Developing a Reliable Shallow Supervised Learning for Thermal Comfort using Multiple ASHRAE Databases

Kanisius Karyono, Student Member, IEEE, Badr M. Abdullah, Alison J. Cotgrave, Ana Bras, and Jeff Cullen

***Abstract*— The artificial intelligence (AI) system faces the challenge of insufficient training datasets and the risk of an uncomfortable user experience during the data gathering and learning process. The unreliable training data leads to overfitting and poor system performance which will result in wasting operational energy. This work introduces a reliable data set for training the AI subsystem for thermal comfort. The most reliable current training data sets for thermal comfort are ASHRAE RP-884 and ASHRAE Global Thermal Comfort Database II, but the direct use of these data for learning will give a poor learning result of less than 60% accuracy. This paper presents the algorithm for data filtering and semantic data augmentation for the multiple ASHRAE databases for the supervised learning process. The result was verified with the visual psychrometric chart method that can check for overfitting and verified by developing the Internet of Things (IoT) control system for residential usage based on shallow supervised learning. The AI system was a Wide Artificial Neural Network (ANN) which is simple enough to be implemented in a local node. The filtering and semantic augmentation method can increase the accuracy to 96.1%. The control algorithm that was developed based on the comfort zone identification can increase the comfort acknowledgement by 6.06% leading to energy saving for comfort. This work can contribute to 717.2 thousand tonnes of CO2 equivalent per year which is beneficial for a more sustainable thermal comfort system and the development of a reinforced learning system for thermal comfort.**

***Impact Statement* — The current AI systems that are used in the thermal comfort system have progressively narrowed the acknowledgement of the thermal comfort zone with a negative impact on energy use. Less energy use is among the keys to achieving sustainable residential. The less energy will also contribute to fighting fuel poverty due to the current high energy price. This work introduced the use of the filtered and semantically augmented ASHRAE multiple databases for reliable supervised training data. The result was a wider acknowledgement of the thermal comfort zone by 6.06% with the capability to be deployed in the low-cost residential control system. The wider acknowledgement will contribute to less energy usage for comfort and a more efficient thermal regulation system. This reduction will support sustainable residential areas with a better indoor thermal conditioning system. This solution in general could contribute to the sustainable development goals.**

***Index Terms*— filtering algorithm, heating control systems, semantic data augmentation, supervised learning**

I. Introduction

THE decarbonising heat and buildings has become one of the vital commitments of many countries, including the United Kingdom (UK), as the response to the UN COP26 summit in Glasgow, in 2021. The UK is committed to reaching net-zero emissions by 2050 [1]. The support includes the incentive to decarbonise homes and funding new ideas and technologies. The government will also increase the effort to reduce dependence on burning natural gas in homes and introduce a Future Homes Standard by 2025 with low carbon heating and world-leading levels of energy efficiency [2]. This effort is significant because 86.9% of our time is spent indoors [3], even before the pandemic.

The heat pump is becoming more common to be installed to lower the carbon footprint for residential heating. The UK government will commit to deploying 600,000 heat pumps every year by 2028 [4]. However, implementing the heat pump is not a drop-in replacement for gas boilers [5]. Prior to the installations, an assessment should be done to optimise the heat pump usage and ensure efficiency.

If the heat pump is installed in poorly performed or leaky buildings, the efficiency will decrease. Efficiency

This paper was submitted on 27 September 2023 for review and resubmitted on 10 February 2024 after revision.

Kanisius Karyono is with the Electrical Engineering Department, Universitas Multimedia Nusantara, Scientia Garden Gading Serpong, Indonesia, 15810. E-mail: karyono@umn.ac.id, k.karyono@2019.ljmu.ac.uk

Badr M. Abdullah is with the School of Civil Engineering and Built Environment, Liverpool John Moores University, Byrom Street Campus, Liverpool, United Kingdom, L3 3AF. E-mail: B.M.Abdullah@ljmu.ac.uk

Alison J. Cotgrave is with the School of Civil Engineering and Built Environment, Liverpool John Moores University, Byrom Street Campus, Liverpool, United Kingdom, L3 3AF. E-mail: A.J.Cotgrave@ljmu.ac.uk

Ana Bras is with the School of Civil Engineering and Built Environment, Liverpool John Moores University, Byrom Street Campus, Liverpool, United Kingdom, L3 3AF. E-mail: A.M.ArmadaBras@ljmu.ac.uk

Jeff Cullen is with the School of Civil Engineering and Built Environment, Liverpool John Moores University, Byrom Street Campus, Liverpool, United Kingdom, L3 3AF. E-mail: J.D.Cullen@ljmu.ac.uk

improvements can be made, for example, by adding internal or external insulation and improvements to fabric airtightness [6]. In the UK, 53.36% of houses were built prior to 1964' which usually have poor performance in heating compared to modern houses, which are better insulated [7].

Using the other types of electric heating with intelligent control will be crucial to fill in the gap while still maintaining the energy efficiency target. Infrared radiant heaters for a specific heating area, a ceramic heaters to accommodate off-peak loads or other kinds of heaters that give an additional benefit are becoming crucial. Electric heating will be among the technologies considered to be carbon-friendly. This trend highlights the importance of continuous research on thermal comfort even before 1920 [8].

There were two significant research in thermal comfort, the human physiology approach and the human psychology/ adaptive behavioural approach [8]. Each approach has benefits and limitations. There has been much research to develop each approach until now, but combining the two approaches is not an easy job. Due to the advancement in artificial intelligence (AI) technology, combining both approaches are now open.

This work contributes to the thermal comfort advancement by elaborating the human physiology and adaptive approach which can:

- harvest the benefit of energy conservation whilst maintaining human comfort.
- introduce the base component for electric heating control that can acknowledge adaptive /psychological aspects. This was because of the multiple behaviours that faced by thermal comfort controls [9].
- uses the physiology measurement data from the multiple ASHRAE datasets to train the adaptive model to accommodate the human behavioural factors. This is to answer the gap of the need for larger datasets to achieve higher performance [10], [11].
- set the baseline dataset to support the reinforced learning by using multiple ASHRAE databases with proposed inconsistency filtering and semantic augmentation.
- proposes validation methods to check the learning process in the AI system for thermal comfort based on the psychrometric chart analysis. This validation is crucial to avoid overfitting problems.
- use simple artificial neural networks (ANN) to fit with a residential controller that has computational limitations and can be deployed as a part of the Internet of Things (IoT) based system.
- minimise the need to use the explainable AI, which consumes more resources by using a more transparent weighting process. Auburn's Nguyen even mentions that the field of explainability is getting somewhat stuck [12]. On the other hand, European legal requirements also mandate transparency in the use of automated or AI systems [13].
- has the ability to acquire a wider comfort zone that will impact lowering energy for comfort. This is to answer the gap in the comfort zone that has been progressively narrower over the past several decades [14].

The paper is presented by focusing on the research motivation, novelty, and contributions in the introduction section, followed by the literature review and previous research in the background section. This section also highlights the possibility of acknowledging wider comfort to reduce energy consumption and the reason for choosing the data set along with the process of inconsistency filtering and semantic augmentation. The next chapter is the methodology for this work followed by the research and discussion. The conclusion of this work is presented in the last part of the paper.

II. Background

A. Thermal Comfort

Thermal comfort is a state of mind expressing satisfactory adaptation to the immediate thermal environment. Fanger set up the milestone model for thermal comfort in 1970 [15]. This work was based on the measurement of human physiology and uses the climate chamber for the measurement. In this chamber, the indoor conditions can be simulated and controlled. This model is formulated with the relations among the affecting parameters (1). The approach then expanded with the Predicted Mean Vote (PMV)/Predicted Percentage of Dissatisfied (PPD). This method becomes a standard reference work on thermal comfort, the basis of the ISO 7730–2005 [16] and acknowledged by the ASHRAE-55 Standard [17].

$$M - W = C + R + E + (Cres + Eres) + S \qquad (1)$$

Where: M: the metabolic rate of the occupant
W: mechanical work done by the occupant
C: convective heat loss from the clothed body
R: radiative heat loss from the clothed body
E: evaporative heat loss from the clothed body
Cres: convective heat loss from respiration
Eres: evaporative heat loss from respiration
S: the rate at which heat is stored in the body tissues

Since thermal comfort is about a state of mind and Fanger's trial was done in the chamber with limited subjects, it could not represent the different variations in the subject preferences, especially the person with special needs and the difference in the dwelling types. Nicol and Humphreys propose the adaptive method to overcome the problem [18]. The adaptation can be physiological, which is related to the body's adaptation to the temperature change, psychological which is formed by previous human experiences and behaviour-related adaptation [19]. Comfort is achieved if people have sufficient opportunities to adapt [20]. The ASHRAE-55 Standard has also acknowledged this adaptive method [21].

Deploying the adaptive thermal comfort models is often done through the black box approach [22]. This deployment makes the model incompatible and cannot be precisely calculated as in the thermal physiology models. Complete thermal adaptation also requires a more extended period. If the training data is taken directly from the occupants during the system implementation, there will not be enough data to cover the whole extreme condition, putting the user in an uncomfortable situation and triggering the user's scepticism. This user experience can negatively impact the system because comfort is about the state of mind [23].

B. The Possibility of Widening the Comfort Zone

Thermal comfort plays an essential role in the design process of the indoor artificial climate due to its significant impact on health and safety. Productivity is also affected by the thermal condition [24]. The comfort zone can be predicted using the

PMV-PPD method [21].

The PMV-PPD comfort zone is included in the ASHRAE-55 Standard [17]. Besides PMV-PDD, the comfort zone is also defined by Givoni [25]. The PMV-PPD is prescriptive, based on the measurement in the thermal chamber. Due to the nature that comfort is based on satisfactory human adaptation, there is the potential to widen the comfort zone based on the adaptive approach, which is a goal-based method.

