant particles and transmit this data to smart home systems. However, new smart home owners need to adapt to situations like not being able to manually open windows, which is a significant adjustment that many find challenging.

According to the research, we will need to further investigate how to create healthy living environments [3]. In the United States, a group of health experts has extensively addressed how climate change affects our lifestyle [4]. Upon reviewing the report of this research, it was found that climate change significantly increases health risks for the elderly, children, and people with cardiac/respiratory diseases. Study [5] also focused on the impacts of climate change on indoor air quality and emphasized that systems regulating indoor air quality (HVAC) are a significant factor affecting the prevalence of diseases.

Health-based smart technologies form a complex network with various technological devices, introducing the concept of Ambient Assisted Living (AAL) into our lives. This network integrates with personal healthcare monitoring and tele health systems [6]. In addition to helping reduce the increasing demands in the healthcare sector, this technology also supports improving energy efficiency and reducing negative environmental impacts.

The data collected by sensors also aims to enhance user comfort. Study [7] shows that the need for humidity and temperature sensors in smart home systems has increased over the years. The number and placement of sensors are crucial. Sensors placed inside the home can be of different types, and their measurement capacities and performances bring along various advantages and disadvantages [7]. To better illustrate this situation, Figure 2 compares different humidity sensors, aiding experts in making informed choices.

Another important issue is IoT protocols. In smart home systems, data collected from sensors should be securely, quickly, and comprehensively transported in an appropriate format. Protocols such as Wi-Fi (802.11), Bluetooth, Zigbee, and Z-Wave are commonly used technologies for transmitting data in smart home automation systems [8]. While Wi-Fi provides broad coverage and high bandwidth, Bluetooth is ideal for short-range communication. Zigbee and Z-Wave, on the other hand, are preferred for establishing wireless communication among devices with low power consumption. These protocols are selected based on the features of the devices and user preferences, each offering different advantages.

Active Material	Electrodes	Substrate	Range (%RH)	Sensitivity [1]	Ref.
Gravure printed MWCNTs	Screen-printed, Ag	PI	30–60	0.96%/% RH	[10]
Gravure printed MWCNTs	Screen-printed, Ag	PI	10–90	–	[13]
Drop-cast TiO_2 nanoflowers	Gravure, Ag	PI	20–95	485.7/RH%	[14]
Screen-printed MEPAB/CMDAB/MMA copolymer	Screen-printed, Ag/Au	PI	20–95	0.0586 logΩ/% RH	[15]
Screen-printed epoxy/IPN polyelectrolyte	Chemical Etching–Plating (Ni/Au)	PI	20–95	0.046 logΩ/% RH	[18]
Drop-cast SnO_2/rGO	Chemical Etching–Plating (Cu/Ni)	PI	11–97	15.19–45.02%	[19]
Spin-coat PEDOT:PSS (15%) + PVA (SAW)	Photolithography	$LiNbO_3$	0–80	350 Ω/% RH	[27]
Spin-coat PEDOT:PSS (5 wt%) + $ZnSnO_3$ (5 wt%)	Photolithography, Au	$LiNbO_3$	0–90	–	[28]
Screen-printed MDBBAC/MMA (70/30) Polyelectrolyte	Screen-printed, Ag–Plating (Cu/Ni/Au)	Glass Epoxy	20–95	0.0349 logΩ/% RH	[25]
EHD Graphene/methyl-red	Inkjet-printed, Ag	PET	5–95	96.36%	[20]
Drop-cast Pt/MoS_2 (0.25:1)	Photolithography, Au	Ceramic	35–85	~4000 times (85 % RH)	[26]
Gravure printed CNT	Screen-printed, Ag	PET	20–80	0.1%/% RH	[16]
Screen-printed TiO_2-Cu_2O-Na_2O	Screen-printed, Pt	Al_2O_3	20–95	–	[32]
Inkjet-printed PANI	–	Polyester	20–95	–	[23]
Micro-pipette deposited Nafion	Screen-printed Ag on screen-printed PU	Polyester Cotton Fabric	30–90	–	[24]
Gravure printed FMWCNT/HEC (1:6 *w/w*)	Screen-printed, Ag	PET	20–80	0.048/% RH	[21]
Inkjet-printed PEDOT:rGO-PEI/Au NPs	–	PET	11–98	7.41–51.60%	[22]
Spin-coated Fe_2O_3	Inkjet-printed, Ag	PET	0–100	~88.89%	[33]
Substrate	Inkjet-printed, Ag	Paper	18–88	–	[12]
Substrate	Inkjet-printed Ag & PEDOT: PSS	Paper	0–85	0.0492 & 0.0551 logΩ/% RH	[17]

Figure 2: Comparison of humidity sensors.

