Semi-Supervised Learning Techniques for Automated Fault Detection and Diagnosis of HVAC Systems

Maitreyee Dey, Soumya Prakash Rana, Sandra Dudley
Biomedical Engineering and Communications (BiMEC) Research Group,
School of Engineering, London South Bank University,
London, United Kingdom

Abstract—This work demonstrates and evaluates semi-supervised learning (SSL) techniques for heating, ventilation and air-conditioning (HVAC) data from a real building to automatically discover and identify faults. Real HVAC sensor data is unfortunately usually unstructured and unlabelled, thus, to ensure better performance of automated methods promoting machine-learning techniques requires raw data to be pre-processed, increasing the overall operational costs of the system employed and makes real time application difficult. Due to the data complexity and limited availability of labelled information, semi-supervised learning based robust automatic fault detection and diagnosis (AFDD) tool has been proposed here. Further, this method has been tested and compared for more than 50 thousand TUs. Established statistical performance metrics and paired t-test have been applied to validate the proposed work.

Keywords-Smart buildings, HVAC terminal unit, Automatic fault detection and diagnosis, Semi-supervised learning.

I. INTRODUCTION

Energy monitoring and performance degradation of building heating, ventilation and air-conditioning (HVAC) systems are often ignored until they result in significant impact on occupant comfort, trigger an equipment-level alarm, deteriorates equipment life or results in excessive energy consumption. Thus, a building energy management system (BEMS) is installed or retrofitted into many new and existing buildings to overcome these issues and help building managers pave the way for greater energy efficiency and occupant comfort improvements. BEMS units can comprise of many sensors with thousands of sensors common in large buildings. Manual fault finding has become a problem that only highly qualified staff can address leading to prolonged BEMS fault and maintenance issues. Thus, automatic and remote identification of real unit faults plays a crucial role in both improving BEMS - building manager relationships as well as creating a "fit for purpose" buildings that match their design criteria.

The study of HVAC systems began in 1980's [1], [2], and since then significant development has been made by means of data mining techniques for fault detection and diagnosis (FDD). The FDD research has been categorized into multiple approaches and data-driven based techniques gained more attention among them as it is appropriate for modern HVAC systems and being used in huge number of commercial buildings [3]. Data-driven based methods do not need any explicit model to build relationships between different data patterns and find out faulty units of a building [3]. Generally, this approach is suitable for fault detection rather than fault diagnosis. Integrated or hybrid approaches are considered to solve this limitation of FDD [4]. Signal processing based methods such as wavelet transformation, short time Fourier analysis and a combination with principle component analysis (PCA) is proposed to diagnose faults for air handling unit (AHU) [5], [6]. While, expert knowledge based techniques are limited due to the unavailability of real data thus, machine-learning algorithms like Hidden Markov Model (HMM) [7], Kernel Machines (KM) [8] are applied to deal with this, where knowledge is automatically extracted from data. In addition, physical characteristics classification-based techniques are employed to build non-linear correlations between non-faulty and faulty units in the absence of strong prototypes. Machine-learning classification algorithms such as, Bayes classifier [9], artificial neural networks (ANN) [8], [10], support vector machine (SVM) [11], and fuzzy logic [12] are too applied for efficient FDD models in large buildings.

These models have been constructed to handle specific fault types (e.g. fan failure, stuck valve) and these TU data analysis has been given little to no attention in recent research.

In this paper, the proposed experiments have been conducted on a specific sub-unit of the HVAC terminal unit (TU) which is a "final delivery" section of a fan coil unit (FCU). There are hundreds of these devices installed in a building, and if a single TU malfunctions, it may result in performance deficiencies causing excessive energy use over time. Manual fault finding in devices such as these is very difficult, thus, data driven based automatic fault detection and diagnosis (AFDD) has been employed on historical building data previously processed by the authors with the aim of remote fault detection and/or prediction to generate real time notifications on faulty TUs. This notification would help for example, building managers to take appropriate action and save timing by fixing faults and reducing multiple investigational visits or worse, fault obliviousness. Because such TU data may be infrequent, BEMS can be badly maintained and raw data is mostly unlabelled thus the main challenges are to discover a faulty TU without appropriate knowledge or labelled information and make a reasoned assumption that a TU drawing high power demand

to maintain control strategies is most likely a faulty TU for example.