The specific group of people like young, elderly, disabled, or temporarily ill people can have a different comfort range than defined in the PMV-PPD method. For example, the elder people group tend to have higher comfort temperatures [26], [27]. On the other hand, there can be a request for a lower comfort temperature due to the health recommendation. It has been proven that frequent cold exposures can help people have less body fat [28]. Our previous research also shows that the lower temperature setpoint can lower the carbon footprint while maintaining the indoor humidity within the healthy range [29]. Regular exposure to cold acclimation will improve the subjective responses to cold [30]. In the long periods, this might alter the thermal preferences of the occupants.

The comfort zones are shown in Fig. 1. The boundary presents the prescriptive approach, and the number represents the adaptive potential comfort zone. The comfort zone of PMV-PPD is presented as data1 for activity value of 1.0<met<1.3 and clothing value of 1.0 Clo. Data 2 presents the PMV-PPD comfort zone for 1.0<met<1.3 and 0.5 Clo. Data 3-6 represent Givoni's comfort zone. Data 3 represents the still air condition for winter, data4 for still air condition for summer, Data 5 for airspeed of about two m/s in winter and data6 for summer as seen in Fig.1. Our proposed system aims to gain a wider comfort zone that has been identified as progressively narrower over decades [14] to conserve the energy for comfort.

III. METHODOLOGY

This work focused on the shallow learning AI for controlling, for example, the electric radiant panel to be deployed as part of the Internet of Things (IoT) system for residential dwellings. This work is focused on the three TSV values or thermal preferences (no change, warmer and cooler). The diagram that shows the methodology of this work is shown in Fig. 2.

The multiple ASHRAE databases are used for the training

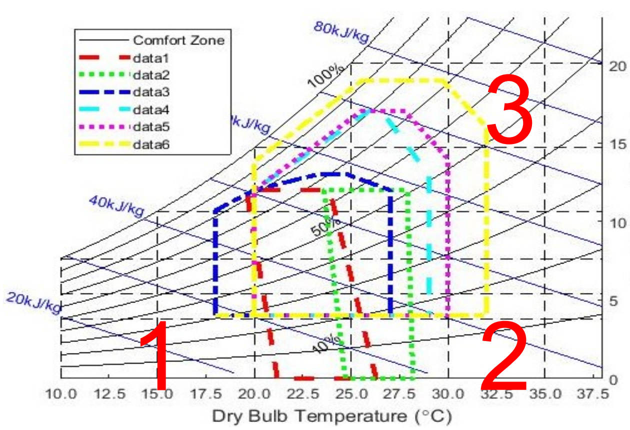


Fig. 1. Psychometrics chart showing the comfort zone [17], [25] and the potential comfort zone [8].

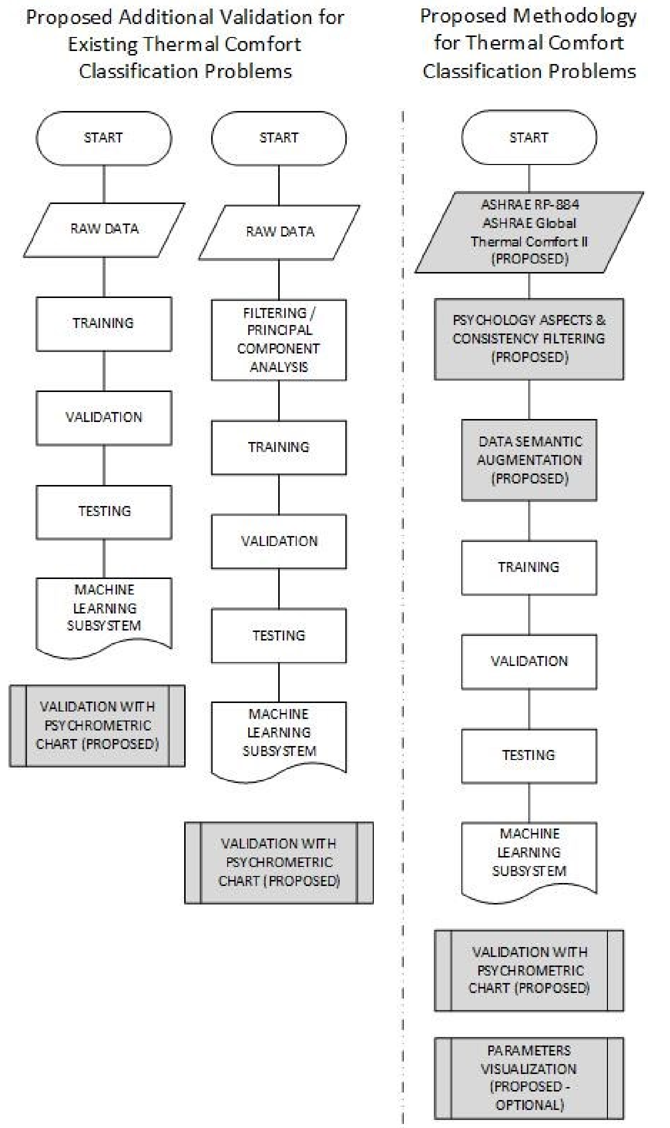


Fig. 2. Proposed validation methods and proposed methodology for thermal comfort AI training

data source, and the filtering process is implied to maintain the data consistency without eliminating the psychological aspect variations to the data. The semantic augmentation process is added to the data to balance the feature and enrich supervision learning. This work will implement four and five parameters from six dominant parameters due to the availability of IoT sensors.

This work proposed a method to visually check the AI learning result using a psychrometric chart. Visual result validation with the psychrometric chart can be done with a predefined input range to map the thermal comfort zone and identify the learning result performance. The notions of human psychological/behavioural aspects can also be shown by altering the input parameters and identifying the impact visually via the chart.

A. The Multiple ASHRAE Databases

This work uses the most comprehensive thermal comfort dataset, ASHRAE RP-884 and ASHRAE Global Thermal Comfort Database II. The reason for choosing these databases:

- available online as open-source databases

- the most comprehensive database in thermal comfort; The ASHRAE RP-884 consists of 25,616 entries, and the ASHRAE Global Thermal Comfort Database II includes 81,967 entries [31].
- consists of objective indoor climatic observations and subjective questionnaire-based evaluations.
- the data were taken from the field experiments, with the people doing their activities. Even the data captured the PMV-PPD values; these values differed from the Fanger experiments done in a controlled indoor environment of a climate chamber [32].
- consists of different cities and countries, seasons, climate zones, building types, cooling and heating strategy
- include personal information about the subjects such as sex, age, height, and weight. Other subjective essential factors are thermal sensation, acceptability, and preference. These subjective factors are taken with the specific metabolic rate (met) and clothing insulation level (clo). The comfort indices such as PMV, PPD and Standard Effective Temperature (SET) were calculated uniformly and included in the database.
- the parameter measurements included in this dataset are various types of temperatures, air velocity, relative humidity, and monthly average temperatures. Some indoor environmental controls include blinds, fans, operable windows, doors, and heaters [32].

This dataset has developed many approaches to predict the Thermal Sensation (PTS) [33] regarding the location. Another recent model that is proposed based on this database is done to assess the PMV-PPD accuracy against the database [34]. This work concludes that accuracy varied enormously between ventilation, building types and climate. The authors have seen this gap to propose better models using the power of AI.

B. Elaborating ASHRAE Data for Supervised Learning

In supervised learning, the role of the training data is crucial. Getting the field data during the system initialisation is not practical because the amount and variety of data will not be adequate. It requires the proper instruments and consent from the occupants. This approach will not give a comfortable experience for the user. Doing their own data collection will require calibrated sensor equipment, the subject consent and awareness of the thermal sensation and the subject questionnaire. The data collection also should be done in the different environmental conditions and building types to get a broad combination of training data. Controlling all these parameters is challenging in field studies. On the contrary, measuring these parameters in the climate chamber will be easier but not represent the building types and the actual occupants' conditions. There is also the approach to using the simulated or generated data, but the validation of the data can trigger another hurdle.

Previous studies have shown that the AI approach using the ASHRAE RRP-884 has limited diversity and unbalanced distribution. This unbalanced distribution results in a model which is not reliable in extreme conditions [10]. This work uses 20,954 data entries from 25,616 entries after data cleansing. The other research uses ASHRAE Global Thermal Comfort Database II, which consists of a more significant amount of data [11]. The other works use the combination of both ASHRAE RP-884 and ASHRAE Global Thermal Comfort Database II [35], [36]. However, in the previous research, the data items were selected to represent each class or label. This selection is because the more extensive data does not guarantee higher accuracy and can cause overfitting issues [11].

Not all data from this dataset can be directly elaborated in the training data. This limitation is due to the nature of the human psychological factor that the thermal comfort is personal or individual [35]. The ambiguity of the data inconsistency can be high. The previous work considers this data illogical and an anomaly [36]. From the 107,583 entries, this work only uses 16,795 based on four thermal metrics. The work by Luo also populates only 10,618 entries out of 81,967 entries [11]. This work uses the thermal sensation vote (TSV) for the label of the learning target. Eighty percent of the data are allocated for training and 20% for testing. This work also mentions that allocating more data percentages for testing can decrease accuracy, indicating the lack of data available for training.

The work by Luo discovered that even with 66,3% maximum accuracy using the Radom Forest, the approach already got 10–20% higher prediction accuracy than the PMV model. The model also got 60–66% for 3-point TSV accuracy and 52–57% for 7-point TSV prediction [11]. This work also introduced the six most influencing variables: temperature, relative humidity, clothing insulation, airflow speed, subject age, and activity/metabolism level. The work from Wang also has a similar result. The scalability and sample number are mentioned as limitations, although the accuracy can be increased to 87% [36]. The accuracy in predicting thermal acceptability is also higher than thermal preference.