With all these mentioned details, IoT systems and automation systems emerge and are utilized alongside machine learning methods. These systems impact humans comprehensively, with effects ranging from health to security, and even entertainment areas. For these reasons, they will continue to be among the focus areas for researchers and individuals in the near future.

2 Literature Review

As an early study, a research conducted in 2011 used multivariate analysis of variance (MANOVA) to predict the indoor air quality of a metro system [9]. In this study, the data was divided into three categories: spring and fall as the first category, summer as the second category, and winter as the third category. The variance analyses were conducted separately for these categories. This study indicates that indoor air quality varies depending on the seasons.

Another study is based on an optimization work conducted by Gaurav Priyadarshi and B. Kiran Naik [10]. In this study, they perform regression prediction using the K-nearest neighbors (KNN) algorithm. For these predictions, they utilize a model called Finite Difference Transient Model (FDM). The study thoroughly discusses the potential limits and advantages of this model.

The study conducted by T. Akilan and K.M. Baalamurugan [11] focuses on predicting indoor weather conditions in agriculture using GRU (Gated Recurrent Unit) and LSTM (Long Short-Term Memory) to develop a warning system that alerts before plants encounter unfavorable conditions. This IoT (Internet of Things) based study is significant for understanding how GRU and LSTM models are used in time series analysis and how sensor data interacts.

Another study conducted by Elnaggar and other researchers presents a risk analysis of the indoor air quality required for the preservation of artworks in a museum located in the Mediterranean region [12]. This study also provides a comprehensive examination of how risk analysis is conducted in the preparation of automation systems. By identifying the necessary indoor air quality parameters for the preservation of artworks, the study details how these parameters can be integrated into automation systems and how these systems can be managed effectively.

A. Gururaj and colleagues conducted a classification study on air quality prediction using artificial neural networks and gradient boosting models, and they developed a website to present their findings [13]. This study is an important resource for our project to compare HMI (Human-Machine Interface) design with literature studies and to learn from current applications.

In another study conducted by Nyoman Kusuma Wardana and other researchers, a comprehensive study was presented on the simultaneous display of predictive and real-time data in SCADA (Supervisory Control and Data Acquisition) devices [14]. This study thoroughly explores how data can be integrated and presented simultaneously in SCADA systems.

The study numbered [15] compared the performance of various machine learning methods, such as Dummy, Huber, Ada Boost, GBoost, Random Forest, XGBoost, KNN, Decision Trees, LSTM, and ARIMA models, in indoor air quality regression problems. This research has a significant impact on selecting the most appropriate model.

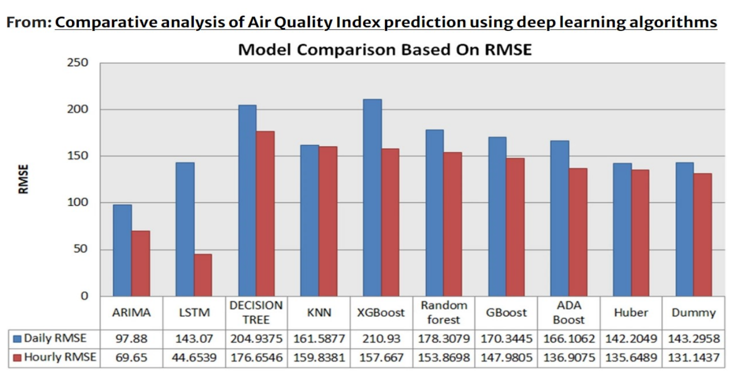

Figure 3: Model Comparison on RMSE.

3 Automation and Prediction Systems

The automation pyramid is used because it provides a better approach for explaining automation processes [16]. Figure 4 presents the automation pyramid. The automation process has a hierarchical structure. If any part of the pyramid is missing or has significant problems, it can hinder the automation processes. From a general perspective, the fundamental parts of the automation process are hardware, software, and management levels. When we talk about hardware, we refer to various sensors, PLC (Programmable Logic Controller), and PLD (Programmable Logic Device) levels. The software part is the Supervisory level, which includes SCADA (Supervisory Control and Data Acquisition) devices. MES (Manufacturing Execution System) and ERP (Enterprise Resource Planning) are the management levels.