Previously unsupervised and supervised machine learning algorithms were investigated by the authors to classify large data sets from TUs over a given period from real buildings. Due to the limited availability of appropriate "fault type" information, enormous levels of data needed to be pre-processed and labelled before learning could be executed. This of course is time consuming but common with real world scenarios where data is not always in appropriate formats [13], [14], [15]. This valuable real-world labelled data created, previously is now used as a training set by the authors as well as new unlabelled data from the same buildings under test for the experiments described in this paper. This previous work is now augmented to investigate semi-supervised learning (SSL) methods for future unlabelled datasets can be useful if historical labelled sets are available. Therefore, a SSL based multi-class support vector machine (MC-SVM) is employed for AFDD and established through training, testing, and validation process.

The paper is outlined as follows; Section II describes the TUs, their working principle and associated faults. Section III explains the proposed FDD tools and semi-supervised learning algorithm. Section IV provides the discussion of result analysis. Section V provides the conclusion and the future directions of this research.

II. TERMINAL UNIT MODEL

A terminal unit (TU) is a small sub-unit within a HVAC system. It is commonly located on the ceiling and manages the flow of hot or cold air to a room. Primarily a TU consists of heating coil, cooling coil, valve, and a fan. Based on the thermostats temperature sensor, it sends signals to the main plants (either a boiler or chiller). If the thermostat sense the room temperature is too warm, then it sends signals to the chiller to start the flow of cold water, which is passed through the cooling coil, and circulate the cool air via the fan to the room. Conversely, if the rooms becomes too cold then the same process repeats but generates signals to the boiler to pass hot air to the room through the fan. The schematic of a TU is shown in Fig.1.
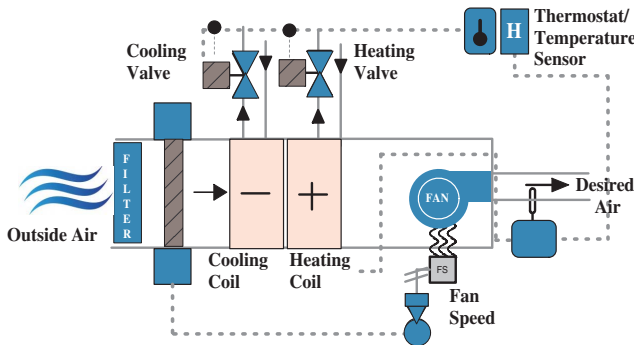

Fig. 1: FCU-TU schematic diagram.

These TUs are distributed throughout the building across the different floors and via a beta virtual meter sensor these data are then collected. The sensors estimate the floor-by-floor heating, cooling and fan energy consumption, several TUs valve position, fan speed data, the boiler or chiller, and pump supply chain so that heating or cooling energy use is only indicated if it appears that a TU is actually being supplied with hot or cold water. A single TU generates multiple data streams; here we have considered control temperature, set point, dead band, heating and cooling power, enable signals. This experiment has been considered around 20 million data points, an example of produced signal from TU is shown in Fig. 2. The blue lined graph in the figure denotes control temperature variation with respect to the heating and cooling set point and corresponding power demand (shown in red) over a month during winter from a building based in central London.

Here, TU data are analysed on daily basis and some of the issues can result in faulty pattern as follows:

1) Incorrect TU sizing to real demand.
2) TU is not receiving adequate temperature.
3) A stuck open valve.
4) Unachievable set point.
5) Poorly positioned temperature sensor.
6) Out of hours operation.