Learning from the previous work based on the ASHRAE Database, this paper proposes the use of the database which are:

- based on the TSV for the learning target, where three labels for the thermal conditions are used (no change, prefer warmer and prefer cooler)
- uses as much entry data as possible and does not pick pointing data to represent as many individual variations or preferences as possible
- compare the use of the four and five most significant variables, which are temperature, relative humidity, clothing insulation, activity/metabolism level, and subject age
- uses simple filtering to minimise the ambiguity of data by considering the psychological aspect of human
- fills the data conditioning gap to prepare the data to become the training data for the learning process.

C. Data Filtering

Previous works that use the ASHRAE database did not use the complete entries, but selected entries based on each class to achieve the balanced features. The data used for training was only less than 20% of the whole ASHRAE database. This selection makes the human psychological aspects not easily captured by the supervised training. On the other hand, the risk of overfitting is also implied in using this database. Since it was published, few AI developers have been willing to use this data to develop their AI systems.

This work proposes simple yet powerful methods to filter the data based on human perception consistency. The need for filtering is because the data was based on precise measurement, but the human perception data was based on the questionnaire that was more prone to error and subjective judgment than the measured data. This filter worked based on the comparison of parameters and omitted the data that was considered to be inconsistent as follows:

1. The filtering is based on the consistency between the thermal acceptability (0 unaccepted 1 accepted) and thermal preference (warmer, no change and cooler). The warmer and cooler should have a thermal acceptability value of 0, no change should have the value of 1.
2. The filtering is based on the consistency between thermal acceptability (0 unaccepted 1 accepted) and thermal sensation (-3 to +3). The value between -2 to 2 should have a thermal acceptability value of 1, and the other should have 0.
3. The filtering is based on the consistency between the thermal sensation (-3 to +3) and thermal preference (warmer, no change and cooler). The thermal sensation value less than -2 should correspond with warmer, and more than 2 should associate with cooler. The state of no change should have a value between -2 to +2.
4. The filtering is based on the consistency between the thermal preference (warmer, no change and cooler) and thermal comfort (1-very uncomfortable, 6-very comfortable). The value 1-very uncomfortable and should not have the value of no change. The value of 6-very comfortable should not have the value of warmer or cooler.
5. The filtering is based on the consistency between the thermal sensation (-3 to +3) and thermal comfort (1-very uncomfortable, 6-very comfortable). The 1- very uncomfortable value should not have the value of -2 to +2. The thermal comfort value of 6-very comfortable should not have a thermal sensation value below -2 or more than 2.

The target labels for the AI are based on the three states of TSV, -1 means that the occupants need a warmer indoor environment, 1 means that the occupants need a cooler temperature and 0 means that the occupants are satisfied with the indoor temperature. This approach is the most straightforward arrangement for the subject because they are still comfortable in the current temperature, need a lower temperature, or need a warmer temperature.

Not many people can define their thermal preferences using seven scale levels. There is no crisp border between each scale, and even the same temperature can be mapped into the different seven-scale TSVs. The border between these three scales is not crisp either. However, this map will better understand people's thermal sensations due to its simplicity.

Previous research shows that there are six dominant parameters, and compared to the complete twelve parameters, it only increases 2.6% in the performance compared to elaborating six dominant parameters [11]. The IoT low-cost sensors can detect two of six dominant parameters: temperature and relative humidity. The occupant data entries can introduce the clothing insulation, metabolic rate or activities, and age. Low-cost sensors do not easily detect air velocity. This work also tried to narrow the parameters into five parameters for easier deployment with a residential IoT system. The precise air velocity sensor and the sensor placement will not be feasible for the residential IoT system. This work identifies that the simplification is acceptable. Further simplification on the occupant's age was also assessed. The simple algorithm for filtering is as follows:

Algorithm 1: Simple Data Filtering for the ASHRAE Database

```
//% simple filtering based on five inconsistency check
// entry will be marked as 0 (M=0) to be filtered and M=1 to retain
// the field with "NA" entries will be skipped
Input: ASHRAE database
for ctr=1 to size (ASHRAE database) do
   //%ThermalAcceptability (TA) vs ThermalPreference (TP)
   If (TA=1)*(TP="no change") then M=1
      ElseIf (TA=0)*(TP="cooler") then M=1
         ElseIf (TA=0)*(TP="warmer") then M=1
            ElseIf (TA="NA")OR(TP="NA") then "NA" Else M=0
  //%ThermalSensation (TS) vs ThermalAcceptability (TA)
  If (TA=1)*(ABS(TS)>2 then M=0
      ElseIf (TA=0)*(ABS (TS)<=1) then M=0
         ElseIf (TA="NA")OR(TS="NA") then "NA" Else M=1
   //%ThermalSensation (TS) vs ThermalPreference (TP)
   If (TP="warmer")*(TS<-2) then M=1
      ElseIf (TP="cooler") *(TS>2) then M=1
         ElseIf (TP="no change")*(ABS(TS)<=1) then M=1
            ElseIf (TS>1)* ((TP="cooler")OR(TP="no
               change"))*(TS<2) then M=1
              ElseIf (TS<-1)*((TP="warmer")OR (TP="no change"))
                 *(TS>- 2) then M=1
                ElseIf ((TS="NA")OR(TP="NA")) then "NA" else
                 M=0
  //%ThermalComfort (TC) vs ThermalPreference (TP)
  If (TC=1)*(TP= "no change") then M=0
     ElseIf (TC=6)*((TP="cooler")OR(TP="warmer" )) then M=0
         ElseIf ((TP="NA")OR(TC="NA")) then "NA" Else M=1
  //%ThermalComfort (TC) vs ThermalSensation (TS)
  If (TC=1)*(ABS(TS)<=2) then M=0
     ElseIf (TC=6)*(ABS(TS)>2) then M=0
        ElseIf ((TC="NA")OR(TS="NA")) then "NA" Else M=1
end for
Output: Marked ASHRAE database
```

D. Data Semantic Augmentation

Most of the entries in the ASHRAE database fall under the "no change" label (43,441 entries). Only about 14,966 entries need a warmer temperature, and about 27,093 entries need a cooler indoor temperature. This case is similar to the image processing and classification problem with highly imbalanced data. The training data for supervised methods are usually difficult to collect due to the costly human efforts and particular domain expertise. A data augmentation strategy is introduced to balance the feature space to enrich the supervision. The augmentation strategy can normalise the supervision process to improve the robustness by embedding such that the features of the same instance under different augmentations should be invariant and forcefully separated from the other instance features [37].

The previous work shows that data augmentation can be more powerful in the image classification problem if the class identity is preserved, for example, with semantic transformations. Each class in the training set can be added with the samples from the generator. The procedure is computationally intensive and lengthens the training procedure. The training set data can be effectively augmented by searching the semantic directions.

The random directions that may result in the meaningless transformation can be reduced [38].

This work aims to develop data augmentation using the approach of semantic data augmentation. The class "no change" remains untouched while the "warmer" and "cooler" classes are added with the data in the semantic direction of the value that is not covered by the ASHRAE database. The "warmer" class is augmented with the lower temperature value under the value of mapped ASHRAE data. On the contrary, the "cooler" class is augmented with the data, which is higher than the mapped ASHRAE data. This method helps to balance the feature space to enrich the supervision. The benefit of this method is that the data obtained from the ASHRAE database is unaffected due to the non-overlapped semantic augmentation direction. In this case, the data related to the psychological aspects are still maintained, and the essence of using the ASHRAE database is sustained.

The base for semantic augmentation is the temperature data. This data is chosen because the class that needs the augmentations are "warmer" and "cooler" classes. Fig. 3 shows the map of both classes, the augmentation and the data map. The class "no change" remains untouched due to the adequate data. This method also retains the psychological aspects and the accurate measurements in the ASHRAE database.

The augmentation data range was decided based on the notion that the augmentation data will not change the original data obtained from the ASHRAE database. The “warmer” class is augmented with the lower temperature value under the value of mapped ASHRAE data, which is 10 degrees Celcius. The “cooler” class is augmented with the data higher than 40 degrees Celcius. It is shown that the semantic augmentation direction is non-overlapped. The essence of the ASHRAE database is sustained. It is shown in Fig. 1. that this range of the augmented data is also outside the comfort zone. The algorithm for generating the semantic augmentation is as follows:

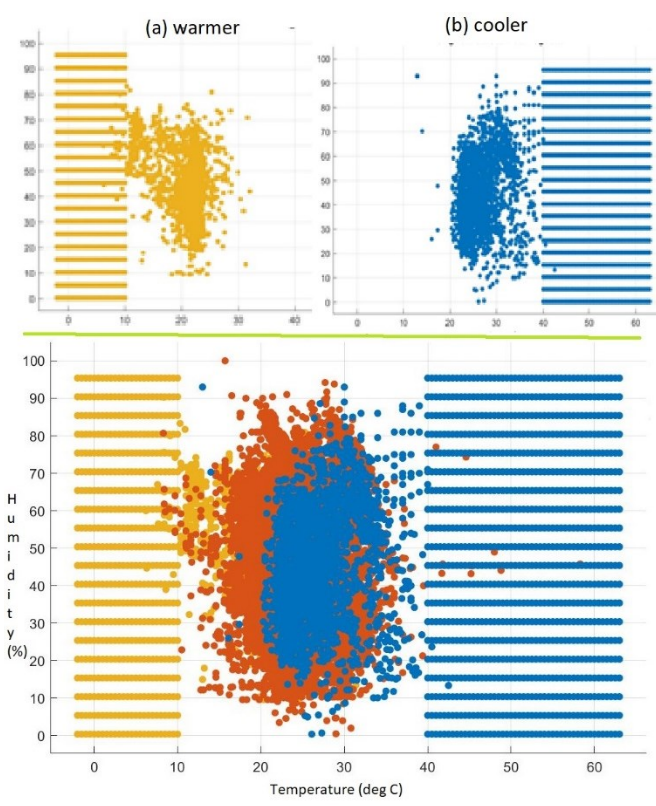


Fig. 3. Database map after the filtering process and semantic augmentation, (a) warmer class (b) cooler class

Algorithm 2: Semantic Augmentation Data

```
//%”for colder augmentation class”
Input: data intervals
row=0;
for clo=0 to 2.89 step clo_intervals do
   for met=0.65 to 6.83 step met_intervals do
      for tem=40 to 63.2 step tem_intervals do
         for RH=0.4 to 100 step RH_intervals do
            for age= 6 to 99 step age_intervals do
               row=row+1;
               AugMat(row,1)=clo;
               AugMat(row,2)=met;
               AugMat(row,3)=tem;
               AugMat(row,4)=RH;
               AugMat(row,5)=age;
               AugMat(row,6)=”colder”;
            end for
         end for
      end for
end for
Output: Augmentation Matrix: AugMat(row,[1:6])
//%“for warmer augmentation class”
// similar with warmer except this line:
//       for tem=0 to 10 step tem_intervals do
//
//               AugMat(row,6)=”warmer”;
// temp can be expanded for more extreme temperature
```

E. Psychrometric Based Verification and Parameter Visualization

Testing and validation for supervision learning are usually done using the data fraction randomly populated. The typical value for testing and validation can be 10% to 20%. A higher percentage can lead to an accuracy decrease [11]. However, more extensive data does not guarantee higher accuracy and can cause overfitting issues.