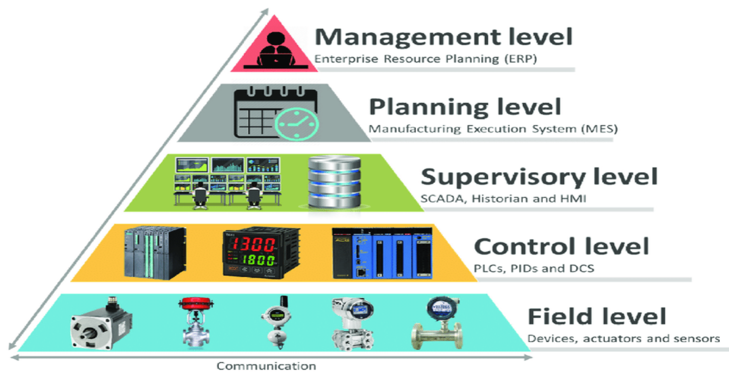

Figure 4: Automation Pyramid.

ISA has a significant role in the creation, development, and adoption of the automation pyramid [17]. According to the hierarchy presented by the pyramid, a five-step process is adopted. Implementation starts from the bottom level. The bottom level is called the "field level," where various sensors interact. One of its most important features is the real-time acquisition and modification of data.

The second part is the control level. Here, the sensors are organized using a programmable logic controller (PLC), microcontrollers, automatic controllers, or a dedicated computer. Control devices can begin their tasks within milliseconds after receiving the information in the decision-making process [18].

The third part is the supervision level. In this part, data from the control level is received within a very short time and filtered. Filtering data is an important step because not all data needs to be sent to and processed by the control system; not all data is included in the higher-level processes. Supervision systems can interact with other devices in the field and ensure the execution of commands generated by certain decisions. However, if there is a discrepancy between the levels, the commands generated at the supervision level may lose their validity.

The planning level is the fourth tier, encompassing manufacturing execution and operation management systems. Compliance with the production plan is observed based on Key Performance Indicators (KPIs). Since the information comes from the control level, it has been filtered; additionally, data is updated in specific patterns. The frequency of data updates may vary; sampling of data is tightly dependent on ISA 95 standards.

At the top of the pyramid is the management level. Commercial systems such as Altai, SAP, and Oracle are utilized to actively operate Enterprise Resource Planning (ERP) systems. The management of all processes is organized from a central point. This can be likened to the main control center that encompasses all home systems for smart home systems. The aim is to easily account for data, and additional factors such as availability and corrective maintenance can also be integrated into the system at this stage.

Automation systems used in smart home setups provide automatic control in various areas of the house, including lighting, heating, ventilation, security, and entertainment. For instance, lighting automation adjusts room lights based on ambient light, promoting energy efficiency and offering homeowners a more comfortable environment. Heating systems optimize every room in the house through temperature sensors and programmable thermostats, enhancing comfort levels and lowering energy costs. Security systems continuously monitor the surroundings of the house through cameras and motion sensors, allowing users remote access. All these automation systems provide homeowners with the flexibility to remotely monitor and manage their homes, enabling a more comfortable, secure, and energy-efficient living experience.

In the study [19] that examines the potential negative effects and risks of systems in smart homes, possible automation system errors and reasons for user dissatisfaction were investigated comprehensively, and the results were presented in detail. According to the research findings, errors arising from a specific hierarchy in automation systems can affect other processes, and the automation system's response to such errors may take a long time or may not be effective.

If the mentioned problems arise from data-related issues, data security can be ensured, data quantity can be increased, and the system can gain foresight by using prediction data. Prediction data can be incorporated into both the management and control parts of the automation system. In the control part, supervisory devices can reduce response time by providing remote access, viewing, and intervention capabilities for data. This approach can be a crucial step in addressing data-based issues in automation systems and making the system more secure and effective.

Our primary objective within the project revolves around the prediction of indoor air quality, thus necessitating the handling of a regression problem concerning time series data. Within the existing literature, a range of machine learning methodologies including Dummy, Huber, Ada Boost, GBoost, Random Forest, XGBoost, KNN, Decision Trees, LSTM, and ARIMA are frequently employed [15]. The outcomes of research indicate that among these machine learning models, Recurrent Neural Networks (RNNs) exhibit superior performance particularly within short-term intervals. Consequently, our focus has been directed towards leveraging RNNs.

4 Deep Neural Networks

In this section, factors affecting the prediction ability will be explained while constructing the deep artificial neural network.

4.1 Dataset

The study [20] aimed at examining household energy consumption, health, and security impacts, and providing support to researchers, presents an open-access dataset. The data were obtained from an apartment in Beijing, China. Using a cloud-based data collection platform called IDCP, the data were collected covering a period from May 31, 2021, to May 31, 2022. The data were collected on a minute-by-minute basis, including household behaviors, thermal environment information, device usage quantities, and also external weather data from the nearest national weather station. This dataset is significant for being the first publicly accessible dataset that concurrently records household behaviors and electricity usage in China. The dataset is made accessible under the title CN-OBEE.