Previously, these faulty and non-faulty data have been thoroughly analysed, and a novel feature extraction (FE) method [13], [14], [15] was implemented for dimension reduction of multi-stream TU data. Here, semi-supervised multi-class support vector machine is employed for the AFDD purpose.

III. PROPOSED AFDD METHODOLOGY

The proposed AFDD methodology consist of four stages: data collection and pre-processing, feature extraction, learning, and prediction. The proposed architecture is shown in Fig. 3.

A. Feature Extraction Process

A novel feature extraction (FE) method has been proposed by the authors previously [13], [14], [15]. and is described here to provide the information for the reader and to demonstrate the exploration base for the SSL. This FE generate events (E) that are divided into three different stages: (1) Event Discovery, (2) Event Area calculation, and (3) Event Aggregation. The stages have been defined through assuming four different step changes of a TU's temperature and power variations during a day by the BEMS enabling signals. These variations have given the names (event start, response delay, goal achieved, event end). These events can be heating and/ or cooling based on the demand and control strategy of the unit.

When the suitable heating and cooling events have been discovered then the area under the temperature and power curve for each event was estimated. Thus, six areas (three from temperature and three from power curve) are calculated for each heating event and similarly, six areas were calculated
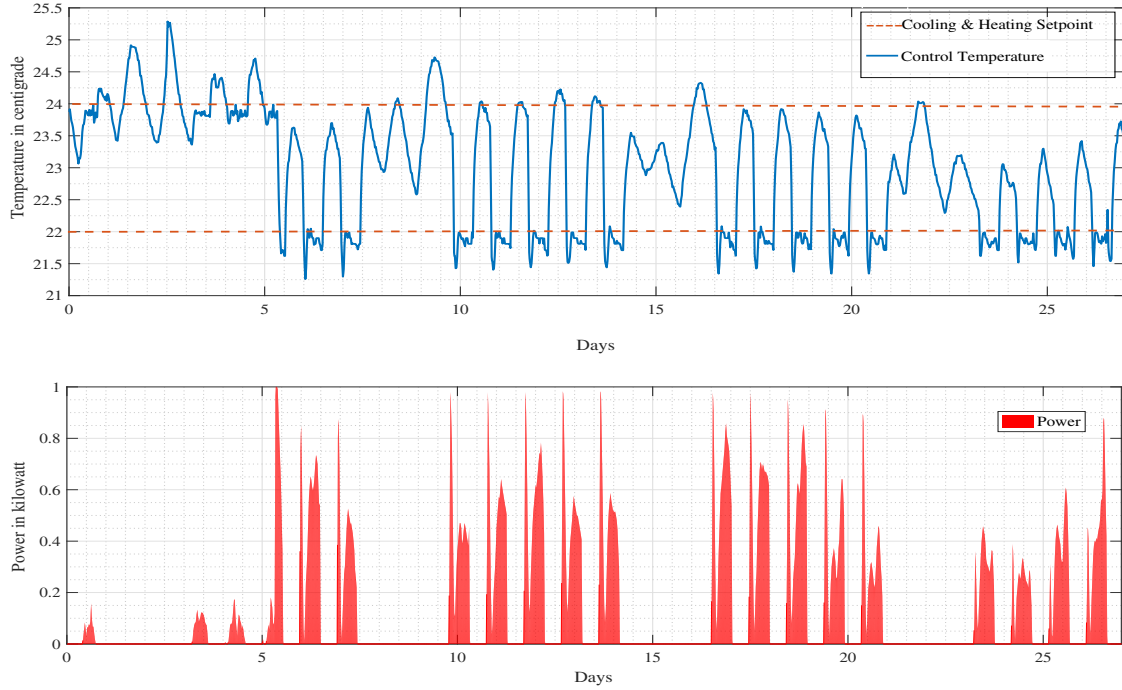
Fig. 2: An example of a TU control temperature and power data signal over 30 days.

for each cooling event. Finally, twelve different areas were derived from a daily TU data stream.