Checking the learning against overfitting issues is not easy. This work proposes using psychrometric chart mapping to validate the supervised learning result. This method is based on the comfort zone map in the psychrometric chart. The overfitting results will lead to the map not showing the correct pattern.

The pattern is generated using the validation data generated with the value range of relative humidity from 0% to 100% and temperature from 10 °C to 40°C. Other parameters like age, clothing insulation and activity can be predefined with median values. The result can be mapped with different colours or symbols to represent the output.

These symbols represented the class that needs warmer temperatures, the class that needs a cooler temperature, and the comfort zone (no temperature change). In the case of overfitting, the pattern generated will show the pattern very differently compared to the comfort zone shown in Fig.1.

This verification process method can also be used to compare the effects of the parameter change. One parameter value can be altered while the other parameters are constant. The impact of each parameter on the comfort zone can be captured and simulated. This method can simplify the representation of the multidimensional parameters that impact thermal comfort.

IV. Result and Discussion

The aim of data filtering is to use the data entries from the ASHRAE database as much as possible by removing inconsistent data while still capturing the psychological aspects of human comfort that are registered in the database. This filtered data can be used as base training data so that the AI developer does not have to capture their data which needs much effort or will decrease the occupants' comfort. The user can then override the settings of the system with their personal preferences. Their personal preferences can be entered in the later stage of the system development.

The ASHRAE RP-884 database has 25,616 entries, and the ASHRAE thermal comfort database II (1995 – 2015) has 81,967 entries. After the filtering process, the amount of data in ASHRAE RRP-884 is 14,970, with filtered entries of 10,646 or 41.56%. The ASHRAE Thermal Comfort Database II entries are 50,286 after the filtering process, with a filtered value of 31,681 entries or 38.65%. In total the ASHRAE database after filtering is 65,256 entries or 60.66% (filtered value is 42,327 entries or 39.34%). This entry has at least three times as much data as the previous work. More elaborated data means that the system can better capture the occupants' variations. The risk of overfitting can be eliminated with the further processing of the data.

The database that has been filtered is mapped and compared with the original ASHRAE database. Fig. 4 shows the mapping with the temperature as the x-axis and relative humidity as the y-axis. The original database map is shown on the left side, whereas the filtered database is shown on the right side. This figure shows the data mapping based on the three TSV class groups, which are "no change", "warmer", and "cooler". The ASHRAE database is shown to have major overlaps between classes. This significant overlap is the cause of the difficulties in training the AI using this database. It is challenging to have a proper classification process with the risk of overfitting.

The class overlap is reduced with the filtered database shown on the right side. This method gives the learning process a better chance to generate better training results for proper classification. The mapping position between the "warmer" class and the "cooler" class looks physiologically better than the original database.

This map also shows a gap in the data availability for the "warmer" class and "cooler" class. Not only to make the feature space to be equal. Furthermore, the system needs a different range of data to be registered in the database for the "warmer" and "cooler" classes. It needs more data outside the comfort temperature zone for a better learning process. The answer to this problem is semantic augmentation. The result of this method will be covered in subchapter 4.C.

A. Excluding Air Movement to Accommodate the IoT Systems and Excluding the Age

One of the six crucial parameters in thermal comfort is air movement. However, this parameter cannot be easily obtained from the IoT sensor system. This work deploys the system with five parameters and four parameters without the age data of the occupant. This case study is for simulating in case the age information of the occupants is not available. This combination of the original and filtered ASHRAE database was then used for the training data for 29 well-known AI algorithms for classification. The accuracy result for each database and method can be seen in the **Appendix**. The average of the accuracy results is given in Table 1.

The result shows that the data filtering increases the accuracy of the training results. The proposed filtering methods can make the ASHRAE database better for each RP884 database, Thermal Comfort Database II and the combination of both databases. The accuracy increases for all methods of training. Besides this data filtering, data normalization is already included in the AI process to gain better results. The result also shows that reducing the parameter (age parameter) can maintain accuracy. This result can be caused by the overfitting or the unbalance of the feature space. The semantic data augmentation

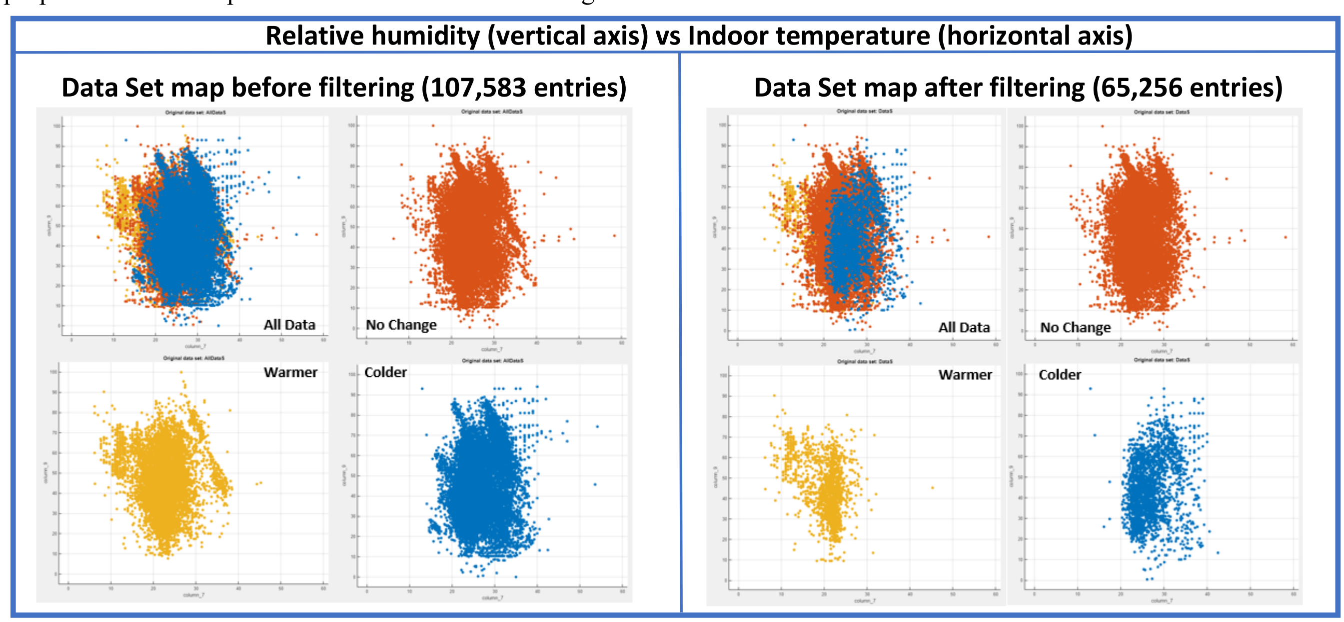


Fig. 4. The ASHRAE Database Mapping for Relative Humidity vs Indoor Temperature Before Filtering and After Filtering

will give better accuracy results for this problem. This semantic augmentation performance will be discussed along with the psychrometric chart in subchapter 4.B.

TABLE 1
THE AVERAGE ACCURACY FROM 29 CLASSIFICATION METHODS

Database	Average (%)
DB1	51.87
Filtered DB1	76.09
DB1 without age	55.05
Filtered DB1 without age	80.40
DB2	42.00
Filtered DB2	70.23
DB2 without age	50.51
Filtered DB2 without age	81.57
DB1 and DB2	43.69
Filtered DB1 and DB2	74.90
DB1 and DB2 without age	50.53
Filtered DB1 and DB2 without age	81.52
Augmented & Filtered DB1 and DB2	86.43
Augmented & Filtered DB1 and DB2 without age	90.12

Note: DB1 = ASHRAE RP-884 database
DB2 = ASHRAE Global Thermal Comfort Database II

B. Validation Using Psychrometric Chart, and The Result with Data Filtering and Semantic Augmentation

Many of the previous AI works were the area near the edge of the comfort zone or the borderline of the predefined zone that was considered as uncomfortable. These areas should be further explored because in some cases, the occupants can still be in comfort. The ability of the system to acknowledge this area will conserve energy. The intelligent system can have the recommender function to lower the energy cost by informing the occupants about clothing or activities that can keep them comfortable but with less energy spending. The occupants still have the probability of staying comfortable with higher clothing insulation during winter to conserve heating energy. On the contrary, the occupants will also have the probability to be comfortable in the hot summer by wearing lighter clothes, using the fan, and consuming fresh beverages to conserve cooling energy. This behaviour is the current gap in the previous work.