The dataset includes data from 6 different rooms: cloakroom, home office, kitchen, living room, master bedroom, and secondary bedroom. Data for each room, such as temperature, relative humidity, pressure, window status (open/ closed), location, and occupancy status, are recorded at a minute-by-minute frequency. External weather data, on the other hand, are recorded at an hourly frequency and include dry-bulb temperature (°C), relative humidity (%), atmospheric pressure (hPa), wind speed (m/s), wind direction, ground temperature (°C), horizontal total solar radiation intensity (W/m^2), and horizontal diffuse solar radiation intensity (W/m^2). Energy consumption data for household appliances are stored at a minute-by-minute frequency and include data from electric kettle, fridge, rice cooker, computer, TV, upstairs water heater, downstairs water heater, washing machine, and 6 different air conditioners.

4.2 Data Preparation

Ensuring that the data is in numerical format is imperative before training the models. Categorical data such as wind direction in the dataset has been converted to numerical values using Label Encoder. However, dealing with missing values presents another challenge. While low rates of missing values may be tolerable, high rates can significantly compromise data quality. Therefore, missing values have been imputed with the median value. Due to the high rate of missing values, data from the master bedroom has been excluded from the analysis.

To optimize the processing of the data, it has been examined in 10-minute intervals. Since the external weather data is available at hourly frequency, it is assumed that the outdoor weather conditions remain constant within each hour. Therefore, the weather data for each hour is utilized to divide the data into 10-minute intervals.

4.3 Data Visualization

Examining the characteristic features of the data before training is essential for better management of the process. Data visualization can facilitate the analysis of statistical information such as arithmetic mean and standard deviation, as well as other problems. For example, anomaly detection is a crucial approach for analyzing automation systems. Upper and lower limits can be determined using Exponential Moving Averages, aiding in the analysis of abnormal behaviors. Figure 5 presents an analysis of abnormal behavior in some of the indoor temperature data in the living room.

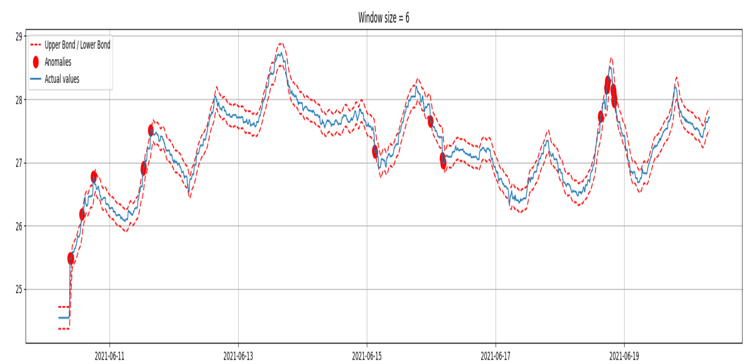

Figure 5: Anomaly detection for living room interior temperature.

4.4 Feature Engineering

The selection of data for training significantly impacts the performance of the training process. In some cases, meaningful potentials can be derived from the data, while in others, training with similar or identical data may exacerbate the tendency for overfitting. Guided by the insights provided by study [9], we are incorporating seasonal information into the data numerically. The high interaction of this information, particularly with relative humidity data, suggests its potential to enhance predictive performance. Conversely, through the use of a correlation matrix, we investigate the interaction among the data, and some data with similar correlation values that are selected for prediction are excluded from the training process.

4.5 Data Preprocessing

In deep neural networks, normalization is a routine process applied to enhance training performance, as also mentioned in study [21]. Normalization procedures reduce the effects of outliers while preserving the fundamental characteristic features of the data. An implementation such as MinMaxScaler rescales the data to a scale between the minimum and maximum values. This process brings the data distribution into a range between 0 and 1, allowing the model to be trained more effectively and converge more rapidly.

The function serves as a data preprocessor, designed to extract batches of samples from a time series dataset for training purposes. It generates sequential samples with a specified "lookback" period, determining how far back each sample should consider, and a "delay" period, indicating the time lag for target values. The function supports shuffling of samples and allows customization of batch size and step size, providing flexibility in data preparation for model training.

Each sample is constructed by looking back over a period of 10 days, with a time interval of 30 minutes between consecutive data points. The time difference between input data and target values is 1 day. These parameters determine how the dataset will be processed and how far back and forward each sample will look. Thus, they play a crucial role in determining how the model will evaluate and predict time series data.

4.6 LSTM

When feedforward artificial neural networks are trained with time series data, recent data is learned strongly while the influence of older data diminishes over time. This problem is called vanishing gradients. Conversely, in situations where the influence of older data is amplified, the exploding gradients problem may occur. Recurrent Neural Networks (RNNs) are used in artificial neural networks to address these issues. RNNs mitigate this effect by feeding a representation of past data alongside new data into the input. This feature enables the prediction of the next data point by deciphering patterns between past and current data.