Eqs. (1) shows the area (A_E) calculation under the curve $f(x)$ at each time interval Δx.

$$A_E = \sum_{i=0}^{n} f(x_i)\Delta x \qquad (1)$$

Features are defined as $F_{H_1} - F_{H_6}$ and $F_{C_1} - F_{C_6}$ that is calculated through (2) - (5). Here the area calculation for temperature (T) and for power (P) are denoted by $A_{H_1} - A_{H_3}$ and $A_{H_4} - A_{H_6}$ for heating type. Similarly, $A_{C_1} - A_{C_3}$ and $A_{C_4} - A_{C_6}$ denoted for cooling type.

Eqs. (2) and (3) shows the area calculations for a heating event.

$$F_{H_1} = \frac{A_{H_1}}{T_{H_1}}, \quad F_{H_2} = \frac{A_{H_2}}{T_{H_1}}, \quad F_{H_3} = \frac{A_{H_3}}{T_{H_2}}$$
$$where, \quad T_{H_1} = max(A_{H_1} + A_{H_2}) \qquad (2)$$
$$and, T_{H_2} = max(A_{H_3})$$

$$F_{H_4} = \frac{A_{H_4}}{P_{H_1}}, \quad F_{H_5} = \frac{A_{H_5}}{P_{H_1}}, \quad F_{H_6} = \frac{A_{H_6}}{P_{H_2}}$$
$$where, \quad P_{H_1} = max(A_{H_4} + A_{H_5}) \qquad (3)$$
$$and, P_{H_2} = max(A_{H_6})$$

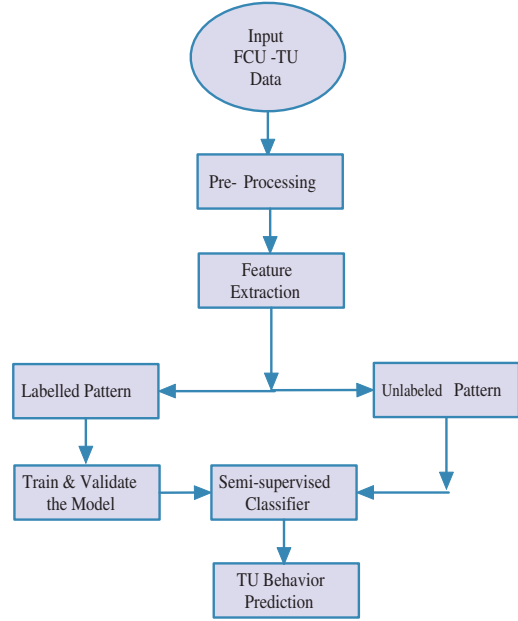Eqs. (4) and (5) shows the area calculations for a cooling event.

Fig. 3: Proposed AFDD architechture.

$$F_{C_1} = \frac{A_{C_1}}{T_{C_1}}, \quad F_{C_2} = \frac{A_{C_2}}{T_{C_1}}, \quad F_{C_3} = \frac{A_{C_3}}{T_{C_2}}$$
$$where, \quad T_{C_1} = max(A_{C_1} + A_{C_2}) \qquad (4)$$
$$and, T_{C_2} = max(A_{C_3})$$

$$F_{C_4} = \frac{A_{C_4}}{P_{C_1}}, \quad F_{C_5} = \frac{A_{C_5}}{P_{C_1}}, \quad F_{C_6} = \frac{A_{C_6}}{P_{C_2}}$$
$$where, \quad P_{C_1} = max(A_{C_4} + A_{C_5}) \quad (5)$$
$$and, P_{C_2} = max(A_{C_6})$$

Thus, due to the occurrence of multiple event types in a single day all the events are aggregated to represent the averaged values. Where, k denotes the event number and n denotes total number of occurrences for event of each type. Therefore, the daily TU is represented by twelve features, calculated in (6) and (7).