This work accommodates these needs by proposing the validation process using the psychrometric chart and test data generator. The test data generator works similarly to **algorithm 2** but with the parameter range to be more specific on the comfort zone map. The temperature can be between 10°C to 40°C, with the relative humidity value between 0 to 100%. The generated data then being fed to the intelligent system, and the result is drawn in the psychrometric chart. Each label can be drawn in the chart with a different colour to show the “no change” class, “warmer” class, and "cooler" class.

The previous work mapped the training result with the psychrometric chart but without the test data generation. This method makes the validation only limited to the available data, and the comfort zone cannot be adequately mapped. The comfort zone can be appropriately mapped using the generated data, with the edges of the thermal comfort zone visually presented.

In order to compare the learning result and the effect of filtering and the semantic augmentation method, the neural network algorithm was used in this work. The neural network structure was used for the training with the original ASHRAE data, the filtered data, and the filtered semantic augmented data. The wide neural network used has one hidden layer with a 100-node layer size. The learning dataset composition for this Neural Network training was 70% training. 15% validation and 15% testing data. The data selection for these groups was based on random selection.

The parameters involved in this trial consisted of indoor temperature, relative humidity, clothing insulation, activity/metabolism, and age. With an accuracy of 45.6%, the AI system is then fed with the generated test data. For this test, the parameters involved were the combination of indoor temperature, relative humidity, clothing insulation, and activity/metabolism. The age parameter is a single value, taken randomly to reduce the training data size. The different age parameter values will be covered in the following subchapter, 4.C. The class result is drawn in the psychrometric chart, as shown in Fig 5 (a). This result shows an incomplete comfort zone. Only class "no change" and "cooler" classes dominate, and the comfort zone is not drawn correctly. The class “no change” is represented with green colour, “cooler” with red colour and “warmer” with blue colour. The comfort zone range is from 10°C up to 30 °C, which is not valid compared with the PMV-PPD results.

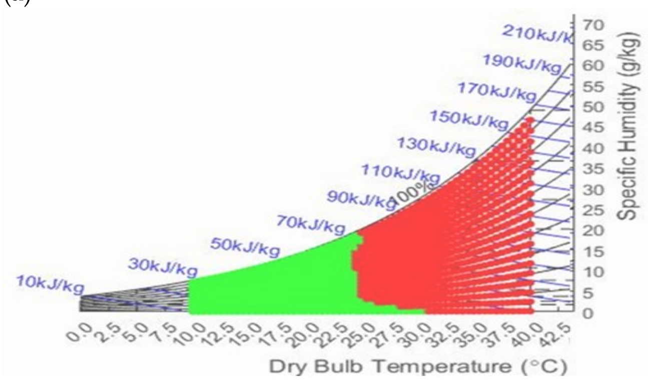
Figure 5 (b) shows the psychrometric map of the generated test data for the system, which was trained using the filtered ASHRAE database. The parameters and dataset grouping method were similar to the trial for Fig. 5 (a). The accuracy of this training was 55.5%. All three classes are visible, but the comfort zone is still not drawn correctly. The class "warmer" only covers a small portion of the chart, and the comfort zone range spans 10°C up to more than 40 °C, which is not valid if compared with the PMV-PPD results. This problem shows that the system needs semantic augmentation to generate the correct result.

The training data with semantic augmentation was deployed with the accuracy of this training data about 98%. The parameters and the data grouping were similar to the trial for Fig. 5 (a) and Fig. 5 (b). The psychrometric map generated for the semantic augmentation filtered ASHRAE data is shown in Fig 5 (c). The comfort zone is better represented in this result. The comfort zone ranges from 17.5°C to about 29 °C. The result is better than the two mapping before, shown in Fig. 5 (a) and Fig. 5 (b). This result represents the comfort zone presented in Fig.1.

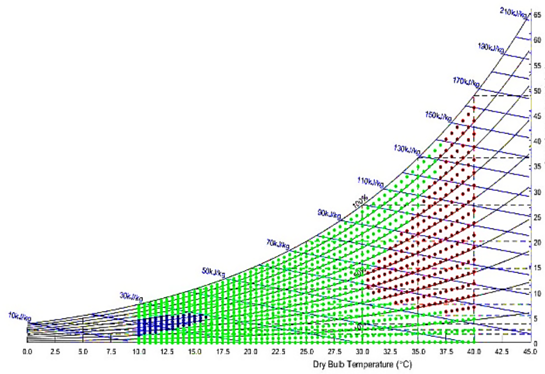
The reason related to the selection of a wide neural network was because of the performance shown in the **Appendix**. The multi-layer feed-forward fully connected neural network with 100 nodes was used with the backpropagation learning algorithm (BP) to update the weight of the neurons and bias. The Ensemble Bagged Trees gave a learning accuracy of 97 %, the Ensemble Subspace K-Nearest Neighbors (KNN) gave the highest performance with 98.2% accuracy and the Wide Neural Network resulted in **96.1%** accuracy. With the slightly lower performance, the Wide Neural Network was chosen over the other methods due to the implementation of the method which is the residential indoor thermal comfort.

The indoor local controller has limitations in memory space and computational power. Since the algorithm should be able to run in the controller node, the Wide Neural Network which

(a)



(b)



(c)

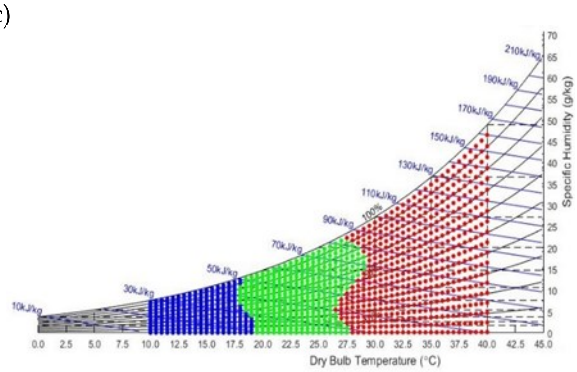


Fig. 5. Psychrometric Mapping for the Comfort Zone Trained with (a) the Original ASHRAE Database (b) the Filtered ASHRAE Database (c) the Filtered ASHRAE Database with the Data Semantic Augmentationentation

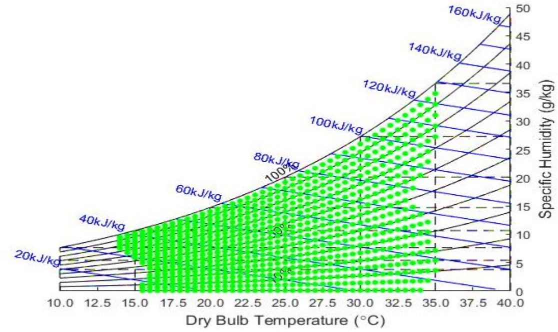
consists of the input layer, one hidden layer and an output layer was chosen due to the benefit of less computationally intensive and less memory usage with better interpretability of the model.

C. *Parametric Visualisation in the Psychrometric Chart*

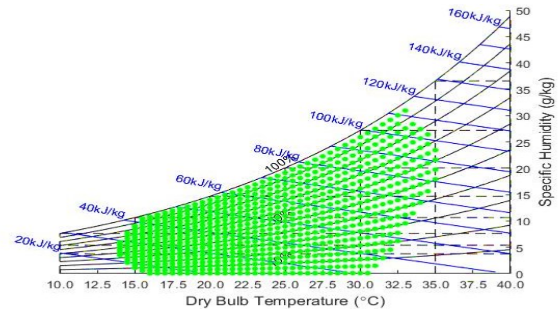
The parameters shown in the previous subchapter can show the differences between classes “warmer", "no change", and "cooler" for the particular value of a parameter such as age to show the potential comfort zone for each parameter value. If the parameter is changed accordingly, the impact of this parameter change on the thermal comfort zones can be mapped and studied. The impact of the parameter change in the thermal comfort zone can be seen in Fig. 6.

This paper assesses the age parameter impact on the thermal comfort zone. This assessment becomes an example of the parameter assessment with this method. The age parameter is one parameter that can show the human condition aspect of

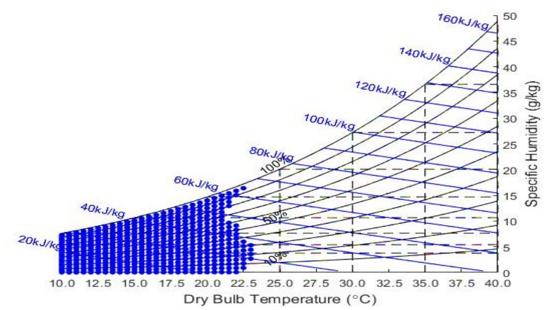
(a)



(b)



(c)



(d)

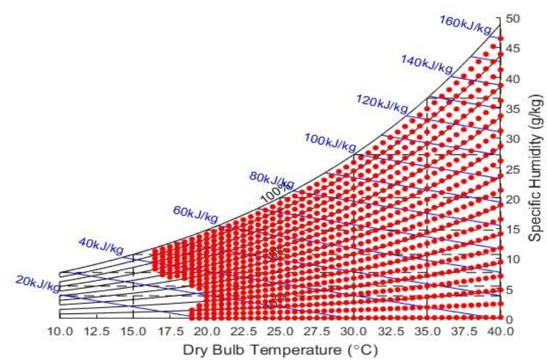


Fig. 6. Psychrometric Mapping for (a) the Comfort Zone Trained with the Filtered ASHRAE Database with the Data Semantic Augmentation for adult, (b) for elderly people group (c) the “warmer” class of the elder people group (d) the "cooler" class of the elderly people group

thermal comfort. The thermal comfort is not standard and is based on personal factors. It has been studied that the young, elderly, disabled, or temporarily ill people group will have a different comfort zone.

Figure 6 (a) displays the representation of the comfort zone from the adult based on the filtered ASHRAE database with the data semantic augmentation. Figure 6 (b) shows the same comfort zone for the elder group. Based on this chart, the system designer will have insight into designing and testing the AI system since the borders between the comfort zones are not crisp and are more personal.