LSTM-based models can be thought of as an extension of RNNs and have the ability to learn longer-term dependencies. However, remembering longer-term dependencies also brings some challenges in updating the weights. LSTM can decide which gradients to forget and which ones to remember. This feature is often clarified using the gate analogy. The fundamental components of LSTM typically consist of three gates: the forget gate, the input gate, and the output gate.

Forget Gate: This gate typically uses the sigmoid function to decide which information should be forgotten from the LSTM memory. This decision is primarily based on the previous time steps, ht1, and the current input, xt. The output of this gate, ft, takes a value between 0 and 1; where 0 indicates complete forgetfulness of the learned information, and 1 implies the preservation of all information. This output is computed as:

$$ft = (Wfh[ht1], Wfx[xt], bf) \tag{1}$$

,where bf is a constant termed as bias.

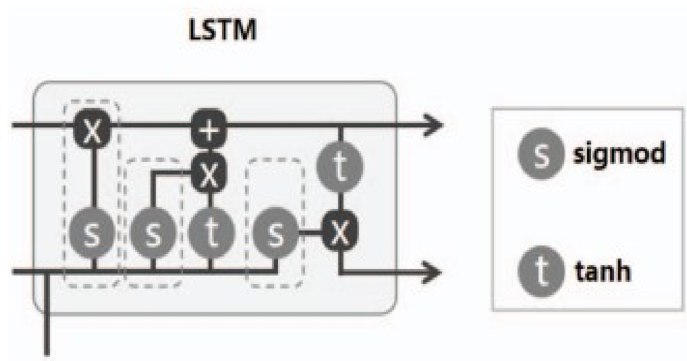

Figure 6: LSTM structure.

Input Gate: In the LSTM model, the input gate determines whether new information should be added to the LSTM memory. This gate consists of sigmoid and tanh layers. The sigmoid layer identifies which values need to be updated, while the tanh layer facilitates the generation of new candidate values. The output of the sigmoid layer indicates the update decision (it), while the output of the tanh layer (c˜t) represents the new candidate values. The outputs of these two layers are used in updating the LSTM memory.

The sigmoid layer is formulated as follows:

$$it = (W_{ih} \cdot h_{t-1} + W_{ix} \cdot x_t + b_i) \tag{2}$$

The tanh layer is formulated as follows:

$$\widetilde{c_t} = \tanh(W_{ch} \cdot h_{t-1} + W_{cx} \cdot x_t + b_c) \tag{3}$$

The combined formula for these layers is as follows:

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \widetilde{c_t} \tag{4}$$

Output Gate: This gate primarily utilizes a sigmoid layer to decide how the LSTM memory contributes to the output. It then applies a non-linear tanh function to map the values between -1 and 1. Finally, it obtains the result by multiplying it with the output of a sigmoid layer. The following equations are used to compute the output:

$$o_t = \sigma(W_{oh} \cdot h_{t-1} + W_{ox} \cdot x_t + b_o) \tag{5}$$

$$h_t = o_t \cdot \tanh(c_t) \tag{6}$$

4.7 GRU

Gated Recurrent Unit (GRU) is another type of recurrent neural network (RNN) similar to LSTM but with a simpler cell architecture. GRU consists of two gates: a reset gate and an update gate. The reset gate controls the flow of new input to the previous memory, while the update gate determines how much of the previous memory to retain.

The equations representing the single cell structure of GRU are as follows:

Reset Gate:

$$r_t = sigmoid(W_{xr}x_t + W_{hr}h_{t-1} + b_r) \tag{7}$$

Update Gate:

$$z_t = sigmoid(W_{xz}x_t + W_{hz}h_{t-1} + b_z) \tag{8}$$

Candidate Activation:

$$\tilde{h}t = tanh(Wxhx_t + W_{hh}(r_t \cdot h_{t-1}) + b_h) \tag{9}$$

Output:

$$h_t = z_t \cdot h_{t-1} + (1 - z_t) \cdot \tilde{h}_t \tag{10}$$

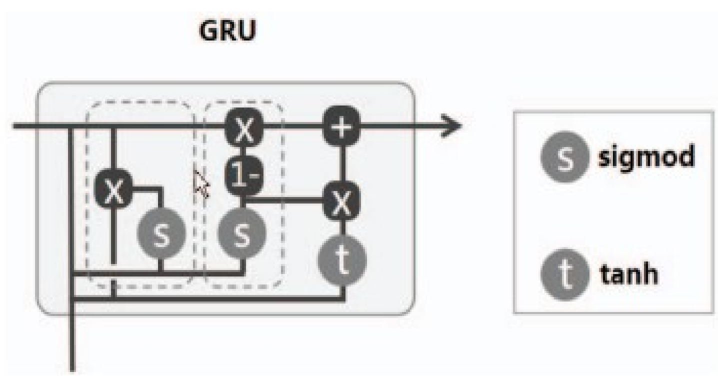

Figure 7: GRU structure.