$$F_{H_k} = \frac{1}{n} \sum_{i=1}^{n} (F_{H_{k_i}}) \quad (6)$$

$$F_{C_k} = \frac{1}{n} \sum_{i=1}^{n} (F_{C_{k_i}}) \quad (7)$$

B. Semi-supervised Learning

These feature extraction steps intend to derive informative and non-redundant values about TU characteristics, which helps the proposed semi-supervised learning framework in the identification of significant TU patterns. In this test, six different classes of faulty and non-faulty TU patterns are available for specific period and used as labelled data. Then, multi class support vector machine (MC-SVM) [16], [17] is employed into SSL framework for classifying the faulty and non-faulty TU patterns. This SSL model is simple yet more efficient and adopts three steps: training, testing and validation. The obtained data have been randomly divided for training and testing phase. Subsequently, the training and testing accuracy of the proposed model have been measured through precision and recall calculation. Thereafter, unlabelled data are fed into the best scored SSL model to predict the faulty and non-faulty TU patterns. This prediction then validated through paired t-test [18], [19] which has been determined for understanding the correlation between historical data (labelled) and predicted data (unlabelled).

C. Model Validation

Precision and recall have been measured to validate the training and testing phase where label information are available which assist to find out the true TU predictions (truly faulty and non-faulty TU) and false predictions (wrong TU class prediction). Precision and recall are then calculated from these true and false predictions. SSL has been applied to the unlabelled data therefore true and false predictions could not be calculated. Thus, the paired t-test has been estimated to investigate the correlation between a labelled class and the same TU class predicted by the SSL algorithm. Therefore, the null hypothesis symbolise the fitness of a predicted TU class data with the TU belongs to that class in historical data. Test result delivers one to denote the rejection of consideration of predicted data in the same class of labelled data and zero for acceptance based on the probability (p-value) of test observation. Low probability or p-value implies the invalidity of null hypothesis. The results of precision, recall, and t-test have been discussed in the next section.

IV. RESULT ANALYSIS & DISCUSSION

This experiment has been tested and observed on a commercial building of London over a period from 2015 to 2017. Details of this case study is described below:

Case Study:
Data has been collected from a building established in 1960 at London, which has been renovated later in 2009. It covers 149,000 sq. ft. for offices and 8,000 sq. ft. for retail space. The building has 17 floors and 731 TUs in all spread across the different floors. The data have been gathered through a data acquisition device (DAD) at continuous 15 minute intervals and stored in a Cassandra cloud. This TU data is then retrieved for pre-processing by the authors through a secure network to the cloud.

Features have been extracted from all available (old and new) TU data where the old TU data have labels and are used to train the model for AFDD. The rest of the unlabelled data are only used for prediction using the trained model. The data have been extracted from 17th July 2015 onwards, where one whole day has been considered to train the model (training and testing) with the help of labelled information. Then, TU behaviours have been predicted by SSL where a label is unavailable. Two seasons: summer and winter have been considered for this study.

Now, the SSL model has been trained using different training data and investigated by different classification algorithms. Three types of faulty and three types of non-faulty TU patterns have been classified by this model. This classification approach is performed for k-nearest neighbour (kNN) [19] and multiclass support vector machine (MC-SVM). In case of the kNN experiment, the 'k' has been varied by one, three, and five. The MC-SVM has been experimented using two kernel function linear (LMC-SVM) and quadratic (QMC-SVM). The obtained testing accuracy results are tabulated and compared in Table I.