A similar result also happened in the same age group. The differences in clothing insulation and activity/metabolism can have different thermal conditions. This condition is shown in Fig. 6 (b), 6 (c) and 6 (d). Figure 6 (b) represents the class "no change", Fig. 6 (c) represent the class "warmer", and Fig. 6 (d) represents the class "cooler". This map shows that the class "no change" intersect with the class "warmer" and "cooler". This case means one person feels cold, but another person can feel comfortable. This condition highlights the need for occupant features to alter the system setting. Figure 7 shows the ability of the proposed method to acknowledge 98.90% of total

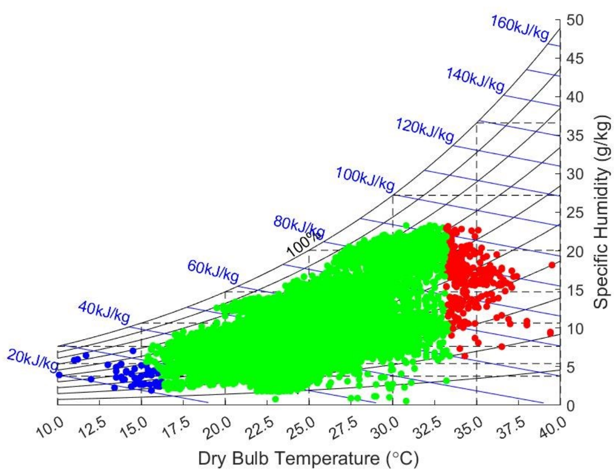


Fig. 7. Proposed method ability to acknowledge 98.9% from ASHRAE comfort data.

ASHRAE comfort data which is 6.06% wider compared to the PMV-PPD and Givoni comfort area that able to capture 92.84% comfort. This result will benefit the system that can deliver lower energy usage for achieving indoor comfort.

Based on the assumption that there are 27.8 million households in the UK [39] and the annual median energy consumption for the UK household is 15,400 kWh/year [40] and the assumption that 61% of energy is used for space heating [41], the total energy spent for the annual domestic heating energy across the UK is 261.1532 billion kWh/year. With this massive amount of value, if the 6.06% wider comfort area acknowledgement is directly associated with the same amount of energy saving, it will equal 14.34 billion kWh/year. If the CO_2 emission factor is 0.309 kge / kWh [42], this work will contribute to 4,432 thousand tonnes of CO_2 equivalent. If the emission factor used is 50 g CO_2 eq/kWh, which is the target for 2030 [43], the contribution of this work will be about 717.2 thousand tonnes of CO_2 equivalent per year.

D. Potential Refinement of the Model and the Future Works

Semantic augmentation has proven to be robust in the processing of thermal data. There is the possibility that the semantic augmentation can be implemented in other parameters without changing the notion of comfort that is stored in the ASHRAE database. Another potential development of the system is implementing the recommendation and gamification system to lower energy use but maintain comfort. Since thermal comfort is the state of mind related to memory and not just physiology, the gamification feature and the intelligent system preset can help achieve the goal of lower energy use either for heating or cooling. The system can influence the user to feel comfortable with the gamification and recommendation, but it will need a long adaptation process [8]. For low temperatures, for example, exposure to cold acclimation can improve the subjective responses to cold [30]. This is why this research for the use of the ASHRAE database is essential, to give the fundamental ability to the intelligent system to deliver comfort. There can also be a healthier target set in the system, like exposing the user to lower temperatures to decrease body fat [28].

V. Conclusion and Future Work

Obtaining the dataset for thermal comfort is not an easy task. This work develops the filtering and semantic augmentation for the multiple ASHRAE databases, the most reliable databases for thermal comfort. This work proves that the use of an Artificial Neural Network with proper data training can outperform other methods in the thermal comfort zone acknowledgement at least by 6.06% leading to energy saving for comfort. This work can contribute to 717.2 thousand tonnes of CO2 equivalent per year which is beneficial for a more sustainable thermal comfort system and the development of a reinforced learning system for thermal comfort.

This concludes that the simple ANN can perform well with the accurate training process and can be the solution for the base system that does not have large memory or computational power to run complex AI algorithms with better interpretability. The validation part of this work shows that even though the training percentage results were high, it needs to be validated with mapping to the comfort zone, to identify the borders of acknowledgement. This work shows the benefit of validation methods based on the visual psychrometric comfort zone mapping to avoid false claims on AI performance for thermal comfort.

The future work will be the field trial of this method in the occupied dwelling and the system development by elaborating on the gamification features of the system.

Appendix

CLASSIFICATION METHODS ACCURACY COMPARISON FOR DATABASE ALLOCATED FOR TRAINING AND VALIDATION (IN %)

Classification Methods	Accuracy with all available data for training and validation (%) without PCA													
	DB1	Filtered DB1	DB1 without age	Filtered DB1 without age	DB2	Filtered DB2	DB2 without age	Filtered DB2 without age	DB1 and DB2	Filtered DB1 and DB2	DB1 and DB2 without age	Filtered DB1 and DB2 without age	Augmented & Filtered DB1 and DB2	d & Filtered DB1 and DB2 without
Fine Tree	59.20	89.60	60.70	89.40	42.50	90.20	48.50	90.30	45.00	89.80	49.00	89.7	93.90	93.80
Medium Tree	57.70	88.00	57.80	87.90	41.10	89.90	45.40	90.00	43.60	89.30	46.50	89.3	93.70	93.70
Coarse Tree	53.70	87.20	54.70	87.50	39.70	89.80	41.70	89.80	42.30	89.10	42.30	88.9	93.60	93.60
Linear Discriminant	51.40	87.30	52.00	86.80	38.50	89.80	40.70	89.50	42.00	89.00	43.60	88.7	77.40	89.20
Quadratic Discriminant	49.50	86.70	52.80	86.20	36.80	89.30	42.50	88.20	37.80	88.50	43.70	87.5	77.50	89.10
Gaussian Naïve Bayes	52.90	85.10	52.30	85.20	43.30	88.00	42.70	88.00	43.60	86.80	43.30	86.7	95.50	95.50
Kernel Naïve Bayes	55.70	87.50	55.70	87.30	43.20	90.20	42.60	90.10	47.40	89.30	46.00	89.1	95.60	95.60
Linear SVM	36.70	50.10	41.00	67.50	36.20	38.10	43.90	73.80	35.40	41.50	42.90	71.4	78.30	89.60
Quadratic SVM	38.20	50.30	35.10	67.50	36.40	38.10	37.80	72.50	33.40	41.60	29.90	59.1	78.40	72.10
Cubic SVM	27.10	47.20	21.00	48.90	26.70	37.80	21.70	32.10	24.80	36.50	28.00	62.8	78.20	81.70
Fine Gaussian SVM	45.40	53.60	52.10	70.50	41.20	38.70	56.10	74.80	41.40	42.70	53.60	73.2	78.50	89.70
Medium Gaussian SVM	38.50	51.80	45.20	68.70	38.70	38.20	49.70	74.00	38.10	41.90	47.40	72.2	78.40	89.60
Coarse Gaussian SVM	37.40	50.30	42.50	67.50	36.90	38.10	46.00	73.80	36.50	41.60	44.00	71.8	78.30	89.60
Fine KNN	73.00	95.60	87.20	97.90	50.70	91.10	79.30	97.10	56.00	92.40	81.20	**97.4**	78.80	92.40
Medium KNN	57.40	89.30	62.50	89.10	43.70	90.00	55.60	90.50	46.70	89.60	56.20	89.9	77.70	89.60
Coarse KNN	54.20	87.80	57.70	87.70	42.20	89.80	49.70	90.00	44.50	89.00	50.50	89.2	77.70	89.30
Cosine KNN	56.60	88.90	60.90	88.40	43.40	89.90	53.80	90.30	46.30	89.40	54.50	89.4	77.60	89.40
Cubic KNN	57.30	89.30	62.20	89.00	43.70	90.00	55.50	90.50	46.60	89.60	56.10	89.9	77.70	89.60
Weighted KNN	73.00	95.60	87.30	**97.90**	50.90	91.10	70.80	**97.20**	56.10	92.40	81.60	97.4	78.80	92.50
Ensemble Boosted Trees	58.30	88.30	58.20	88.10	41.60	89.90	45.70	90.00	43.70	89.30	47.30	89.3	96.10	96.10
Ensemble Bagged Trees	78.10	95.50	**88.20**	97.60	53.70	91.10	**80.40**	96.60	56.70	92.30	**81.90**	96.9	97.00	**98.70**
Ensemble Subspace Discriminant	52.70	87.20	52.70	87.00	43.90	90.00	41.70	90.10	44.90	89.20	43.80	89	94.40	94.80
Ensemble Subspace KNN	**94.10**	**99.10**	78.50	94.80	**79.80**	**97.60**	59.50	93.80	**86.30**	**98.60**	58.50	93.2	**98.20**	64.50
Ensemble RUS Boosted Trees	41.20	49.20	42.90	57.00	41.10	57.20	42.40	43.80	36.30	48.90	35.00	46.7	93.70	93.70
Narrow Neural Network	40.00	88.50	45.90	68.50	38.00	35.50	47.30	73.10	37.30	89.20	45.90	71.4	96.00	89.70
Medium Neural Network	40.80	52.30	46.90	69.00	38.40	37.90	49.50	74.10	37.80	89.40	47.20	72.2	96.00	96.00
Wide Neural Network	42.20	90.40	50.10	70.10	34.60	36.20	49.30	90.30	38.80	42.20	49.40	72.5	**96.10**	**96.10**
Bilayered Neural Network	40.50	88.80	48.50	88.50	38.20	89.70	49.50	73.30	37.80	89.30	47.80	89.2	96.00	89.70
Trilayered Neural Network	55.50	52.30	48.60	89.10	26.80	89.70	50.50	90.20	38.00	89.40	48.20	89.5	78.50	89.70
Average	**52.36**	**77.68**	**55.21**	**81.40**	**41.79**	**72.86**	**49.65**	**82.68**	**43.62**	**76.82**	**49.84**	**82.53**	**86.47**	**90.16**
Classification Methods	DB1	Filtered DB1	DB1 without age	Filtered DB1 without age	DB2	Filtered DB2	DB2 without age	Filtered DB2 without age	DB1 and DB2	Filtered DB1 and DB2	DB1 and DB2 without age	Filtered DB1 and DB2 without age	Augmented & Filtered DB1 and DB2	Augmente d & Filtered DB1 and DB2