Here, x_t, h_t, r_t and z_t represent the input vector, output vector, reset gate, and update gate respectively. All W variables denote weight matrices, and b are biases. The activation functions used are the sigmoid function ($sigmoid$) and the hyperbolic tangent ($tanh$) function.

These equations define the single cell structure and output of GRU. This architecture, designed similarly to LSTM, is suitable for modeling long-term dependencies.

5 Air-Smart Controller

Air-Smart Controller is a web interface developed using Python's Streamlit library. This interface allows users to view forecast data at ten-minute intervals. These forecast data are connected to the control level of the automation system and work integrated with SCADA systems.

Through this web interface, users can view data for the next 44 days at ten-minute intervals. Data is available for 5 different rooms. The data displayed includes temperature, relative humidity, and pressure information. Additionally, it shows which modes the automation system will operate in, synchronized with the data.

Users can request changes to the modes of the automation system. They can suspend the operation of the automation system or request it to operate in different modes. The manipulated data is evaluated by the automation system, and as a result of this evaluation, the request may be deemed valid or invalid.

Figure 8: Air-Smart Controller Data Visualisation.

6 Results

6.1 Metrics

MAE:

MAE measures the magnitude of the difference between actual and predicted values. It does not consider whether the predictions are higher or lower than the actual values. Instead, it measures the average amount of error across all predictions.

Formula:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|P_i - O_i| \tag{11}$$

P_i is the predicted value at index i,
O_i is the actual value at index i,
N is the number of data points.

MAE is straightforward to calculate and interpret.Compared to RMSE, MAE is less affected by outliers, making it a more robust metric. The absolute value function is not differentiable at zero, making it challenging to incorporate into optimization problems.

MSE:

MSE is the mean of the squares of errors. Squaring the errors gives more weight to larger errors.

Formula:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(P_i - O_i)^2 \tag{12}$$

P_i is the predicted value at index i,
O_i is the actual value at index i,
N is the number of data points.

Using MSE helps to deal with larger error magnitudes, thereby improving the generalization performance of the model. Additionally, its differentiability at every point extends its applicability to various domains. However, compared to MAE, MSE may become more sensitive to outliers.

RMSE:

RMSE is simply the square root of the MSE error metric. It measures the magnitude of the resulting error rates.

Formula:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(P_i - O_i)^2} \tag{13}$$

P_i is the predicted value at index i,
O_i is the actual value at index i,
N is the number of data points.

Similar to MSE, RMSE is sensitive to large error magnitudes. This characteristic is important to consider in certain applications. RMSE is sensitive to outliers. However, the combination of average error and variance in RMSE can complicate the application in some scenarios.

6.2 Results

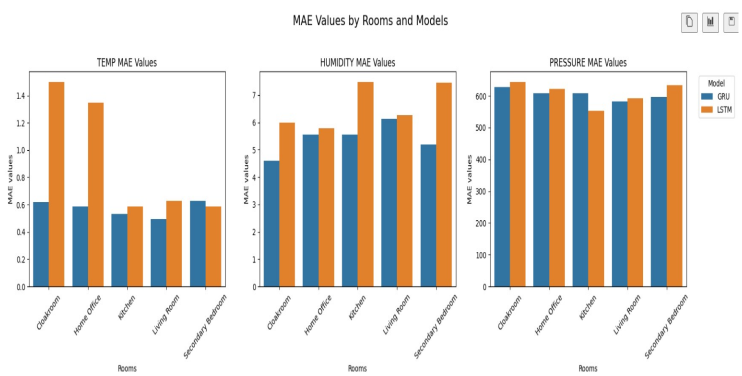

Figure 9: MAE results.

With these data, you can compare the performance of LSTM and GRU models based on the values of rooms and metrics (MAE, MSE, RMSE). MAE indicates how much the model's predicted values deviate from the actual values. The smaller the value, the better the model's performance. In temperature predictions, LSTM models generally have higher MAE values, indicating that GRU models perform better. A similar situation is observed in humidity predictions, where LSTM models typically have higher MAE values. In pressure predictions, LSTM models also show higher MAE values, while GRU models generally perform better.