The experiment has been executed using randomly selected data for training and testing phase. The training data have been varied from 10% to 60% and vice versa for testing. The training and testing both performances have been calculated separately to check the robustness of the proposed model. The highest precision (0.998) for training performance has achieved in 1NN using 30% of training data. kNNs have performed well than SVMs in training phase because kNNs find the distance between data points in feature space and a nearest neighbour is the data point itself in the training period, where SVMs find the inner product or solve quadratic function to find out the best margin among support vectors which doesn't deliver as good result as kNN. One nearest neighbour has worked well because of data compactness. In case of testing phase, five nearest neighbour votes deliver better TU pattern recognition than 1KNN. On the other hand, LMC-SVM has gained highest recall (above 95%) in both training and

TABLE I: Accuracy results for different methods.

| Methods	Traning data - 10%		Traning data - 20%		Traning data - 30%		Traning data - 40%		Traning data - 50%		Traning data - 60%	
	Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall
1NN	0.692	0.744	0.774	0.902	0.802	0.918	0.734	0.812	0.774	0.887	0.788	0.902
3NN	0.772	0.987	0.798	0.955	0.789	0.939	0.786	0.891	0.809	0.932	0.823	0.934
5NN	0.799	0.976	0.813	0.961	0.829	0.967	0.804	0.955	0.811	0.919	0.815	0.931
LMC-SVM	0.771	0.995	0.785	0.987	0.831	0.973	**0.837**	**0.978**	0.821	0.946	0.824	0.981
QMC-SVM	0.691	0.898	0.745	0.91	0.753	0.919	0.813	0.948	0.741	0.914	0.772	0.892
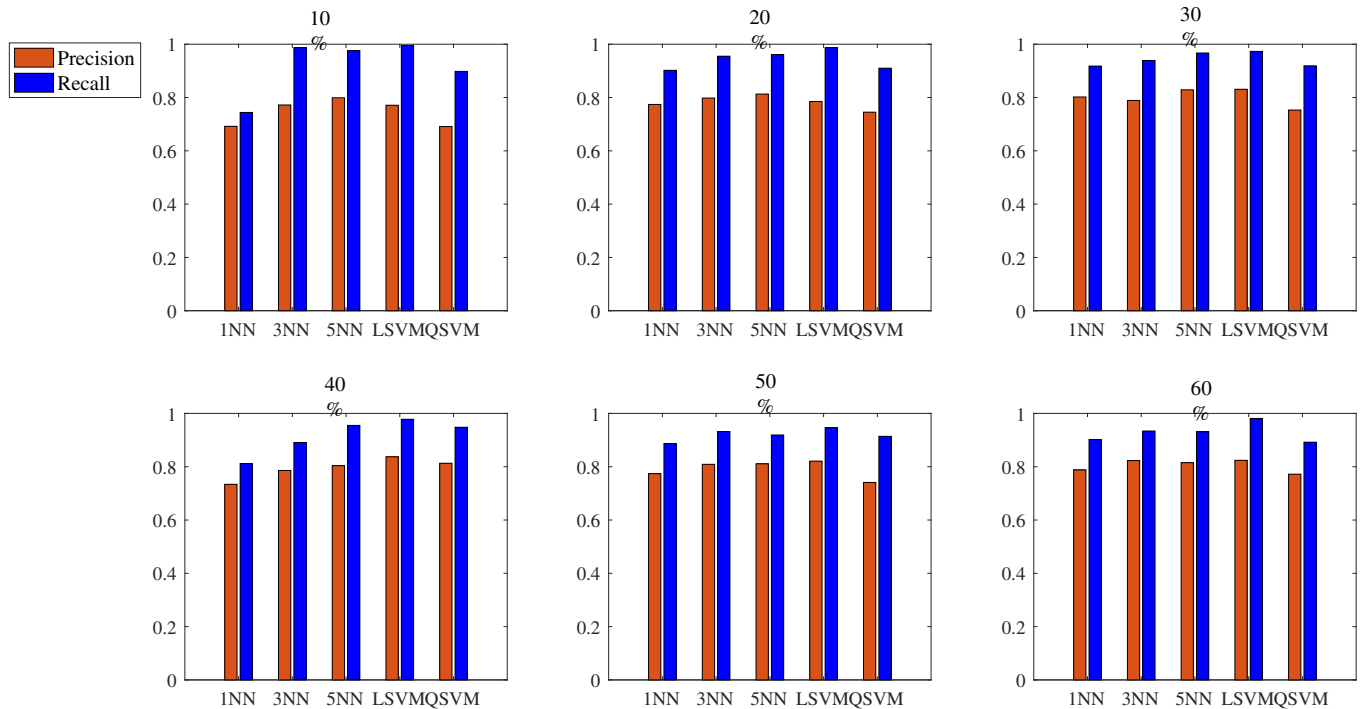
Fig. 4: Comparison of precision and recall score.