Note: DB1 = ASHRAE RP-884 database ; DB2 = ASHRAE Global Thermal Comfort Database II

Methods	Parameters
Fine Tree	Gini's diversity index, 100 max number of splits
Medium Tree	Gini's diversity index, 20 max number of splits
Coarse Tree	Gini's diversity index, 4 max number of splits
Linear Discriminant	Full covariance structure
Quadratic Discriminant	Full covariance structure
Gaussian Naïve Bayes	Gaussian, unbounded support
Kernel Naïve Bayes	Gaussian Kernel type, unbounded support
Linear SVM	1 box constraint level, auto kernel scale mode, one-vs-one multiclass method, standardise data
Quadratic SVM	1 box constraint level, auto kernel scale mode, one-vs-one multiclass method, standardise data
Cubic SVM	1 box constraint level, auto kernel scale mode, one-vs-one multiclass method, standardise data
Fine Gaussian SVM	1 box constraint level, kernel scale: 0.56, one-vs-one multiclass method, standardise data
Medium Gaussian SVM	1 box constraint level, kernel scale: 2.2, one-vs-one multiclass method, standardise data
Coarse Gaussian SVM	1 box constraint level, kernel scale: 8.9, one-vs-one multiclass method, standardise data
Fine KNN	1 Number of neighbours, Euclidean distance metric, Equal distance weight, standardise data
Medium KNN	10 Number of neighbours, Euclidean distance metric, Equal distance weight, standardise data
Coarse KNN	100 Number of neighbours, Euclidean distance metric, Equal distance weight, standardise data
Cosine KNN	10 Number of neighbours, Cosine distance metric, Equal distance weight, standardise data
Cubic KNN	10 Number of neighbours, Minkowski (Cubic) distance metric, Equal distance weight, standardise data
Weighted KNN	10 Number of neighbours, Euclidean distance metric, Squared distance weight, standardise data
Ensemble Boosted Trees	AdaBoost ensemble method, Decision Tree learner type, 20 maximum number of splits, 30 number of learners, 0.1 learning rate
Ensemble Bagged Trees	Bag ensemble method, Decision Tree learner type, 65260maximum number of splits, 30 number of learners
Ensemble Subspace Discriminant	Subspace ensemble method, Discriminant learner type, 30 number of learners, 3 Subspace dimension
Ensemble Subspace KNN	Subspace ensemble method, Nearest neighbours learner type, 30 number of learners, 3 Subspace dimension
Ensemble RUS Boosted Trees	RUSBoost ensemble method, Decision Tree learner type, 20 maximum number of splits, 30 number of learners, 0.1 learning rate
Narrow Neural Network	1 Number of fully connected layers, layer size 10, ReLU Activation, 1000 iteration limit, 0 regularisation strength (Lambda), standardise data
Medium Neural Network	1 Number of fully connected layers, layer size 25, ReLU Activation, 1000 iteration limit, 0 regularisation strength (Lambda), standardise data
Wide Neural Network	1 Number of fully connected layers, layer size 100, ReLU Activation, 1000 iteration limit, 0 regularisation strength (Lambda), standardise data
Bilayered Neural Network	2 Number of fully connected layers, first and second layer size 10, ReLU Activation, 1000 iteration limit, 0 regularisation strength (Lambda), standardise data
Trilayered Neural Network	3 Number of fully connected layers, first, second and third layer size 10, ReLU Activation, 1000 iteration limit, 0 regularisation strength (Lambda), standardise data

References

[1] E. I. S. Department for Business, Prime Minister's Office, "UK's path to net zero set out in landmark strategy," in *The Net Zero Strategy sets out how the UK will deliver on its commitment to reach net zero emissions by 2050.*, ed. London, United Kingdom: Prime Minister's Office, 2021.

[2] T. R. H. P. H. HM Treasury, "Spring Statement 2019: what you need to know," 13 March 2019. [Online]. Available: https://www.gov.uk/government/news/spring-statement-2019-what-you-need-to-know

[3] N. E. Klepeis *et al.*, "The National Human Activity Pattern Survey (NHAPS): a resource for assessing exposure to environmental pollutants," *Journal Of Exposure Analysis And Environmental Epidemiology,* Original Article vol. 11, p. 231, 07/24/online 2001, doi: 10.1038/sj.jea.7500165.

[4] (19 October 2021). *A market-based mechanism for low-carbon heat*. [Online] Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1026607/clean-heat-market-consultation.pdf

[5] C. I. o. B. S. E. (CIBSE). "CIBSE launches design guidance for heat pumps in high density housing." https://www.cibse.org/news-and-policy/september-2021/cibse-launches-design-guidance-for-heat-pumps-in-h (accessed 12 December 2021.

[6] I. N. Joshua Bird, Ilaria Ricci Curbastro and Michael Edwards, *Heat pump installations for multi-unit residential buildings*, A. Deas, ed., London, UK: The Chartered Institution of Building Services Engineers 2021.

[7] (December 2021). *English Housing Survey*. [Online] Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1039214/2020-21_EHS_Headline_Report.pdf

[8] K. Karyono, B. M. Abdullah, A. J. Cotgrave, and A. Bras, "The adaptive thermal comfort review from the 1920s, the present, and the future," *Developments in the Built Environment,* vol. 4, p. 100032, 2020/11/01/ 2020, doi: https://doi.org/10.1016/j.dibe.2020.100032.

[9] S. Baldi, C. D. Korkas, M. Lv, and E. B. Kosmatopoulos, "Automating occupant-building interaction via smart zoning of thermostatic loads: A switched self-tuning approach," *Applied energy,* vol. 231, pp. 1246-1258, 2018.

[10] X. Zhou *et al.*, "Data-driven thermal comfort model via support vector machine algorithms: Insights from ASHRAE RP-884 database," *Energy and Buildings,* vol. 211, p. 109795, 2020/03/15/ 2020, doi: https://doi.org/10.1016/j.enbuild.2020.109795.

[11] M. Luo *et al.*, "Comparing machine learning algorithms in predicting thermal sensation using ASHRAE Comfort Database II," *Energy and Buildings,* vol. 210, p. 109776, 2020/03/01/ 2020, doi: https://doi.org/10.1016/j.enbuild.2020.109776.

[12] C. Q. Choi, "7 Revealing Ways AIs Fail: Neural Networks can be Disastrously Brittle, Forgetful, and Surprisingly Bad at Math," *IEEE Spectrum,* vol. 58, no. 10, pp. 42-47, 2021, doi: 10.1109/MSPEC.2021.9563958.

[13] P. Lisboa, S. Saralajew, A. Vellido, R. Fernández-Domenech, and T. Villmann, "The coming of age of interpretable and explainable machine learning models," *Neurocomputing,* vol. 535, pp. 25-39, 2023.

[14] R. de Dear, J. Xiong, J. Kim, and B. Cao, "A review of adaptive thermal comfort research since 1998," *Energy and Buildings,* vol. 214, p. 109893, 2020/05/01/ 2020, doi: https://doi.org/10.1016/j.enbuild.2020.109893.

[15] P. O. Fanger, *Thermal comfort: Analysis and applications in environmental engineering*. Copenhagen Denmark: Danish Technical Press, 1970, p. 244.

[16] P. Höppe, "Different aspects of assessing indoor and outdoor thermal comfort," *Energy and Buildings,* vol. 34, no. 6, pp. 661-665, 2002/07/01/ 2002, doi: https://doi.org/10.1016/S0378-7788(02)00017-8.

[17] D. Enescu, "A review of thermal comfort models and indicators for indoor environments," *Renewable and Sustainable Energy Reviews,* vol. 79, pp. 1353-1379, 2017, doi: 10.1016/j.rser.2017.05.175.

[18] R. F. Rupp, N. G. Vásquez, and R. Lamberts, "A review of human thermal comfort in the built environment," *Energy and Buildings,* vol. 105, pp. 178-205, 2015, doi: 10.1016/j.enbuild.2015.07.047.

[19] G. S. Brager and R. J. de Dear, "Thermal adaptation in the built environment: a literature review," *Energy and Buildings,* vol. 27, no. 1, pp. 83-96, 1998/02/01/ 1998, doi: https://doi.org/10.1016/S0378-7788(97)00053-4.

[20] M. H. Fergus Nicol, Susan Roaf, *adaptive thermal comfort foundations and analysis*. London: Routledge, 2015.

[21] ASHRAE, "ANSI/ASHRAE Standard 55 Thermal Environmental Conditions for Human Occupancy," ed, 2017.

[22] M. Luo, *The Dynamics and Mechanism of Human Thermal Adaptation in Building Environment: A Glimpse to Adaptive Thermal Comfort in Buildings*. Springer Nature, 2019.

[23] K. Karyono, B. M. Abdullah, A. J. Cotgrave, and A. Bras, "Experience and Memory Principle for Adaptive Indoor Thermal Comfort," in *Intelligent and Reliable Engineering Systems: 11th International Conference on Intelligent Energy Management, Electronics, Electric & Thermal Power, Robotics and Automation (IEMERA-2020)*, 2021: CRC Press, pp. 14-19.

[24] t. H. a. S. Executive. "Thermal comfort." The Health and Safety Executive. http://www.hse.gov.uk/temperature/thermal/ (accessed 13 November 2019, 2019).

[25] B. Givoni, "Comfort, climate analysis and building design guidelines," *Energy and Buildings,* vol. 18, no. 1, pp. 11-23, 1992/01/01/ 1992, doi: https://doi.org/10.1016/0378-7788(92)90047-K.