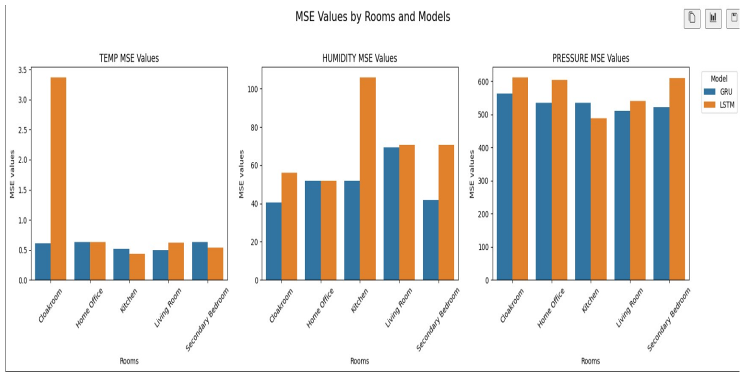

Figure 10: MSE results.

MSE provides the average of the squared errors and shows how sensitive the model is to large errors. The smaller the value, the better the model's performance. In temperature predictions, GRU models generally have lower MSE values, indicating better performance. In humidity predictions, LSTM models typically have higher MSE values, showing that GRU models perform better. Similarly, in pressure predictions, GRU models generally have lower MSE values, indicating better performance.

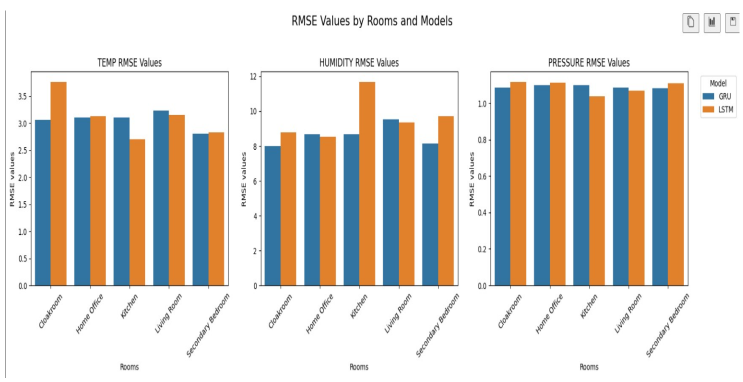

Figure 11: RMSE results.

RMSE is obtained by taking the square root of MSE and, like MSE, is sensitive to large errors. The smaller the value, the better the model's performance. In temperature predictions, GRU models generally have lower RMSE values, indicating better performance. In humidity predictions, LSTM models typically have higher RMSE values, showing that GRU models perform better. In pressure predictions, GRU models generally have lower RMSE values, indicating better performance.

Overall, the data suggest that GRU models perform better than LSTM models in predicting temperature, humidity, and pressure. Generally, lower values were obtained in all three metrics (MAE, MSE, RMSE). This indicates that GRU models work better than LSTM models on this dataset. However, other factors (e.g., training time, computational cost, etc.) should also be considered when selecting a model.

7 Conclusion

This study focuses on the utilization of home automation systems for predicting indoor air quality. Home automation systems measure real-time data such as temperature, humidity, pressure, CO2 levels, occupancy status, and window conditions. However, they may face challenges in data quality and responsiveness to future events. This research aims to predict indoor air quality using deep learning methods.

This study presents the performance of GRU and LSTM models on different datasets. It can be said that GRU generally outperforms LSTM in all three metrics (MAE, MSE, RMSE). Ignoring the metrics where LSTM performs exceptionally well, GRU demonstrates better generalization performance. However, when selecting a model, all scenarios should be taken into account.

Additionally, an interface called Air-Smart Controller has been developed. This interface allows users to view prediction data at ten-minute intervals and change modes of home automation systems.

In conclusion, this research highlights the potential of home automation systems in predicting indoor air quality and compares the performance of different neural network architectures like LSTM and GRU. The study could be a significant step towards the development of automation systems and contribute to better understanding the impact of home automation systems on air quality.

8 References

[1] Phillips TJ, Levin H. Indoor environmental quality research needs for low-energy homes. Sci Technol Built Environ 2015;21:80–90.

[2] Kumar P, Skouloudis AN, Bell M, Viana M, Carotta MC, Biskos G, et al. Real-time sensors for indoor air monitoring and challenges ahead in deploying them to urban buildings. Sci Total Environ 2016;560–561:150–9

[3] Almeida RMSF, Ramos NMM, de Freitas VP. Thermal comfort models and pupils' perception in free-running school buildings of a mild climate country. Energy Build 2016;111:64–75.

[4] IOM The Institute of Medicine. Climate change, the indoor environment, and health. Washington D.C.: The National Academies Press; 2011.