testing cases for different amount of data variations. In addition, LMC-SVM has achieved better testing precision (0.837 with 40% training data) than other classification methods estimated in this work. The graphical representation of the performance in testing phases using different classification algorithm have been shown in Fig. 4. In terms of overall precision and recall, linear kernel has worked better than other algorithms. The linear kernel function defines the optimum margin in feature space. Therefore, LMC-SVM has obtained highest performance score among other classifiers.

Thus, LMC-SVM model with 40% training data has been considered most efficient predictor for SSL approach. Further, this model is used for the TU prediction without label information. Consequently, paired t-test has been implemented to discover the correlation between the predicted TU class and the TUs truly belong to that class. Fig. 5 shows the comparison of p-values obtained by the semi-supervised LMC-SVM for different TU classes where, the first three classes represent the non-faulty TU patterns in terms of control temperature and corresponding power demands. Other three classes represent the different faulty TU patterns.
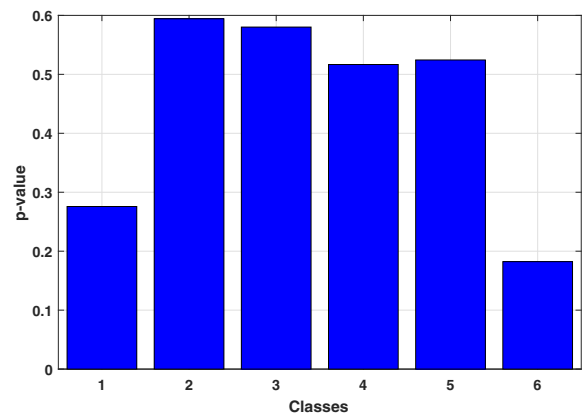

Fig. 5: Obtained p-values for different classes from SSL approach using LMC-SVM.

The significance level 0.5 has been considered and the p-value has been determined for a TU class to justify the

null hypothesis. The null hypothesis has been accepted for a predicted class where p-value is more than the significance level. Fig. 5 shows that the predicted class-1 and class-6 have failed to fit in the actual classes, i.e. the semi-supervised LMC-SVM could predict the class-2, 3, 4, and 5 correctly but unsuccessful in predicting the TUs from class-1 and class-6. It is observed from the results that available training data for class 1 and 6 might not be sufficient to train the LMC-SVM.

V. CONCLUSION & FUTURE WORK

Automated faulty and non-faulty TU prediction has been investigated using two different classification algorithms and the variation of five different parameters. The promising results of semi-supervised learning (SSL) algorithm shows that it performs well (overall accuracy 90%) compared to the supervised learning algorithm for these TU data in previous work by the authors. Thus, unlabelled data can be effectively classified using SSL approaches if historically labelled data is available. It is also found that the performance of LMC-SVM is the best-fitted model among five-tested methods for training these datasets. Based on the paired t-test results, LMC-SVM would need to be improved in the one faulty and one non-faulty classes case. Thus, more training data and other classification algorithms are being investigated to improve future SSL performance.

VI. ACKNOWLEDGEMENTS

REFERENCES

[1] P. Usoro, I. Schick, and S. Negahdaripour, "An innovation-based methodology for hvac system fault detection," *Journal of dynamic systems, measurement, and control*, vol. 107, no. 4, pp. 284–289, 1985.