[26] R. Basu and J. M. Samet, "An exposure assessment study of ambient heat exposure in an elderly population in Baltimore, Maryland," (in eng), *Environmental health perspectives,* vol. 110, no. 12, pp. 1219-1224, 2002, doi: 10.1289/ehp.021101219.

[27] E. H. K. Yung, S. Wang, and C.-k. Chau, "Thermal perceptions of the elderly, use patterns and satisfaction with open space," *Landscape and Urban Planning,* vol. 185, pp. 44-60, 2019, doi: 10.1016/j.landurbplan.2019.01.003.

[28] A. A. van der Lans *et al.*, "Cold acclimation recruits human brown fat and increases nonshivering thermogenesis," *The Journal of clinical investigation,* vol. 123, no. 8, pp. 3395-3403, 2013.

[29] K. Karyono, A. Romano, B. M. Abdullah, J. Cullen, and A. Bras, "The role of hygrothermal modelling for different housing typologies by estimating indoor relative humidity, energy usage and anticipation of fuel poverty," *Building and Environment,* vol. 207, p. 108468, 2022/01/01/ 2022, doi: https://doi.org/10.1016/j.buildenv.2021.108468.

[30] M. J. Hanssen *et al.*, "Short-term cold acclimation recruits brown adipose tissue in obese humans," *Diabetes,* vol. 65, no. 5, pp. 1179-1189, 2016.

[31] V. Földváry Ličina *et al. ASHRAE Global Thermal Comfort Database II*, doi: https://doi.org/10.6078/D1F671

[32] V. Földváry Ličina *et al.*, "Development of the ASHRAE Global Thermal Comfort Database II," *Building and Environment,* vol. 142, pp. 502-512, 2018/09/01/ 2018, doi: https://doi.org/10.1016/j.buildenv.2018.06.022.

[33] W. Ji, Y. Zhu, and B. Cao, "Development of the Predicted Thermal Sensation (PTS) model using the ASHRAE Global Thermal Comfort Database," *Energy and Buildings,* vol. 211, p. 109780, 2020/03/15/ 2020, doi: https://doi.org/10.1016/j.enbuild.2020.109780.

[34] T. Cheung, S. Schiavon, T. Parkinson, P. Li, and G. Brager, "Analysis of the accuracy on PMV – PPD model using the ASHRAE Global Thermal Comfort Database II," *Building and Environment,* vol. 153, pp. 205-217, 2019, doi: 10.1016/j.buildenv.2019.01.055.

[35] Z. Wang *et al.*, "Revisiting individual and group differences in thermal comfort based on ASHRAE database," *Energy and Buildings,* vol. 219, p. 110017, 2020/07/15/ 2020, doi: https://doi.org/10.1016/j.enbuild.2020.110017.

[36] Z. Wang, J. Wang, Y. He, Y. Liu, B. Lin, and T. Hong, "Dimension analysis of subjective thermal comfort metrics based on ASHRAE Global Thermal Comfort Database using machine learning," *Journal of Building Engineering,* vol. 29, p. 101120, 2020/05/01/ 2020, doi: https://doi.org/10.1016/j.jobe.2019.101120.

[37] M. Ye, J. Shen, X. Zhang, P. C. Yuen, and S. F. Chang, "Augmentation Invariant and Instance Spreading Feature for Softmax Embedding," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* pp. 1-1, 2020, doi: 10.1109/TPAMI.2020.3013379.

[38] Y. Wang, G. Huang, S. Song, X. Pan, Y. Xia, and C. Wu, "Regularizing Deep Networks with Semantic Data Augmentation,"

IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1-1, 2021, doi: 10.1109/TPAMI.2021.3052951.

[39] O. f. N. Statistics, "Families and households in the UK: 2020," in "Statistical bulletin," 2021. [Online]. Available: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/families/bulletins/familiesandhouseholds/2020

[40] E. I. S. B. Department for Business, "Energy Follow Up Survey: Household Energy Consumption & Affordability," in "Final report ", 2021. [Online]. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1018725/efus-Household-Energy-Consumption-Affordability.pdf

[41] N. E. F. (NEF), "Energy Tutorial: Energy and Sustainability What's energy used for? ," 2014. [Online]. Available: http://www.energyenvoys.org.uk/sites/default/files/What%27s%20energy%20used%20for_0.pdf

[42] D. f. B. E. I. S. (BEIS), "2018 GOVERNMENT GHG CONVERSION FACTORS FOR COMPANY REPORTING," in "Methodology paper for emission factors: final report," 2018. [Online]. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/726911/2018_methodology_paper_FINAL_v01-00.pdf

[43] T. P. O. o. S. a. Technology, "Carbon Footprint of Electricity Generation," June 2011 2011, vol. Number 383. [Online]. Available: https://www.parliament.uk/globalassets/documents/post/postpn_383-carbon-footprint-electricity-generation.pdf

Kanisius Karyono (Student Member IEEE) currently is pursuing his PhD in the department of the Built Environment, Liverpool John Moores University. He received his bachelor in Informatics from Atma Jaya Yogyakarta University in 2000 and a bachelor in Electrical Engineering from Gadjah Mada University Indonesia in 2002.

He received his Master in Electrical Engineering from Pelita Harapan University Indonesia in 2009. He works at the Department of Electrical Engineering, Universitas Multimedia Nusantara Indonesia.

Kanisius researches Human Comfort (Lighting and Thermal Comfort), Internet of Things, Smart System, Wireless Sensor Network, communication protocol modelling, and Industrial 4.0.

Badr Abdullah is the Programme Leader of Architectural Engineering & Building Services Engineering at the Department of the Built Environment. Badr gained an MEng in Electrical Engineering and Electronics in 2004 and a PhD in Vision Monitoring Systems for Hybrid Welding Processes and Hazardous Environments in Real-time in 2007 from the University of Liverpool.

His main teaching areas are Engineering and Applied Mathematics, Engineering Principles and Electrical Engineering. As well as teaching,

Badr heads the i4Sens3D (Industry 4, Sensors and Drones) Research Group in the Faculty of Engineering and Technology. Badr has extensive research interests spanning a wide range of unique applied science and technologies for the design and development of sensors to suit a wide range of industries such as built environment, energy, oil/gas, environmental monitoring, manufacturing and healthcare.

Alison Cotgrave is the Professor of Built Environment Education and Subject Leader in the School of Civil Engineering and Built Environment with responsibility for Civil Engineering, Architectural Engineering, Building Services Engineering, Construction Management, Architectural Technology, Real Estate, Quantity Surveying and Building Surveying. However, She is currently seconded to the post of Academic Registrar until May 2022.

She has worked in academia for 27 years and was previously employed in the construction industry as a site manager. After graduating with BSc(Hons) Building in 1988 she subsequently completed a Masters of Education at the University of Manchester in 1997, followed by a PhD at LJMU in 2008.

Prof. Alison is a Fellow of the Chartered Institute of Building, Member of the Royal Institution of Chartered Surveyors, Chartered Environmentalist, Fellow of the Leadership Foundation and a Senior Fellow of the Higher Education Academy. She has co-authored a number of books on Construction Technology and have published extensively in the area of Sustainability in Construction Education. She is currently on the supervision teams for a number of PhD students whose topics range from construction pedagogy to the design of a decision making aid for the selection of renewable/sustainable energy systems for buildings. She is the External Examiner at a number of HEIs and a member of the Chartered Institute of Building accreditation panel, which reviews and accredits Construction Management programmes in the UK and internationally.

Ana Bras is a Civil Engineer, CEng, FICE, with 17 years of experience in R&D and Consultancy on bio-based composites for durability and hygrothermal risks minimisation of retrofitting solutions for buildings and RC structures. She is a Reader (Associate Professor) in Bio-materials for infrastructure at the Department of Civil Engineering and Built Environment, Liverpool John Moores University (UK).

She graduated in 2004 with a MEng of a five years degree university programme in Civil Engineering at NOVA University of Lisbon. In 2011, she completed her PhD studies in Civil Engineering at NOVA (speciality in Building Rehabilitation). She is a researcher and research project manager, with active participation in several R&D projects and consultancy for institutions including: NOVA University of Lisbon (Portugal), University of Bath (UK), University of Campinas (UNICAMP, Brazil), Polytechnic Institute of Setúbal (IPS, Portugal), Institute of Welding and Quality (ISQ, Portugal) and more recently, at LJMU. Concurrently, she is taking part in HORIZON 2020 task groups to evaluate EU projects.

Ana develops research work in the field of performance based design with bio-inspiring solutions (laboratory and computational-based) for the selection of the best retrofitting solutions for buildings and structures. She uses probabilistic approaches to increase the effectiveness of sustainable solutions in terms of their service life, and to estimate energy savings and hygrothermal risks.

Jeff Cullen graduated from the University of Liverpool, School of Electrical Engineering and Electronics, with a BEng(Hons) in Computer Electronics and Robotics. He then obtained a PhD at Liverpool in 2000 entitled 'Optical Techniques for Industrial Automated Applications'. His post doctoral research career began at Liverpool working for Prof. Jim Lucas on a tyre pressure monitoring system for trucks. He also worked on a novel UV microwave system, before working on an online spot welding monitor, predicting weld quality and monitoring energy usage.

In 2005 he joined LJMU as a founding member of the RF and Microwave group under Prof. Ahmed Al-Shamma'a, which later became part of the newly formed BEST (Built Environment and Sustainable Technologies) Research Institute.

Here he continued research related to energy, investigating the performance of fridges for the leisure industry. He has also used microwave systems to provide a novel way of heating showers and remediation of waste oil for generator and biodiesel use. He has also lectured at both undergraduate and postgraduate level, delivering courses on Automotive Electronics, Programming and Software Engineering for the School of Engineering. He has helped deliver modules related to Energy Management and Building Monitoring Systems for the School of the Built Environment.