[5] Nazaroff WW. Exploring the consequences of climate change for indoor air quality. Environ Res Lett 2013;8.

[6] Marques G, Pitarma R. An indoor monitoring system for ambient assisted living based on internet of things architecture. Int J Environ Res Public Health 2016;13:1152.

[7] Barmpakos, D.; Kaltsas, G. A Review on Humidity, Temperature and Strain Printed Sensors—Current Trends and Future Perspectives. Sensors 2021, 21, 739. https://doi.org/10.3390/s21030739

[8] Stolojescu-Crisan, C.; Crisan, C.; Butunoi, B.-P. An IoT-Based Smart Home Automation System. Sensors 2021, 21, 3784. https://doi.org/10.3390/s21113784

[9] MinJeong Kim, B. SankaraRao, OnYu Kang, JeongTai Kim, ChangKyoo Yoo, "Monitoring and prediction of indoor air quality (IAQ) in subway or metro systems using season dependent models",Energy and Buildings,Volume 46,2012,Pages 48-55, ISSN 0378-7788, doi: 10.1016/j.enbuild.2011.10.047.

[10] Gaurav Priyadarshi, B. Kiran Naik,"Desiccant coated fin tube energy exchanger design optimization implementing KNN-ML tool and adsorption/desorption kinetics analysis using finite difference based transient model", International Journal of Thermal Sciences,Volume 192, Part B, 2023,108422,ISSN 1290-0729, doi: 10.1016/j.ijthermalsci.2023.108422.

[11] T. Akilan, K.M. Baalamurugan, "Automated weather forecasting and field monitoring using GRU-CNN model along with IoT to support precision agriculture", Expert Systems with Applications,Volume 249, Part A,2024,123468,ISSN 0957-4174, doi: 10.1016/j.eswa.2024.123468.

[12] Elnaggar, A., Said, M., Kraševec, I. et al. Risk analysis for preventive conservation of heritage collections in Mediterranean museums: case study of the museum of fine arts in Alexandria (Egypt). Herit Sci 12, 59 (2024). doi: 10.1186/s40494-024-01170-z

[13] A. Gururaj, V. Nagaraj, R. A S, A. K N, S. M U and A. D. B, "Air Quality Prediction and Analysis using Machine Learning," 2024 5th International Conference on Mobile Computing and Sustainable Informatics (ICMCSI), Lalitpur, Nepal, 2024, pp. 543-550, doi: 10.1109/ICMCSI61536.2024.00086.

[14] N.Kusuma Wardana, Suhaib A. Fahmy, J. W. G,"Low-cost SCADA/HMI with Tiny Machine Learning for Monitoring Indoor CO2 Concentration", https://warwick.ac.uk/fac/sci/eng/people/suhaibfahmy/publications/i2mtc2024-wardana.pdf

[15] Mishra, A., Gupta, Y. Comparative analysis of Air Quality Index prediction

using deep learning algorithms. Spat. Inf. Res. 32, 63–72 (2024). https://doi.org/10.1007/s41324-023-00541-1

[16] Marc-Fabian Körner, Dennis Bauer, Robert Keller, Martin Rösch, Andreas Schlereth, Peter Simon, Thomas Bauernhansl, Gilbert Fridgen, Gunther Reinhart, Extending the Automation Pyramid for Industrial Demand Response,Procedia CIRP, Volume 81, 2019,Pages 998-1003,ISSN 2212-8271,doi: 10.1016/j.procir.2019.03.241.

[17] D. Cortés, J. Ramírez, L. Villagómez, R. Batres, V. Vasquez-Lopez and A. Molina, "Digital Pyramid: an approach to relate industrial automation and digital twin concepts," 2020 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC), Cardiff, UK, 2020, pp. 1-7, doi: 10.1109/ICE/ITMC49519.2020.9198643.

[18] Profibus Profinet International, PROFINET System Description, First. Profibus Profinet International, 2014.

[19]W. He, J. Martinez, R. Padhi, L. Zhang and B. Ur, "When Smart Devices Are Stupid: Negative Experiences Using Home Smart Devices," 2019 IEEE Security and Privacy Workshops (SPW), San Francisco, CA, USA, 2019, pp. 150-155, doi: 10.1109/SPW.2019.00036.

[20] Wang, C., Li, X., Sun, W. et al. Occupant behavior, thermal environment, and appliance electricity use of a single-family apartment in China. Sci Data 11, 65 (2024). https://doi.org/10.1038/s41597-023-02891-9

[21] Ioffe, S. amp; Szegedy, C.. (2015). Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift.Proceedings of the 32nd International Conference on Machine Learning,Proceedings of Machine Learning Research 37:448-456 Available from https://proceedings.mlr.press/v37/ioffe15.html.