[2] D. Anderson, L. Graves, W. Reinert, J. Kreider, J. Dow, and H. Wubbena, "A quasi-real-time expert system for commercial building hvac diagnostics," *ASHRAE Transactions (American Society of Heating, Refrigerating and Air-Conditioning Engineers);(USA)*, vol. 95, no. CONF-890609, 1989.

[3] D. Li, G. Hu, and C. J. Spanos, "A data-driven strategy for detection and diagnosis of building chiller faults using linear discriminant analysis," *Energy and Buildings*, vol. 128, pp. 519–529, 2016.

[4] A. Beghi, R. Brignoli, L. Cecchinato, G. Menegazzo, M. Rampazzo, and F. Simmini, "Data-driven fault detection and diagnosis for hvac water chillers," *Control Engineering Practice*, vol. 53, pp. 79–91, 2016.

[5] R. Isermann, *Fault-diagnosis applications: model-based condition monitoring: actuators, drives, machinery, plants, sensors, and fault-tolerant systems*. Springer Science & Business Media, 2011.

[6] Y. Zhao, S. Wang, and F. Xiao, "Pattern recognition-based chillers fault detection method using support vector data description (SVDD)," *Applied Energy*, vol. 112, pp. 1041–1048, 2013.

[7] S. R. West, Y. Guo, X. R. Wang, and J. Wall, "Automated fault detection and diagnosis of hvac subsystems using statistical machine learning," in *12th International Conference of the International Building Performance Simulation Association*, 2011.

[8] E. Alpaydin, *Introduction to machine learning*. MIT press, 2014.

[9] C.-F. Chien, S.-L. Chen, and Y.-S. Lin, "Using bayesian network for fault location on distribution feeder," *IEEE Transactions on Power Delivery*, vol. 17, no. 3, pp. 785–793, 2002.

[10] W.-Y. Lee, J. M. House, C. Park, and G. E. Kelly, "Fault diagnosis of an air-handling unit using artificial neural networks," *Transactions American Society of Heating Refrigerating and Air Conditioning Engineers*, vol. 102, pp. 540–549, 1996.

[11] J. Liang and R. Du, "Model-based fault detection and diagnosis of hvac systems using support vector machine method," *International Journal of refrigeration*, vol. 30, no. 6, pp. 1104–1114, 2007.

[12] A. L. Dexter and D. Ngo, "Fault diagnosis in air-conditioning systems: a multi-step fuzzy model-based approach," *HVAC&R Research*, vol. 7, no. 1, pp. 83–102, 2001.

[13] M. Dey, M. Gupta, M. Turkey, and S. Dudley, "Unsupervised learning techniques for, hvac terminal unit behaviour analysis," in *2017 IEEE Conference on Smart City Innovation (SCI)*, Aug 2017.

[14] M. Dey, M. Gupta, S. P. Rana, M. Turkey, and S. Dudley, "A pid inspired feature extraction method for hvac terminal units," in *2017 IEEE Conference on Technologies for Sustainability (SusTech)*, Nov 2017, pp. 1–5.

[15] M. Dey, S. P. Rana, and S. Dudley, "Smart building creation in large scale hvac environments through automated fault detection and diagnosis," *Future Generation Computer Systems*, 2018.

[16] K. Crammer and Y. Singer, "On the algorithmic implementation of multiclass kernel-based vector machines," *Journal of machine learning research*, vol. 2, no. Dec, pp. 265–292, 2001.

[17] S. P. Rana *et al.*, "UWB localization employing supervised learning method," in *2017 IEEE 17th International Conference on Ubiquitous Wireless Broadband (ICUWB)*, Sept 2017, pp. 1–5.

[18] P. G. Hoel *et al.*, "Introduction to mathematical statistics." *Introduction to mathematical statistics.*, no. 2nd Ed, 1954.

[19] S. M. Weiss and C. A. Kulikowski, *Computer systems that learn: classification and prediction methods from statistics, neural nets, machine learning, and expert systems*. Morgan Kaufmann Publishers Inc., 1991.